Deep Learning Workshop

Nov. 20, 2015

Andrew Fishberg, Rowan Zellers

Why deep learning?

Deep learning

Yann LeCun, Yoshua Bengio & Geoffrey Hinton

Affiliations | Corresponding author

Nature **521**, 436–444 (28 May 2015) | doi:10.1038/nature14539

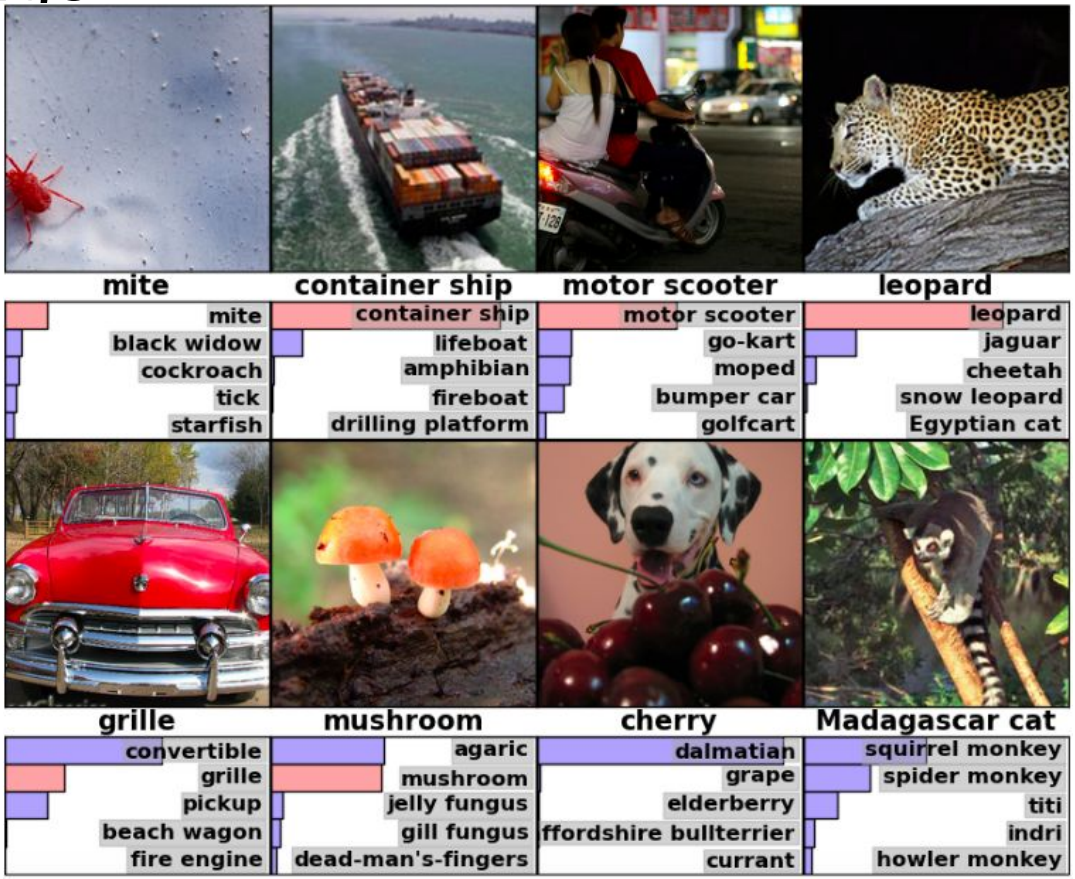
Received 25 February 2015 | Accepted 01 May 2015 | Published online 27 May 2015

learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

The ImageNet Challenge

Goal: image classification with 1000 categories

Top 5 error rate of 15%.



Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

The basics of deep learning

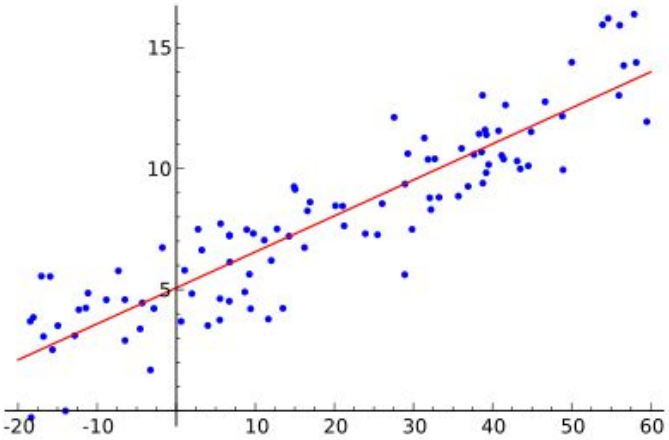
Abstract

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

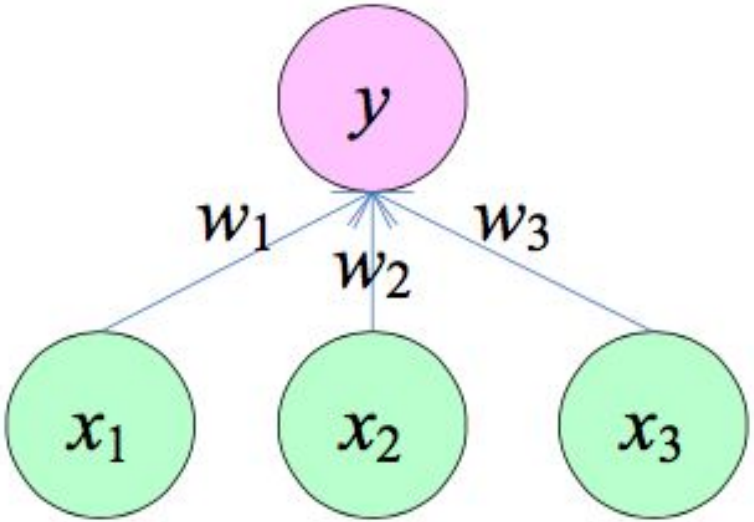
Linear regression

A model which takes an input vector x, and outputs a single number y.

$$y = \mathbf{w}^T \mathbf{x} + b$$



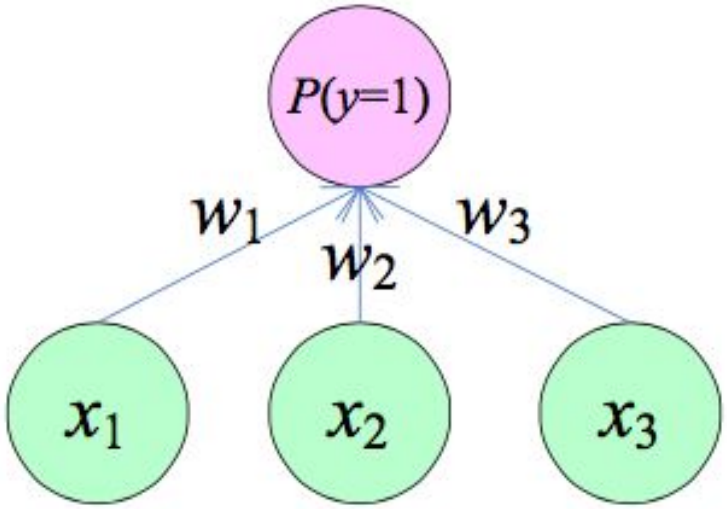
Source: wikipedia



Logistic regression

Like linear regression, except we compute probabilities. A probability must be in the range [0,1].

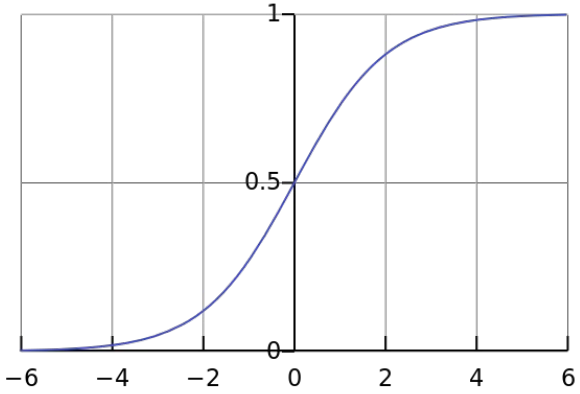
$$P(y=1) = \frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}-b}}$$



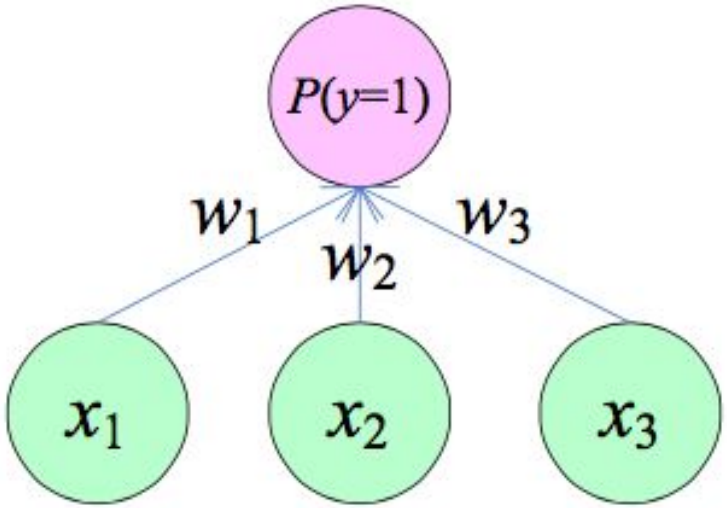
Logistic regression

Like linear regression, except we compute probabilities. A probability must be in the range [0,1].

$$P(y=1) = \frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}-b}}$$



Source: wikipedia



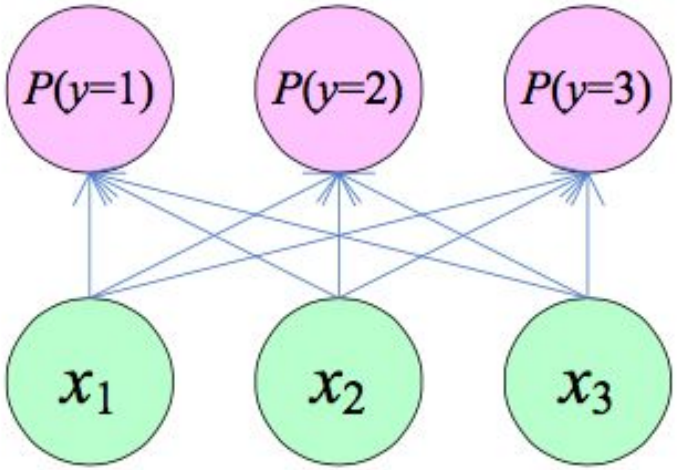
Sigmoid function:

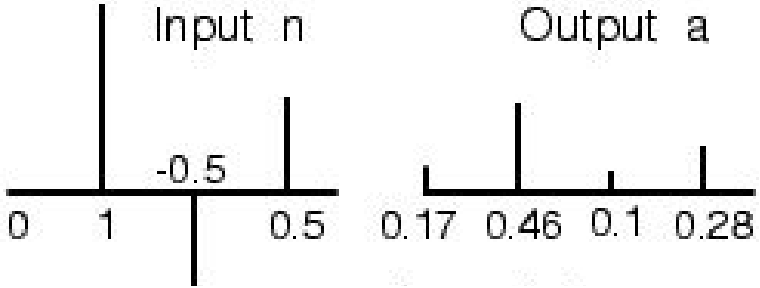
$$\sigma(x) = \frac{1}{1+e^{-x}}$$

Multiclass Logistic regression

We can extend logistic regression to compute the probability of several classes.

$$P(y=k) = \frac{e^{\mathbf{w}_k^T \mathbf{x} + b}}{\sum_j e^{\mathbf{w}_j^T \mathbf{x} + b}}$$





Softmax function:

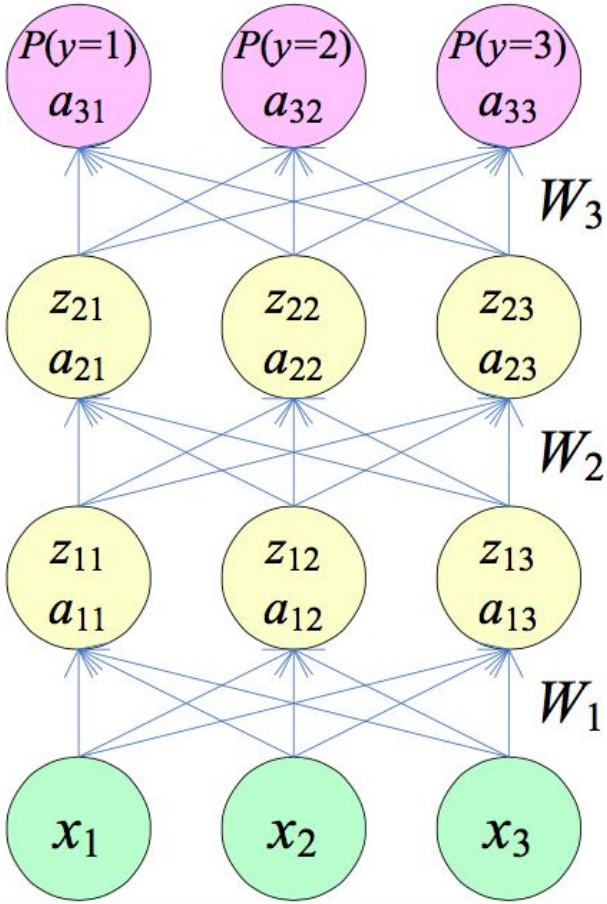
$$\sigma(x)_i = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

Source: http://radio.feld.cvut.cz/matlab/toolbox/nnet/

Deep Neural Networks!

- Big idea: string together a collection “layers”, with multiclass logistic regression at the end, and linear regressors in the middle.

- We need nonlinear “activation functions”



Deep Neural Networks!

Example: to compute the pre-activation a_{2,1}

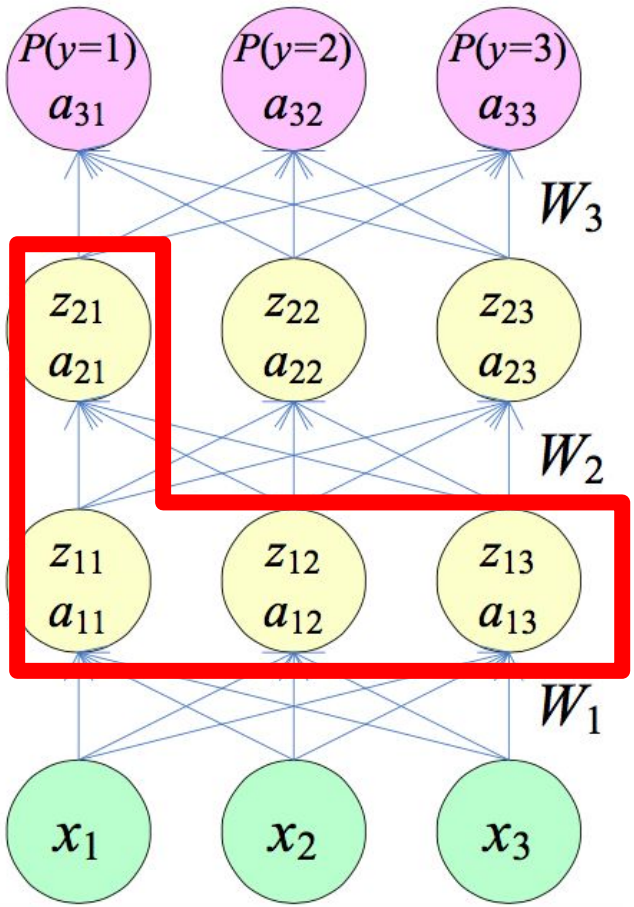
$$a_{2,1} = \mathbf{w}_{2,1}^T \mathbf{z}_{1,1} + \mathbf{w}_{2,1}^T \mathbf{z}_{1,2} + \mathbf{w}_{2,1}^T \mathbf{z}_{1,3} + b_{2,1}$$

To compute the post-activation z_{2,1}, we run a_{2,1} through a nonlinear function.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$$ReLU(X) = max(x, 0)$$

$$z_{2,1} = max(a_{2,1}, 0)$$

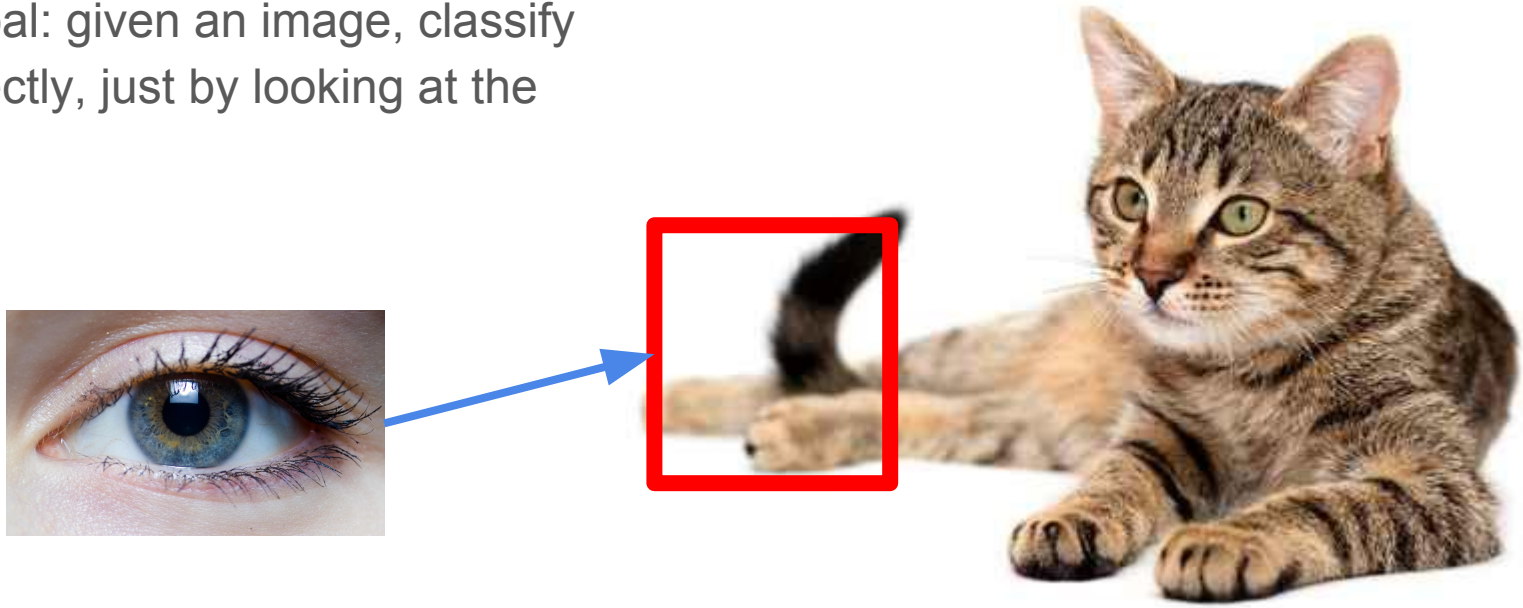


Outline of convolutional neural networks

Our goal: given an image, classify it correctly, just by looking at the pixels!

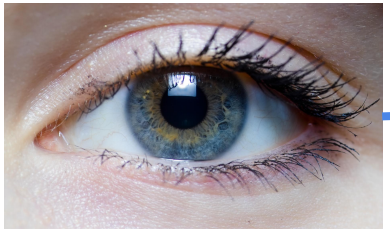
Outline of convolutional neural networks

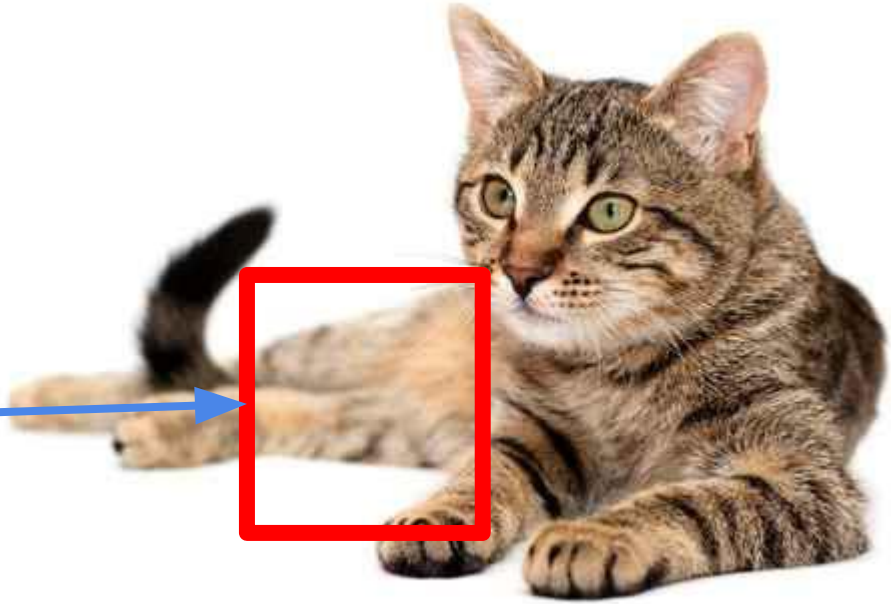
Our goal: given an image, classify it correctly, just by looking at the pixels!



Outline of convolutional neural networks

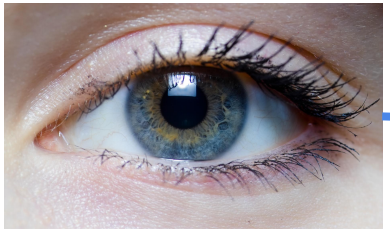
Our goal: given an image, classify it correctly, just by looking at the pixels!

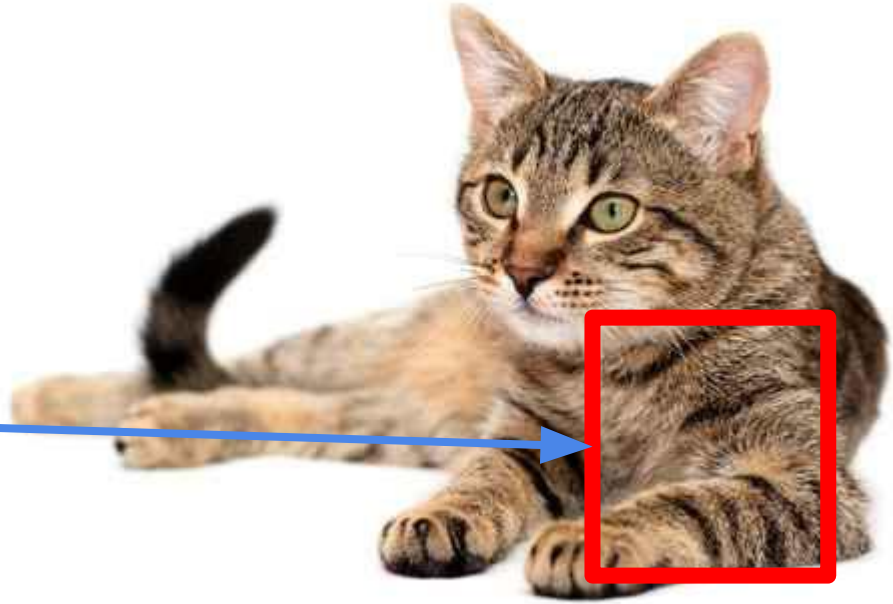




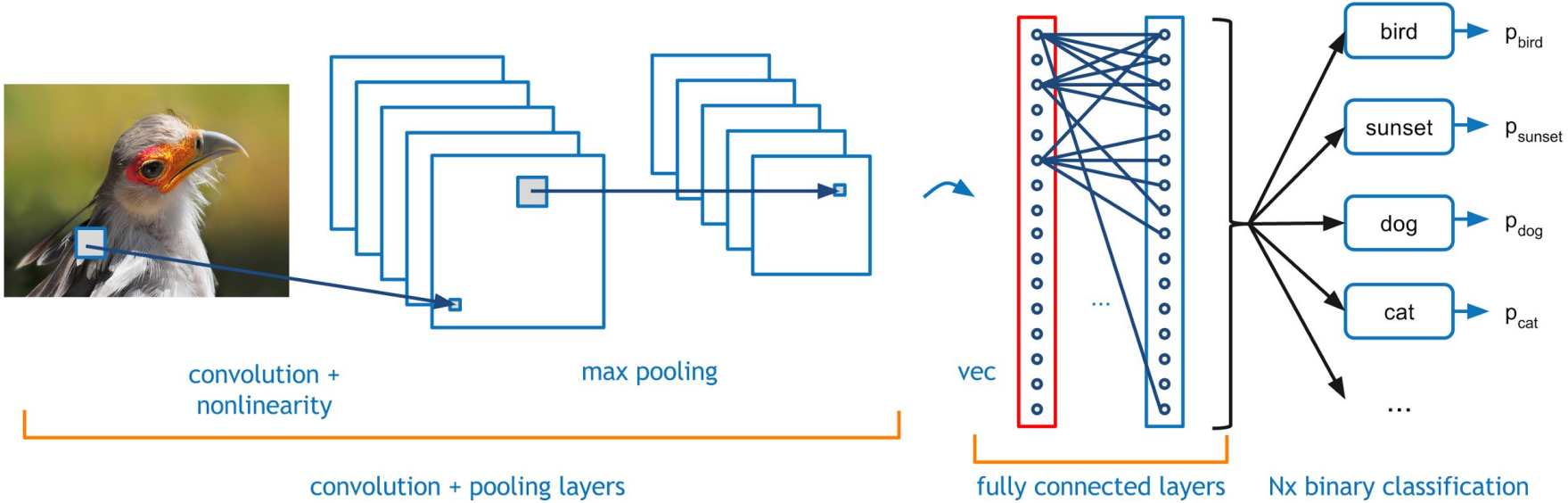
Outline of convolutional neural networks

Our goal: given an image, classify it correctly, just by looking at the pixels!





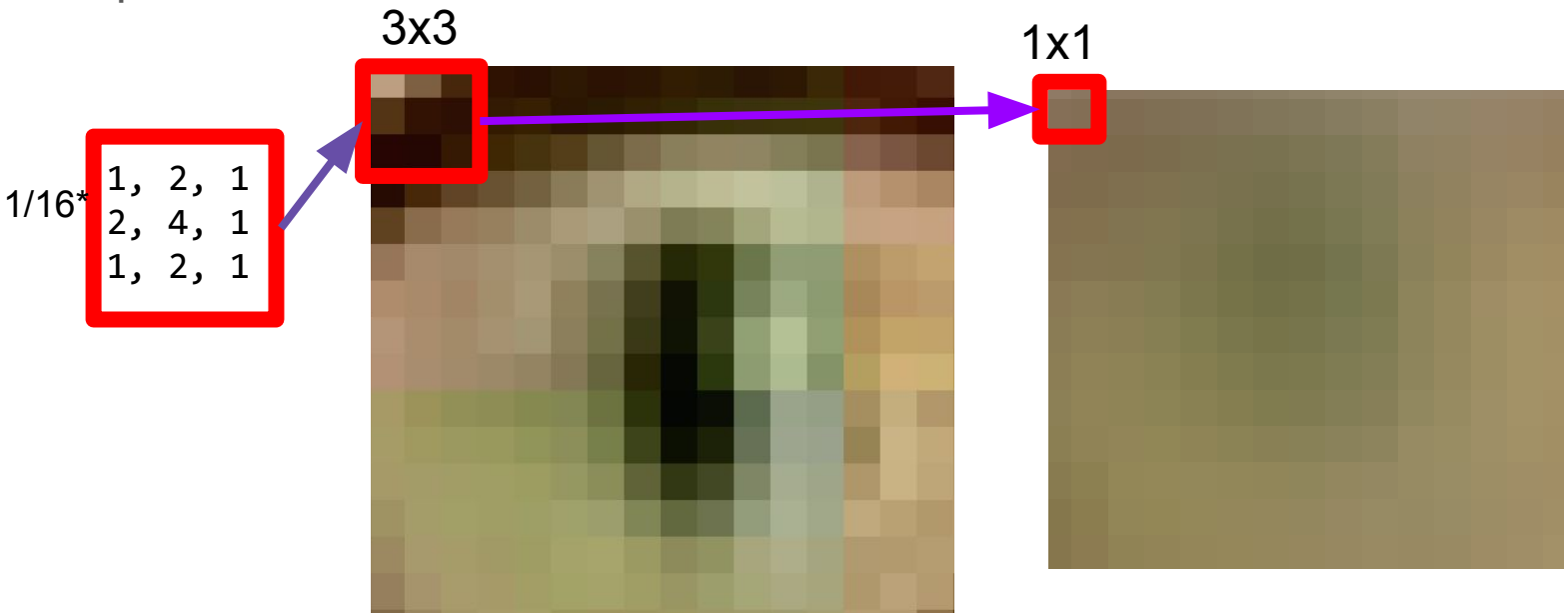
Outline of convolutional neural networks



source: https://code.flickr.net/2014/10/

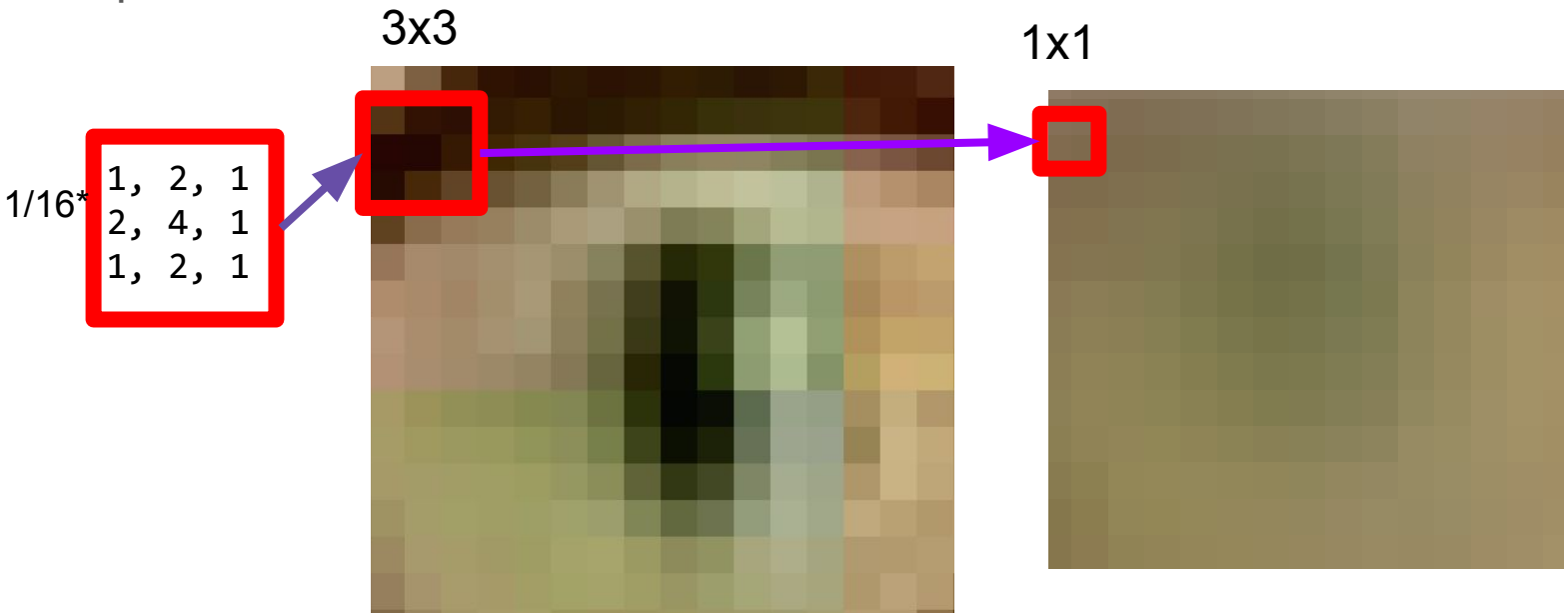
What is convolution?

For each chunk of an image, multiply it elementwise by a kernel, then sum the response.



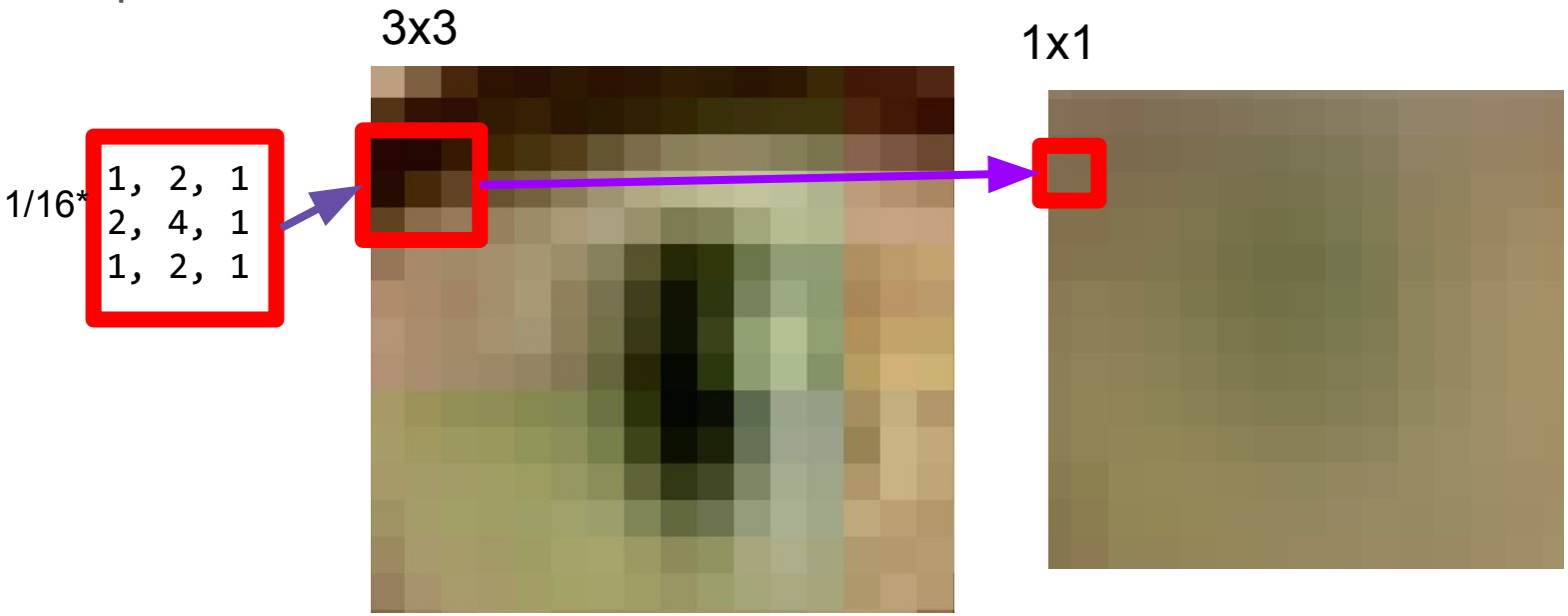
What is convolution?

For each chunk of an image, multiply it elementwise by a kernel, then sum the response.



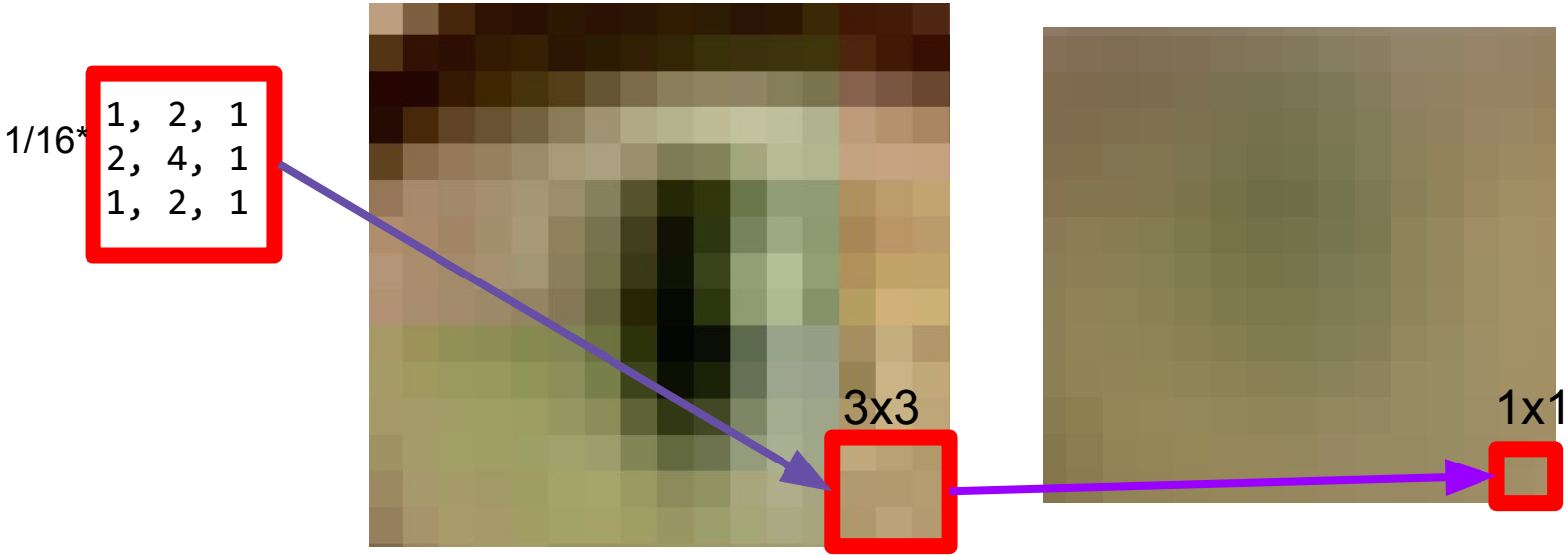
What is convolution?

For each chunk of an image, multiply it elementwise by a kernel, then sum the response.

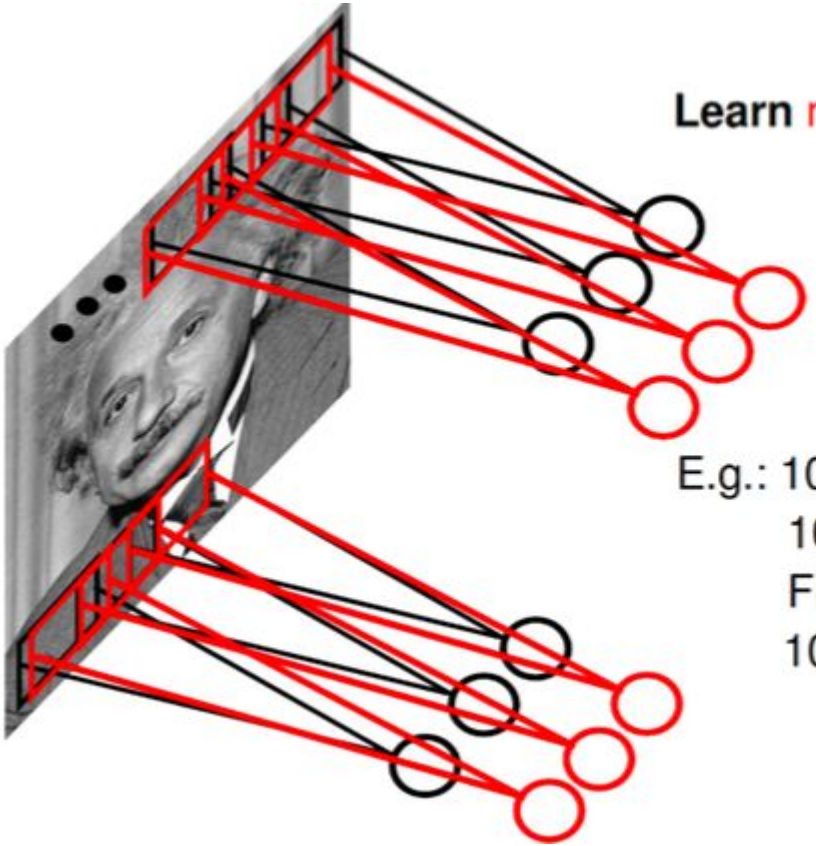


What is convolution?

For each chunk of an image, multiply it elementwise by a kernel, then sum the response.



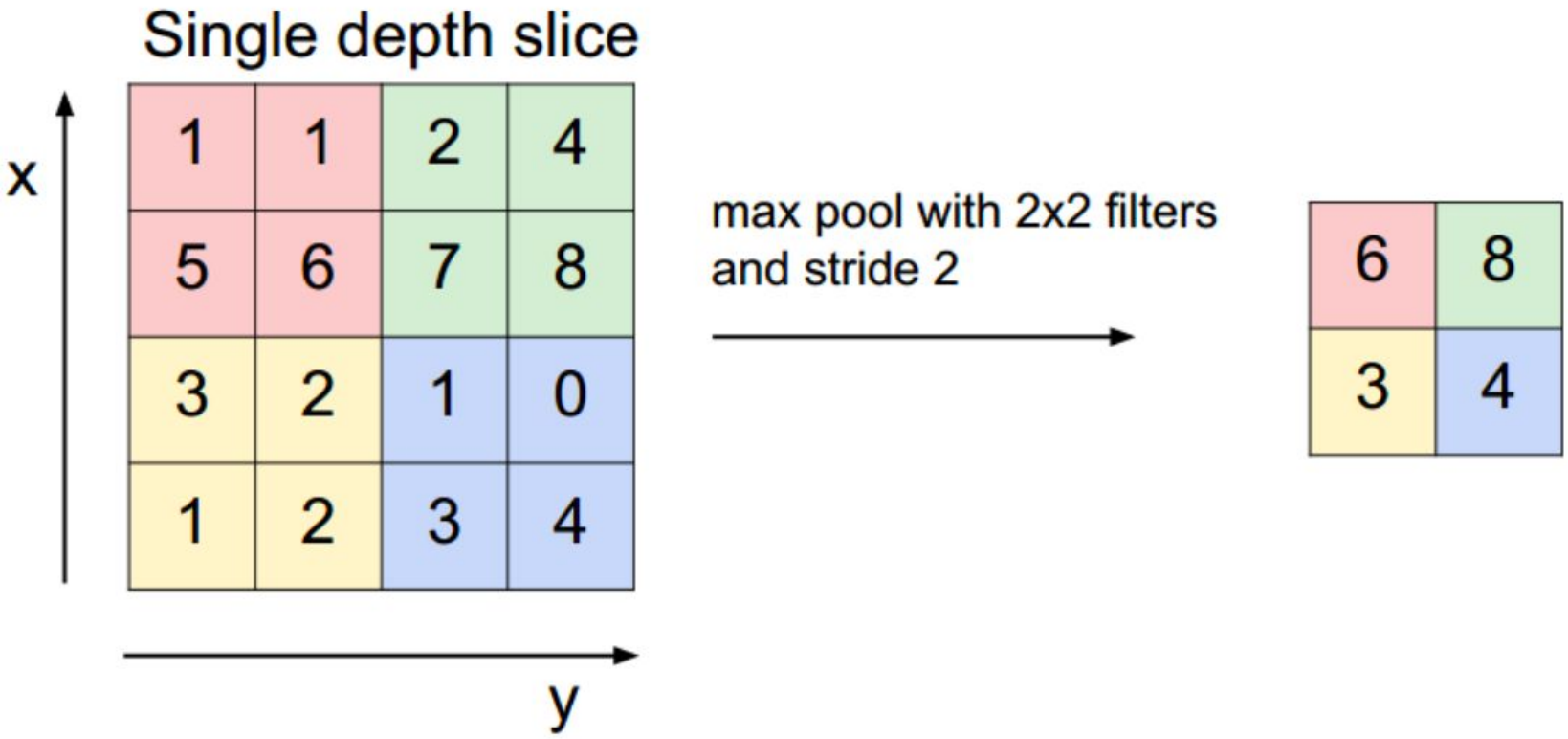
Convolutional layers



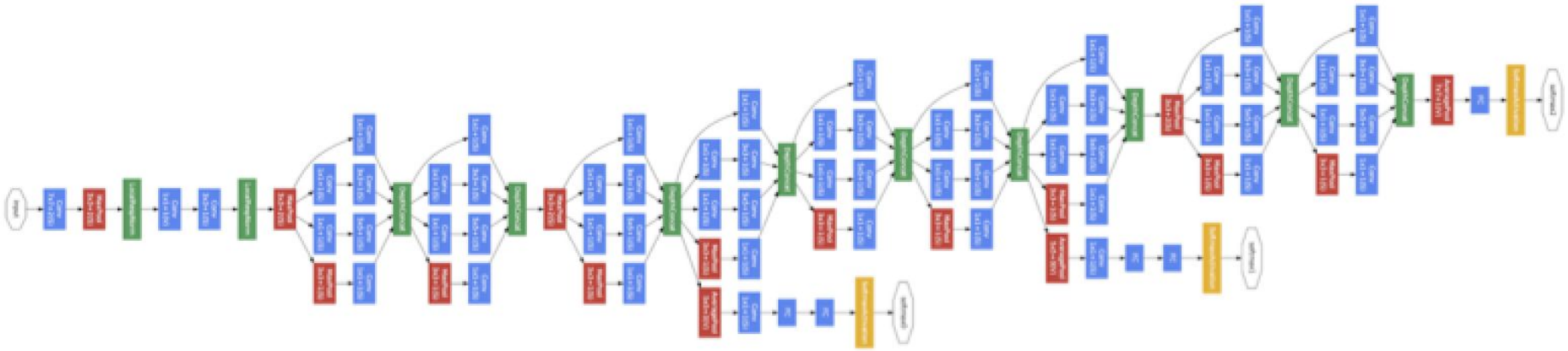
Learn multiple filters.

E.g.: 1000x1000 image
100 Filters
Filter size: 10x10
10K parameters

Max-Pooling


Single depth slice
x
y
1 1 2 4
5 6 7 8
3 2 1 0
1 2 3 4
max pool with 2x2 filters
and stride 2
6 8
3 4

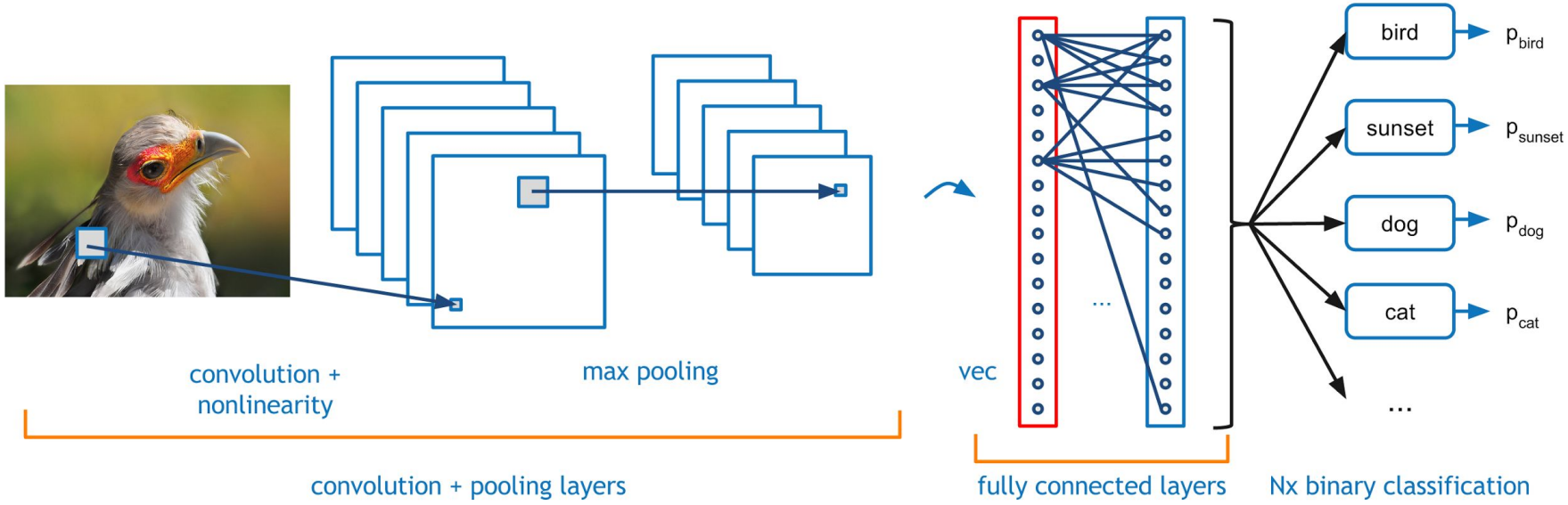
http://www.csc.kth.se/~roelof/deepdream/googlenet2.png



GoogLeNet

Convolution
Pooling
Softmax
Other

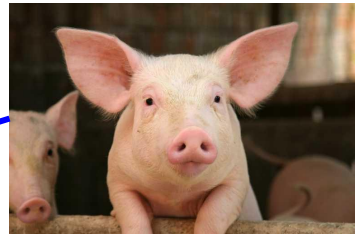
Putting it all together



Training Deep Neural Networks

Stochastic Gradient Descent!

- Calculate the error E for a “batch” of training images
- Update each parameter by moving a little bit (eta) away from the gradient
  - (As we’d like to minimize F(x)).

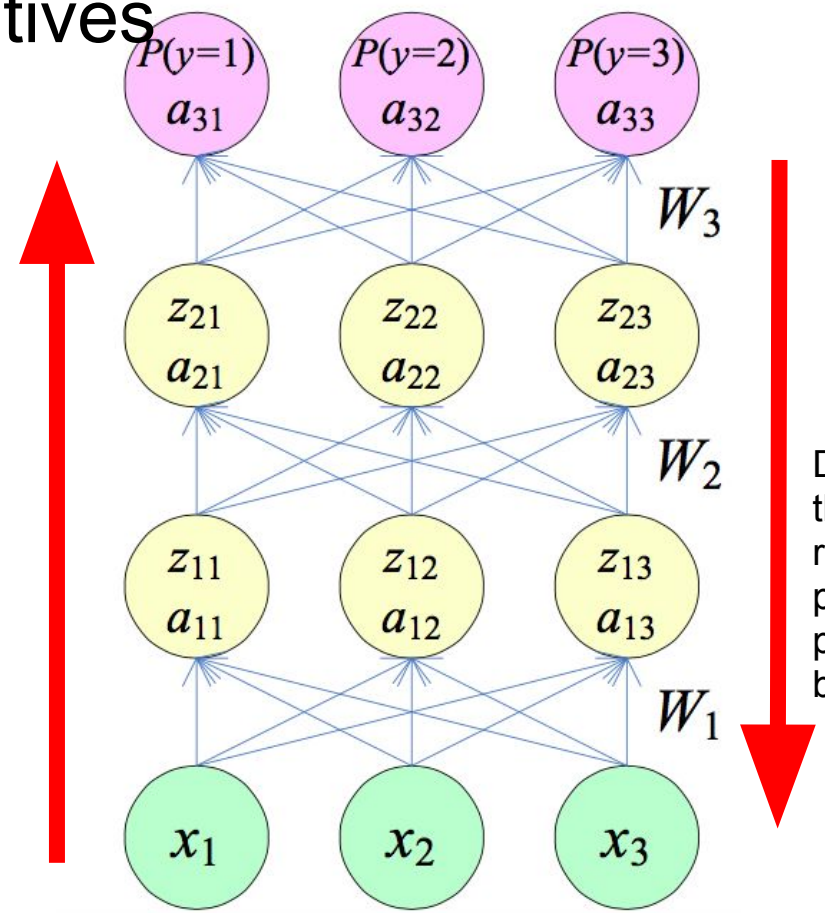


$$w_{n+1} = w_n - \eta \cdot \frac{\partial}{\partial w_n} E(\text{batch})$$

Computing the derivatives

$$w_{n+1} = w_n - \eta \cdot \frac{\partial}{\partial w_n} E(\text{batch})$$

Activations propagate forwards



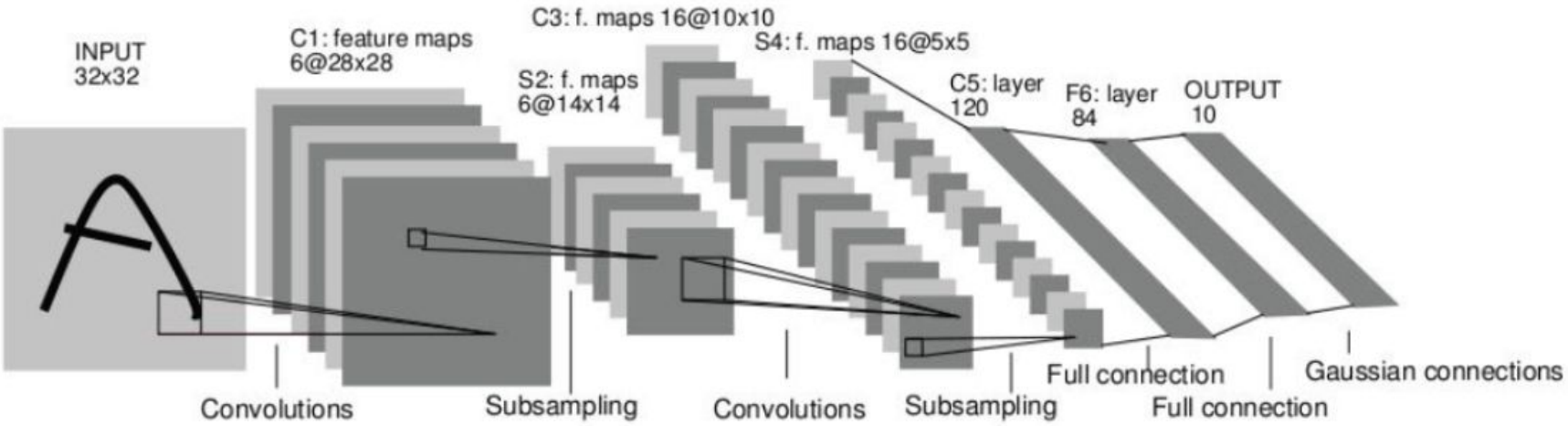
Derivatives of the error (with respect to each parameter) propagate backwards

The MNIST corpus

A benchmark corpus for multiclass classification

- 28x28 grayscale images, scaled and centered
- 10 classes
- 60,000 training + 10,000 test instances

Your job!



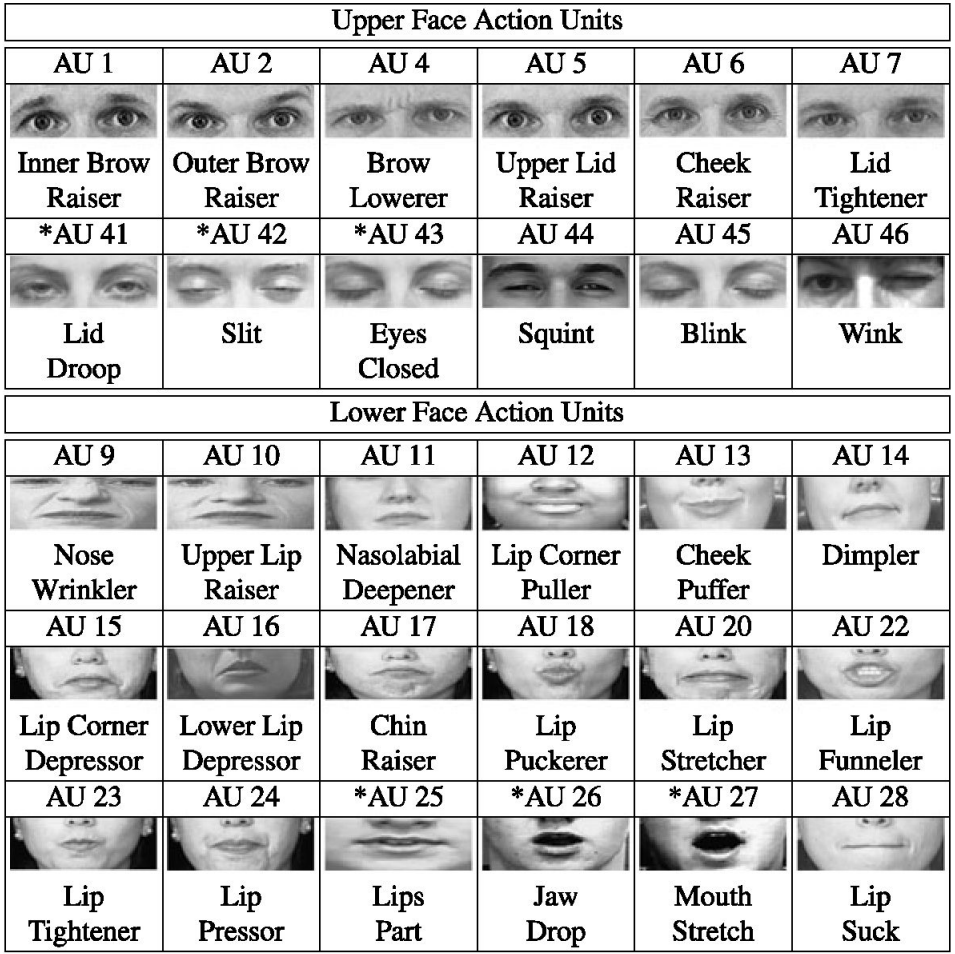
http://rownz.com/deeplearning/

Our own applications of deep learning!

Rowan -- Facial Action Unit detection

Andrew -- Image Processing w/ Deep Learning on Aerial Imagry

Facial Action Unit detection

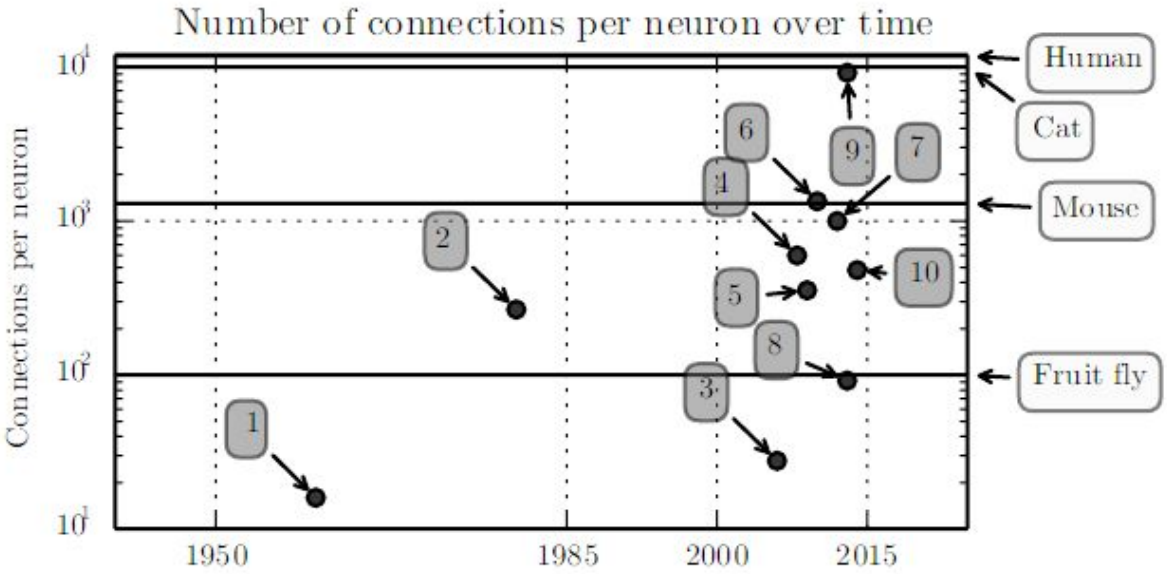


http://goodfeli.github.io/dlbook/contents/intro.html

Source: Deep Learning -- Ian Goodfellow, Aaron Courville, and Yoshua Bengio

Motivation for Andrew's Work

- Field has been around for over 60 years
  - 1940s-1960s known as ***Cybernetics***
  - 1980s-1990s known as ***Connectionism***
  - 2006-Today known as ***Deep Learning***



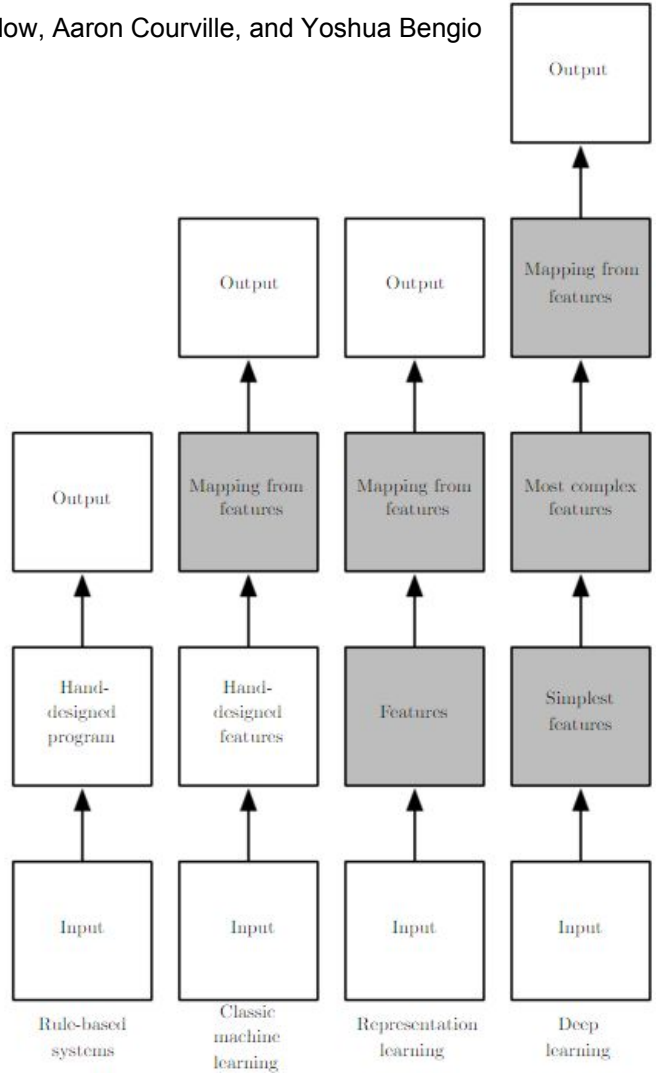
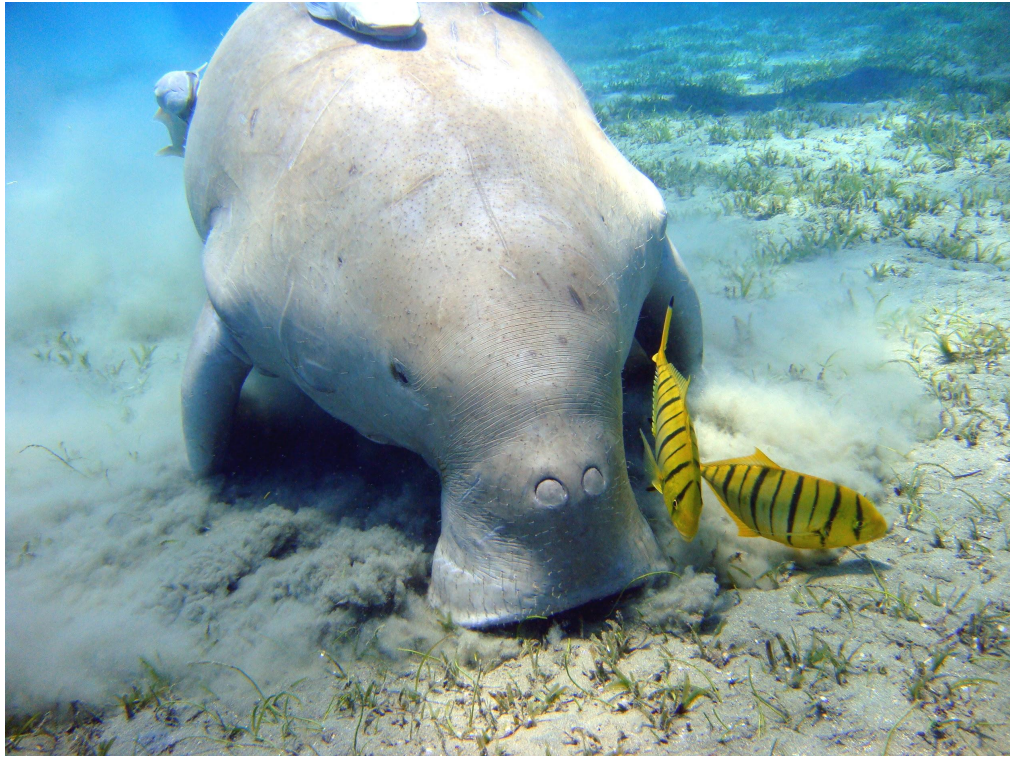


Image Detection from Aerial Imagery



https://upload.wikimedia.org/wikipedia/commons/9/91/Dugong_Marsa_Alam.jpg

A Convolutional Neural Network for Automatic Analysis of Aerial Imagery

Frederic Maire, Luis Mejias
School of Electrical Engineering and Computer Science,
Queensland University of Technology, Brisbane, Australia
f.maire@qut.edu.au, luis.mejias@qut.edu.au

Amanda Hodgson
Murdoch University Cetacean Research Unit,
Murdoch University, Perth, Australia
a.hodgson@murdoch.edu.au

***Abstract*—This paper introduces a new method to automate the detection of marine species in aerial imagery using a Machine Learning approach. Our proposed system has at its core, a convolutional neural network. We compare this trainable classifier to a handcrafted classifier based on color features, entropy and shape analysis. Experiments demonstrate that the convolutional neural network outperforms the handcrafted solution. We also introduce a negative training example-selection method for situations where the original training set consists of a collection of labeled images in which the objects of interest (positive examples) have been marked by a bounding box. We show that picking random rectangles from the background is not necessarily the best way to generate useful negative examples with respect to learning.**

I. Introduction

This paper introduces a technique that can be used to automate the processing of images analysis taken during an aerial survey using a custom payload onboard an Unmanned Aerial Vehicle (UAV). The amount of data produced by such flights is considerable (tens of thousands of images) making the process of manual review intractable [1]. Similar applications of UAVs for surveillance and monitoring in areas such as agriculture, law enforcement, equipment and infrastructure inspection, etc., could benefit from automated image analysis. This type of automation could radically reduce the human hours needed to perform tasks in these fields. Our particular interest was in the automatic detection of marine mammals in images taken from an aircraft (manned or unmanned). We introduce a machine learning approach using convolutional neural networks that automatically detect and annotate marine mammals (dugongs) in images. In the area of vision-based marine mammal identification, most image processing techniques investigated to date fall in the category of low-level image processing [2], [3], [4], [5], [6]. Whilst straightforward, this approach is prone to high rates of false detection which makes the approach in some instances unreliable. Machine learning, and in particular Convolutional Neural Networks (CNNs) has the potential to provide enormous improvements in the automated detection of marine fauna, and other similar applications.

A. The specific challenges of marine mammal detection

What makes the detection of dugongs particularly challenging is that their appearance varies dramatically with the sea conditions. Their apparent color changes with the depth and the turbidity of the water. Although the shape of a dugong is relatively rigid, their tail is not always visible. Moreover, parts of their bodies can be covered by small waves with breaking crests or whitecaps as can be seen in Figure 1. The appearance of the dugongs depends also on the sea floor. In Figure 2, some of the grazing dugongs are hardly distinguishable from the background.

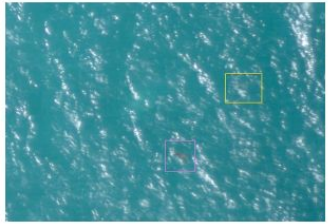


Fig. 1. There are two dugongs in this 1232×840 image. The dugong at the center of the purple rectangle is very salient compared to the dugong at the center of the yellow rectangle. Dugongs can be partially occluded by wavecrests.

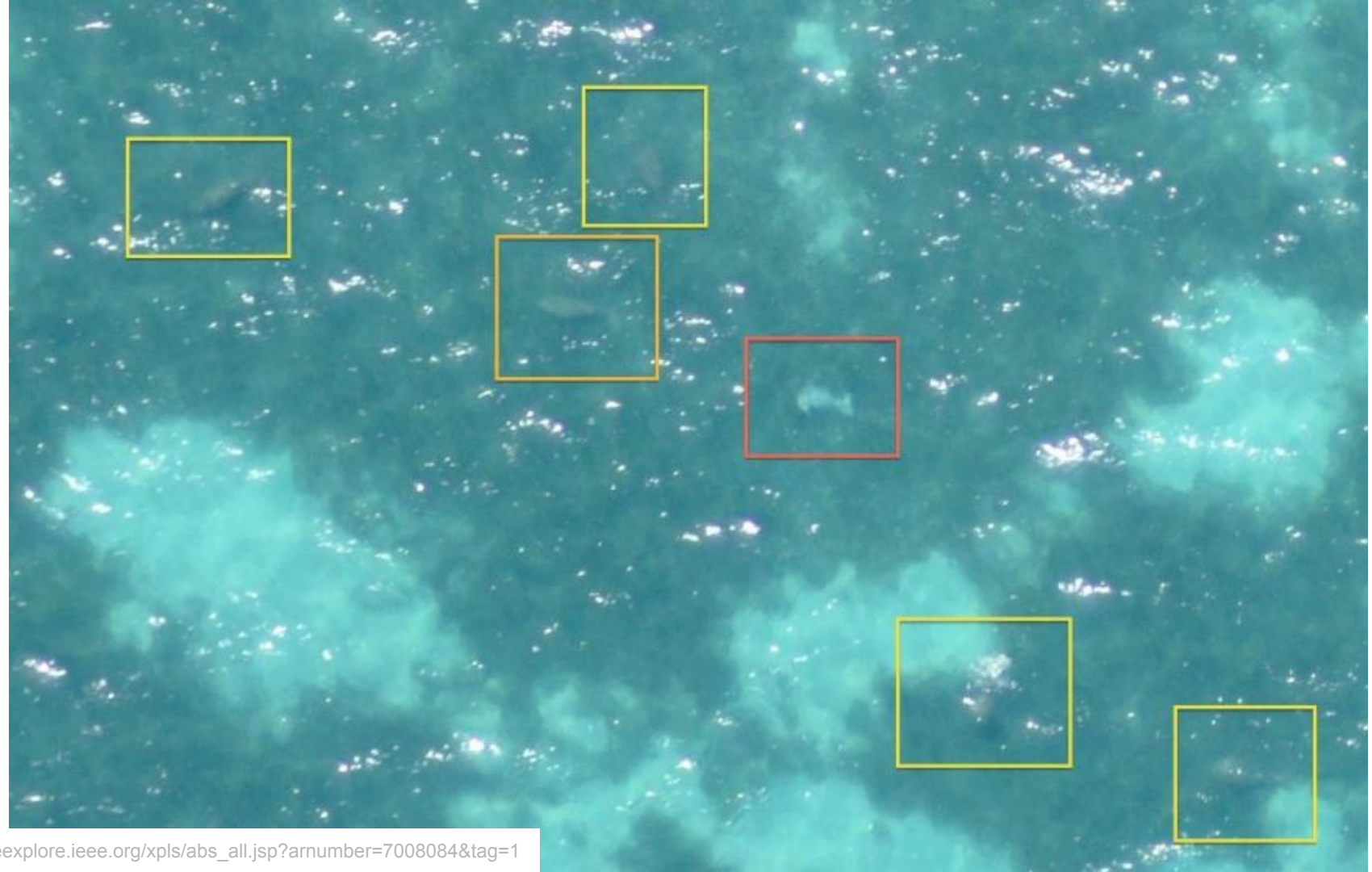
This paper is structured as follows. Section II reviews recent work in CNN for image analysis. Section III describes the CNN approach proposed this paper. Section IV describes how additional training samples are generated. Section V outlines the negative example-selection method. Section VI presents the outcomes and analysis of data. Finally, section VII presents concluding remarks.

II. Related Work

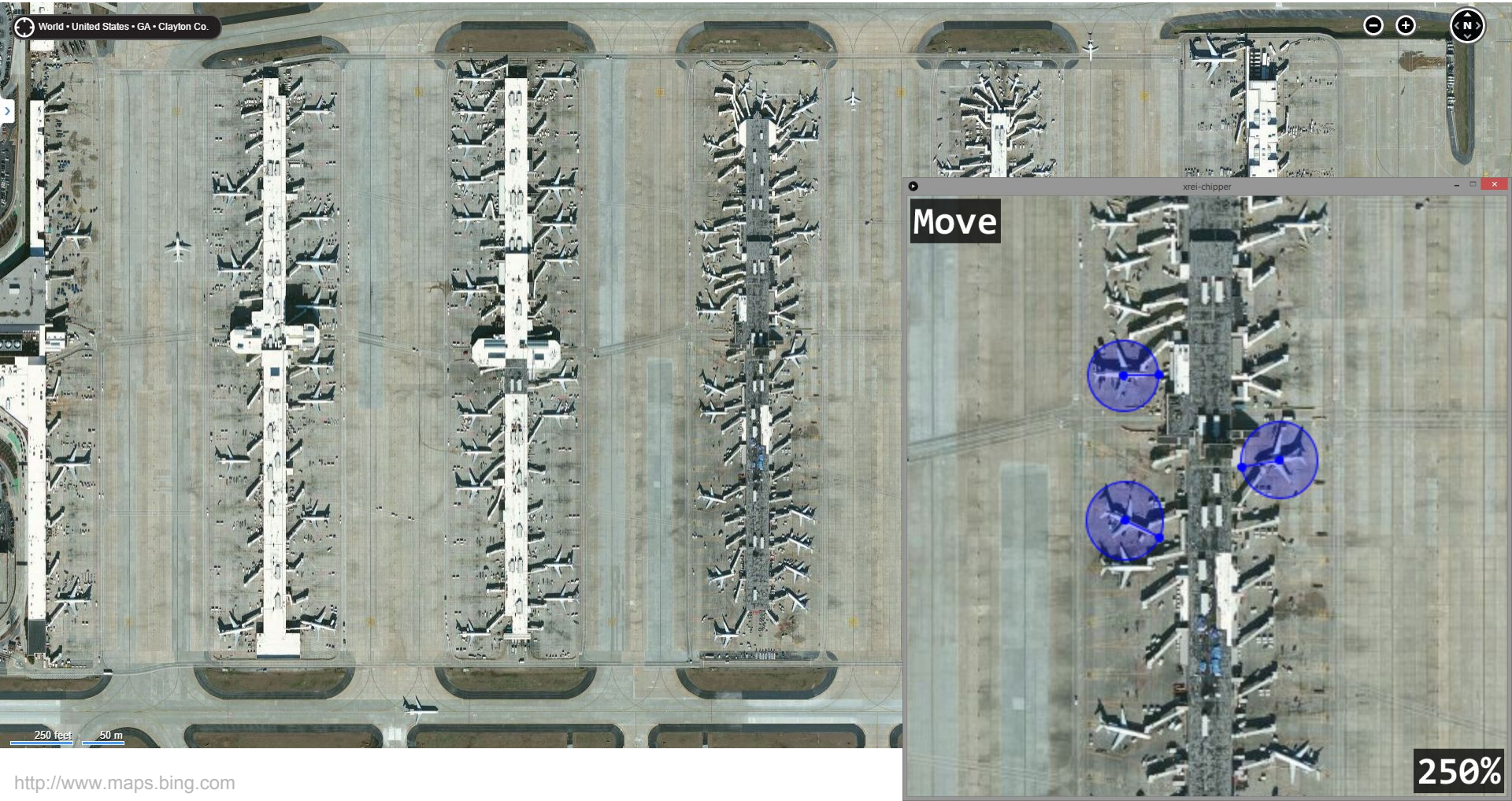
The literature on marine mammal detection using electro optical sensors is not extensive. Infrared and standard cameras have been used to perform detection from aerial platforms [7], [8], [9], [10], [11], [3], [4]. The main limitations of this approach have been identified the environmental conditions, and their effects on illumination within the images. Despite these challenges, visual imagery or vision is an attractive solution given it offers a rich and permanent source of information, and is easily generalisable to many types of aircraft. The analytical approaches presented in [3], [4] uses color segmentation and blob shape analysis. These two attempts represent a significant

978-1-4799-5409-4/14/$31.00 ©2014 IEEE

http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7008084&tag=1

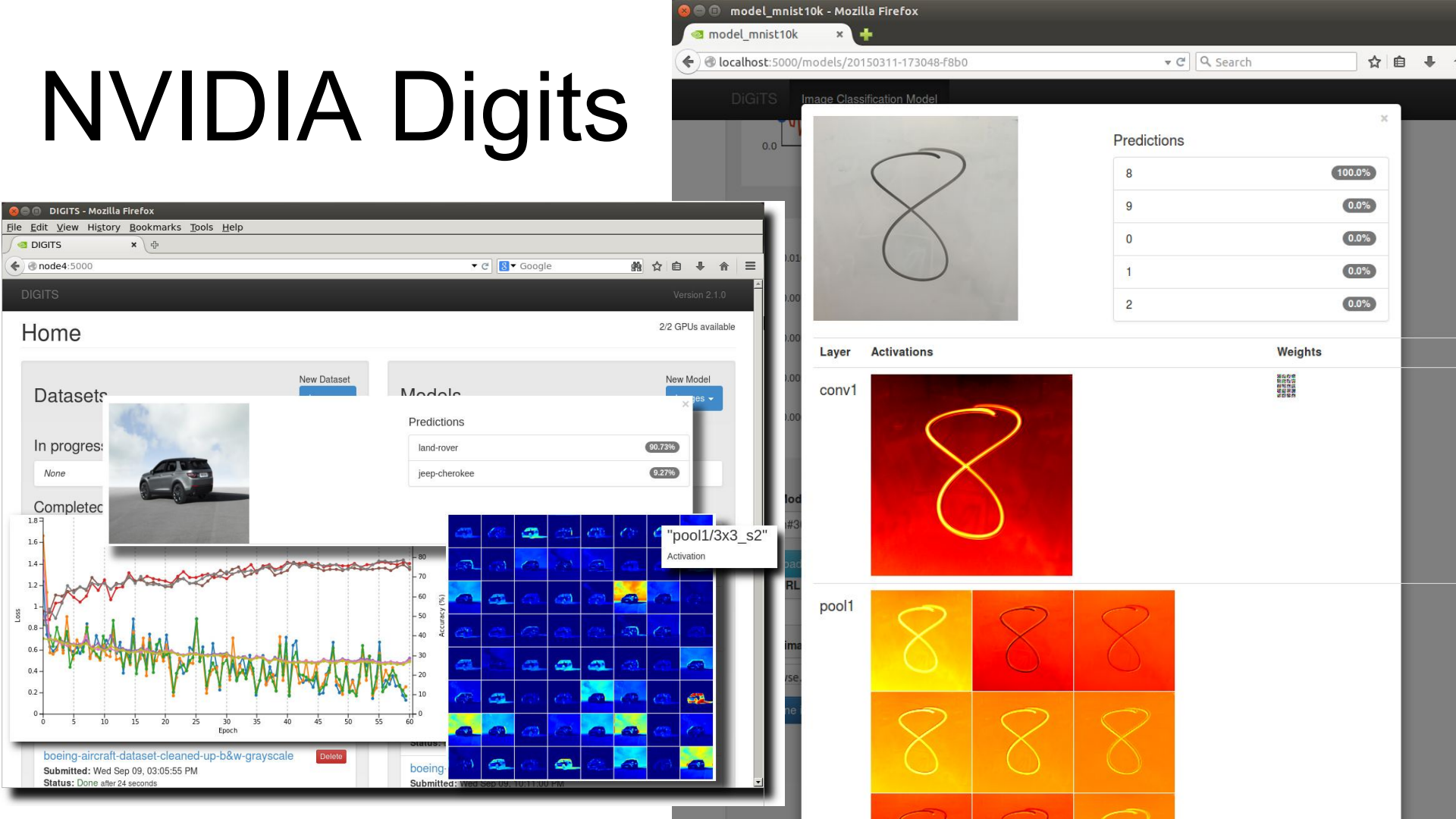


http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7008084&tag=1



http://www.maps.bing.com

NVIDIA Digits

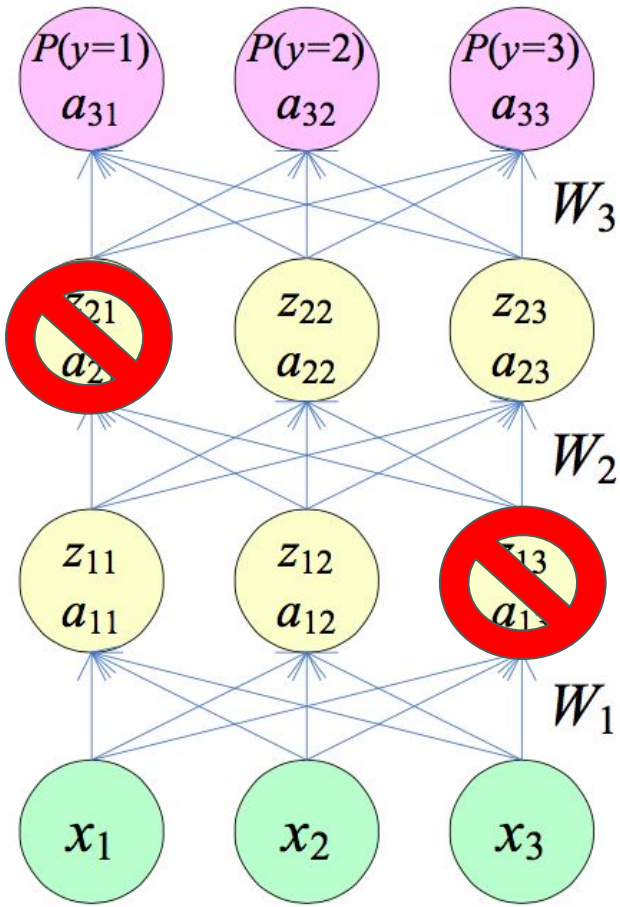


More advanced things!

Black magic is what really makes deep learning work

Dropout

On each training batch, randomly select nodes to temporarily turn off



Dropout

On each training batch, randomly select nodes to temporarily turn off

Tends to massively increase performance!