Imperial College London

Computer Vision Topics

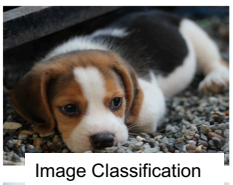
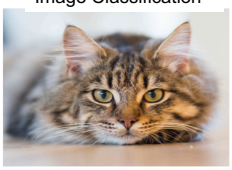


Image Classification



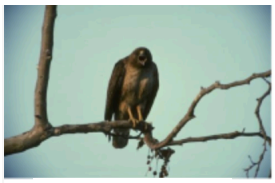
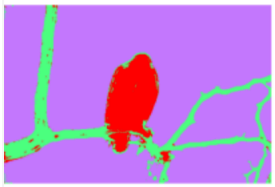
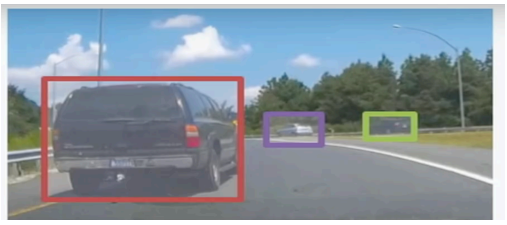


Image Segmentation



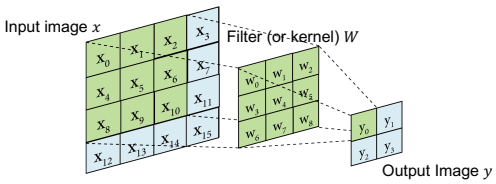


Object Localization (1 object) or Detection (several objects)

1

Imperial College London

Making Each Kernel Operation Non-Linear



$$y_0 = w_0x_0+w_1x_1+w_2x_2+w_3x_4+w_4x_5+w_5x_6+w_6x_8+w_7x_9+w_8x_{10}$$

A bias term can be added

$$y_0 = w_0x_0+w_1x_1+w_2x_2+w_3x_4+w_4x_5+w_5x_6+w_6x_8+w_7x_9+w_8x_{10} + \boldsymbol{b}$$

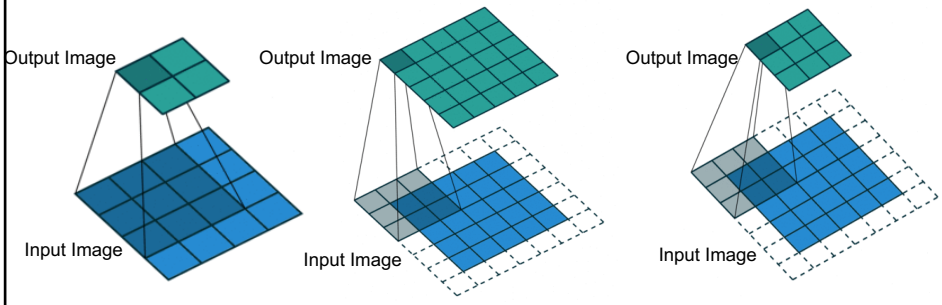
And a non-linear transformation is applied by an activation function g

$$y_0 = \boldsymbol{g}(w_0x_0+w_1x_1+w_2x_2+w_3x_4+w_4x_5+w_5x_6+w_6x_8+w_7x_9+w_8x_{10} + \boldsymbol{b})$$

2

Imperial College London

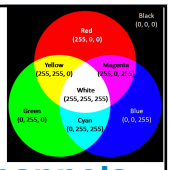
Key Parameters of Convolutional Neural Networks



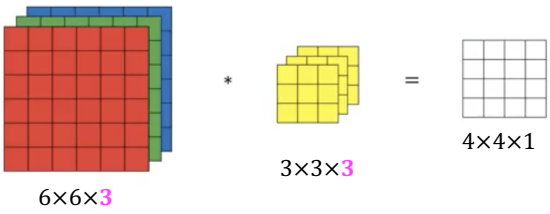
No padding, Stride 1 **Padding 1, Stride 1** **Padding 1, Stride 2**

3

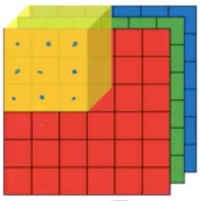
Imperial College London



If the Two-Dimensional image has 3 RGB channels…



27 operations each time (28 if bias term)



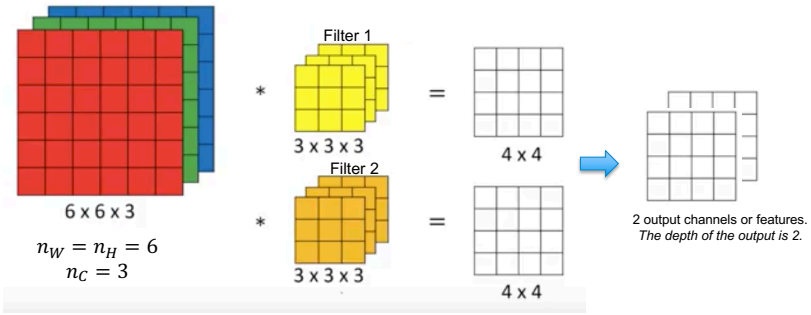
Each slice of the filter can detect different features in each of the 3 colour channels, for instance vertical edges in Red, horizontal edges in Green and vertical edges in Blue!

Andrew Ng https://www.youtube.com/watch?v=KTB_OFoAQcc&list=PLkDaE6sCZn6Gl29AoE31iwdVwSG-KnDzF&t=0s&index=7

4

Imperial College London

If Different Filters are Applied …



Exercise: *if my input Image is 50x50 with 10 channels, and if I apply 25 filters each of size 3x3x10 with a stride of 1, no padding, what is the size and number of channels of the output image?*

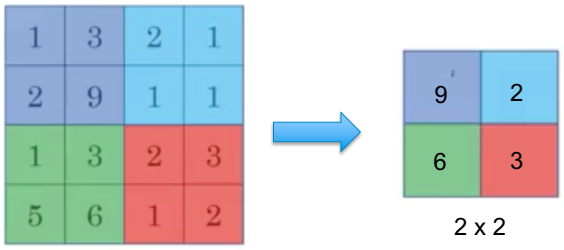
Answer: *size is 48x48. with 25 channels*

Andrew Ng https://www.youtube.com/watch?v=KTB_OFoAQcc&list=PLkDaE6sCZn6Gl29AoE31iwdVwSG-KnDzF&t=0s&index=7

5

Imperial College London

Example of Max Pooling Layer (for Downsampling)



Stride s and Filter size f are the main hyperparameters for pooling.

Here $s = 2$ and $f = 2$ Usually $f = s$

Pooling makes the input representations (feature dimension) smaller and more manageable for the next layer, Max Pooling is used because it may be interesting to keep the high values for the activation of the next layer as they may characterize some important features. Pooling reduces the number of parameters and computations in the network, therefore controlling overfitting.

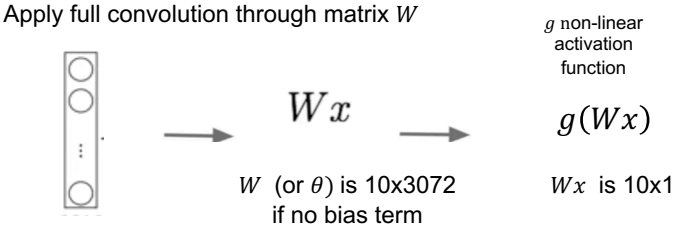
6

Imperial College London

We already know Fully Connected Layers!

Transform for example 32x32x3 image into 10x1 vector.

1. Reshape image into 1D vector x of dimension 32x32x3= 3072
2. Apply full convolution through matrix W



x is 3072x1

W (or θ) is 10x3072 if no bias term

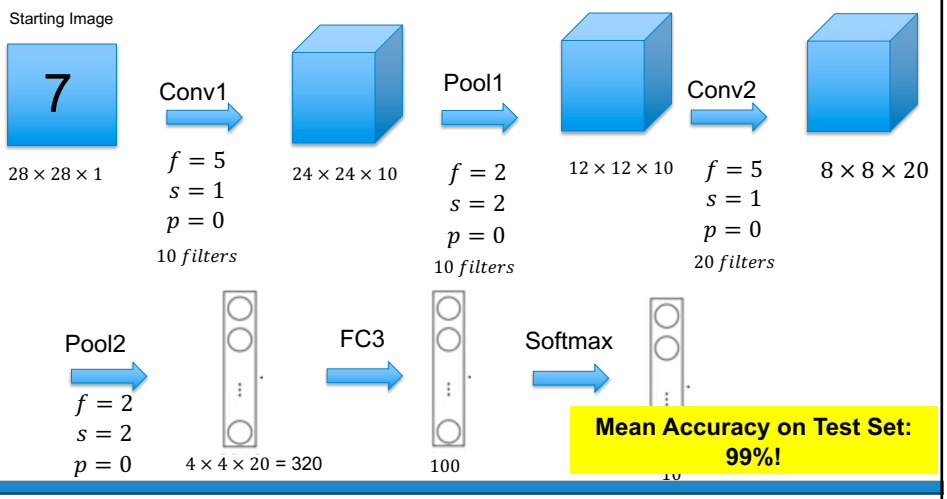
Wx is 10x1

Fully connected layers often have many weights as they connect every neuron in input layer to every neuron in output layer.

7

Imperial College London

MNIST CNN Example (The LeNet-5 Architecture from LeCun)



Mean Accuracy on Test Set: 99%!

8

Imperial College London

MNIST: Summary of CNN Layers (Convolutions with Bias Terms

	Size of input image n	Number of input channels	f	p	s	Size of output image (n+2p-f)/s+1	Number of output channels or filters	Number of output neurons	Size of Filter + 1	Number of Parameters
Conv1	28	1	5	0	1	24	10	5760	26	260
Pool1	24	10	2	0	2	12	10	1440		
Conv2	12	10	5	0	1	8	20	1280	251	5020
Pool2	8	20	2	0	2	4	20	320		
	Size of input							Number of output neurons		
FC1	320							100	321	32100
Softmax	100							10	101	1010
							Total Neurons	8910	Total Parameters	38390

Formula used for last column:
Number of Parameters= Number of output channels × (Size of filter+1) (as we assume there is a bias term)

9

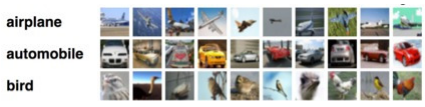
Imperial College London

Some of the Best-Known Datasets

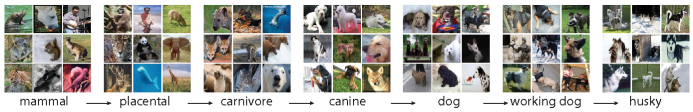
Very deep neural networks work best when trained on very large datasets!

- **MNIST: Handwritten digits, 60000 Training Images, 10000 Test Images**
- **CIFAR-10 / CIFAR-100: 50k Training, 10k Test Images of 10 (CIFAR-10) or 100 (CIFAR-100) classes**

  Color Images are 32x32, Task: Classification https://www.cs.toronto.edu/~kriz/cifar.html



- **Imagenet: > 15 Million Images in 20,000 classes! https://en.wikipedia.org/wiki/ImageNet**

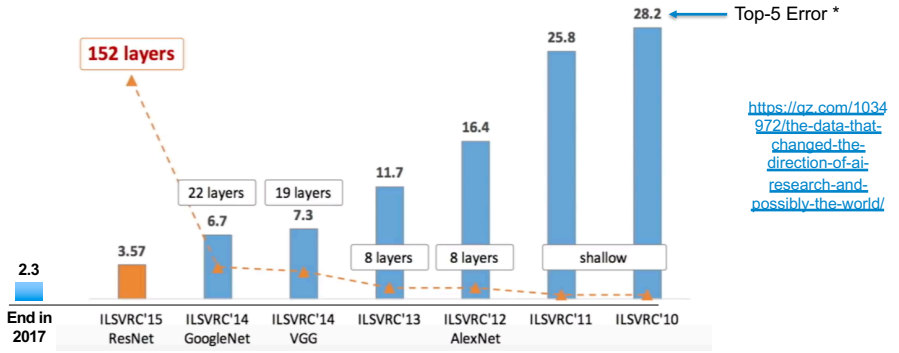


10

Imperial College London

The Image Large-Scale Visual Recognition Challenge

From 2010 to 2017, ImageNet has been running an annual competition (ILSVRC: Image Large Scale Visual Recognition Challenge) in visual recognition where participants are presented with the challenge of classifying 1.2 million images belonging to 1000 different classes from Imagenet dataset



https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/

** The **Top-5 error** is the percentage of the time that the classifier did not include the correct class among its **top 5** guesses.*

https://www.youtube.com/watch?v=DAOcjicFr1Y

11

Imperial College London

Transfer Learning (1)

The most well-known CNN designs are available on-line and have been successfully trained on very large number of images (1,000,000s).

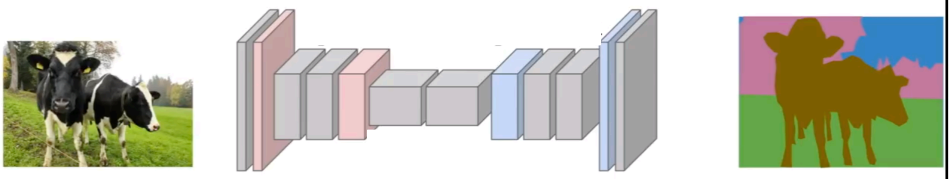
In many applications we often work from a relatively small number of images.

Why not start from an existing trained CNN sharing an objective of a similar nature (such as Classification) and tailor it to our application?

12

Imperial College London

Semantic Segmentation with Down- and Up-Sampling



Input:
3 x H x W

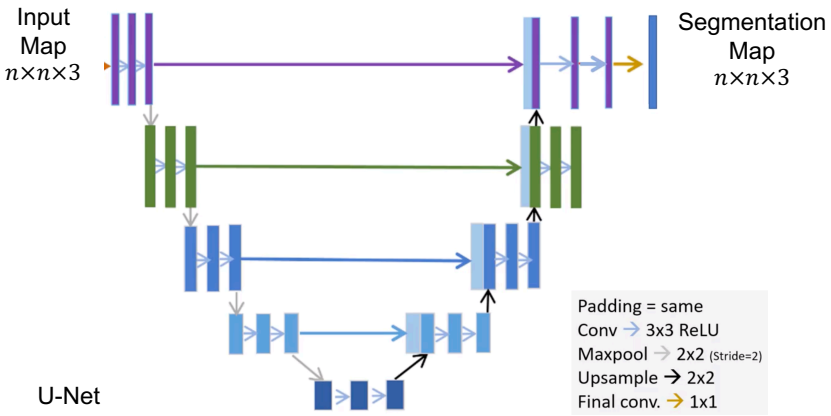
Predictions:
H x W

https://www.youtube.com/watch?v=azM57JuQpQI

13

Imperial College London

Segmentation Model: U-Net (Ronneberger et al, 2015)



U-Net

https://www.youtube.com/watch?v=azM57JuQpQI

14