أكاديمية كاوست
KAUST ACADEMY

Advanced Artificial Intelligence: Computer Vision

May 6, 2023

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

Course Website

أكاديمية كاوست
KAUST ACADEMY

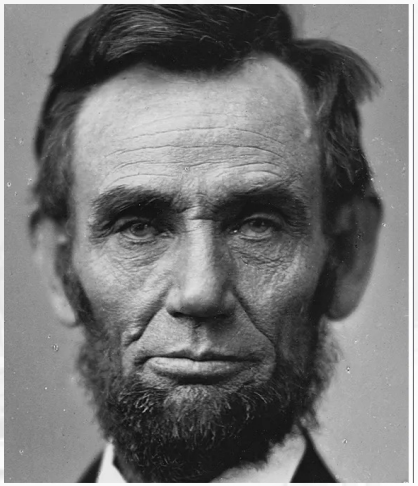
https://naeemullah-khan.github.io/AAI23

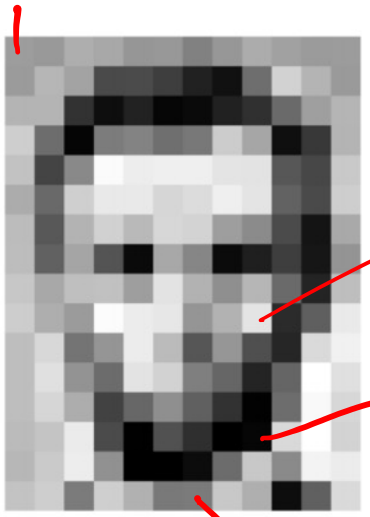
Convolutional Neural Networks

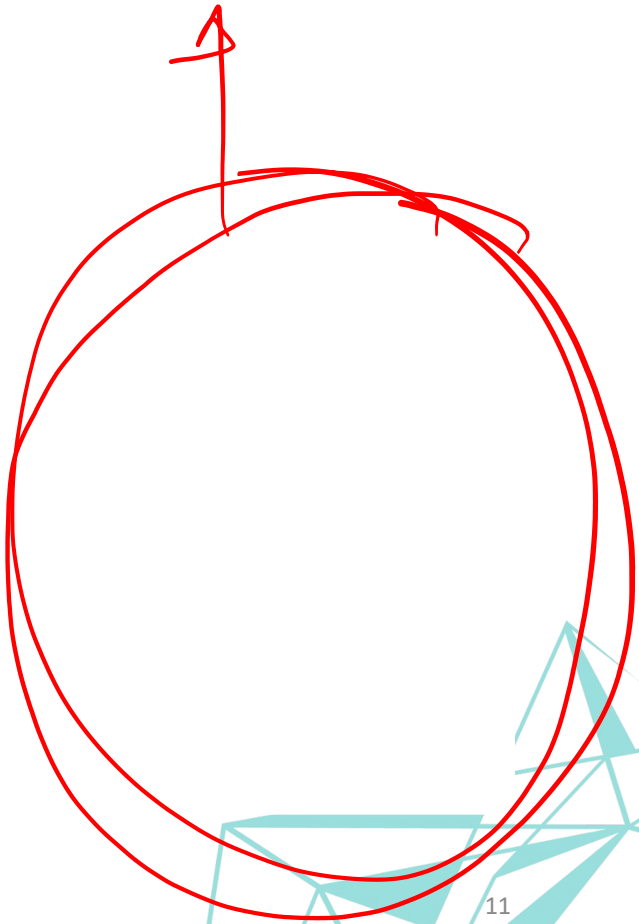
جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

How to represent an image?

- Matrix with elements in [0,255]





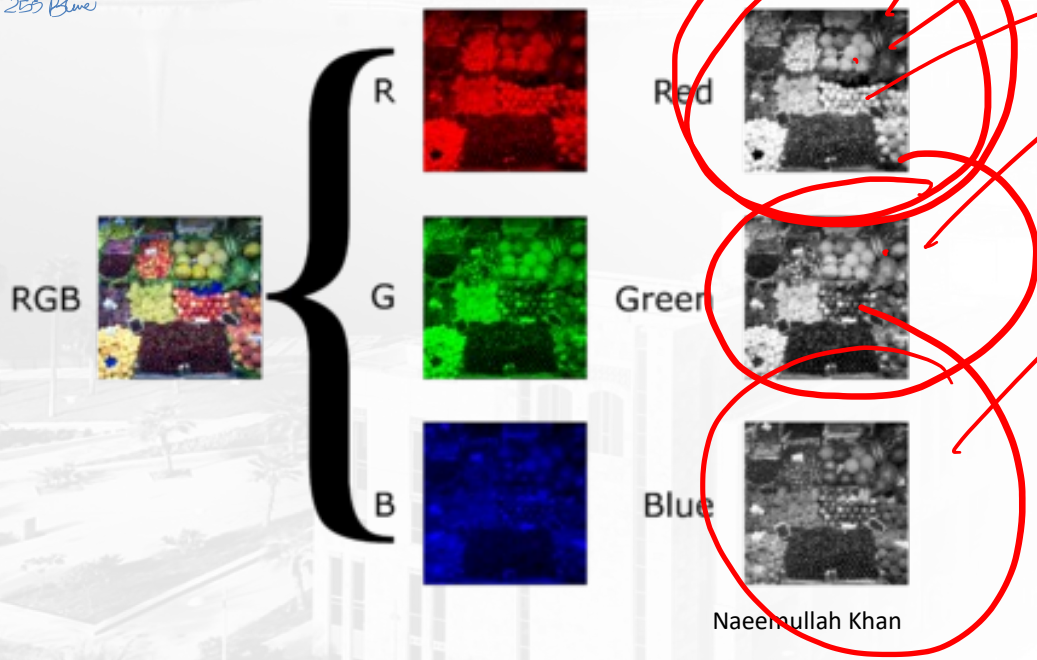


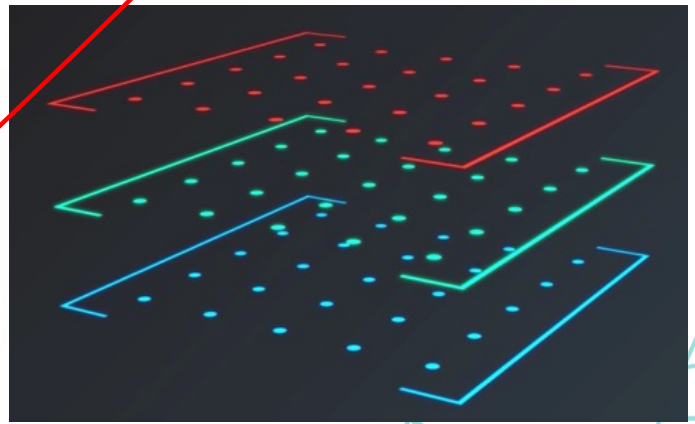
How to represent an image?

- Matrix with elements in [0,255]
- Tensor with RGB channels, each being a matrix

How to represent an image?

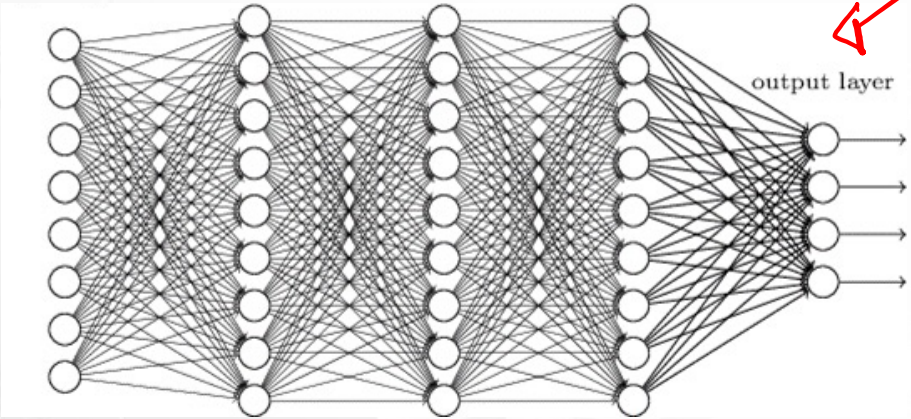
- Grayscale images are matrices
- Color images are Tensors





Convolutional Neural Network (CNN)

Full-connected network



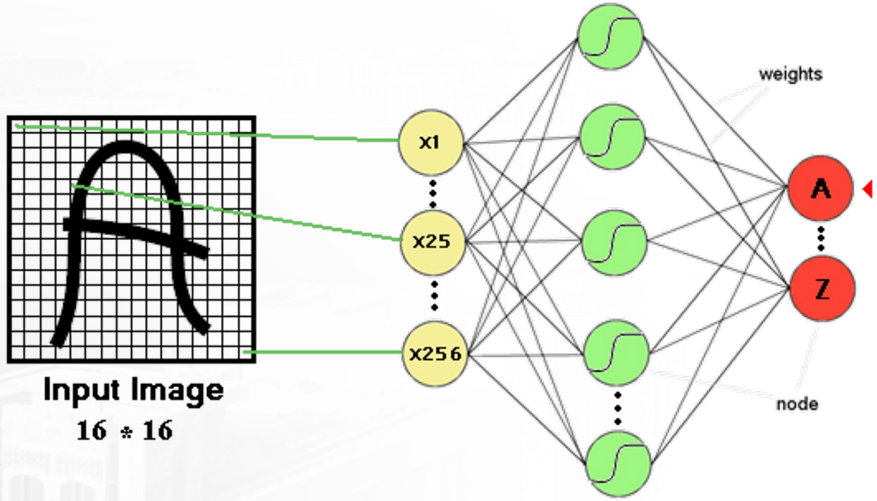
$$z = W_1x_1 + W_2x_2 + \cdots + W_nx_n + b$$

Convolutional network

$$z_{ij} = W * x_{i,j} = \sum_{a=0} \sum_{b=0} W_{ab}x_{(i+a)(j+b)}$$

Multi-layer perceptron and image processing

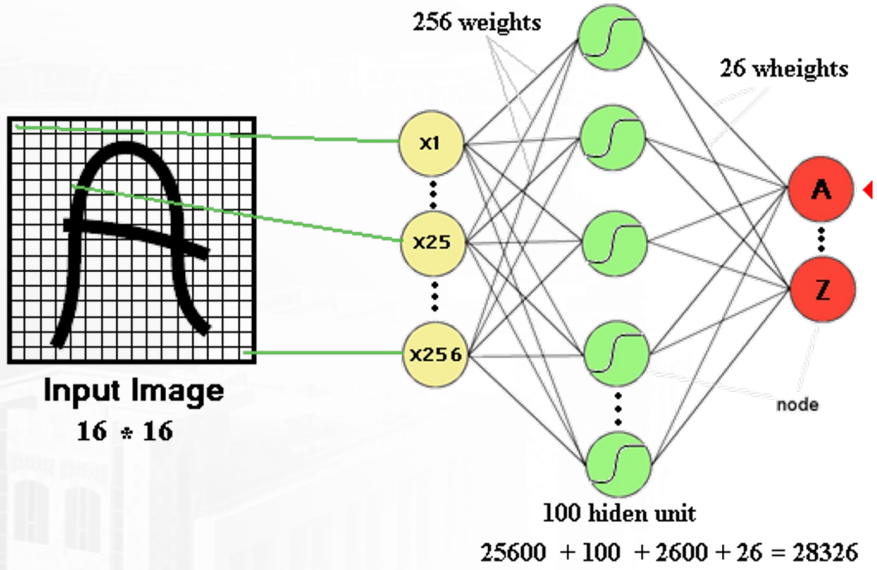
- One or more hidden layers
- Sigmoid activations functions



Multi-layer perecptron is another name for Neural Networks

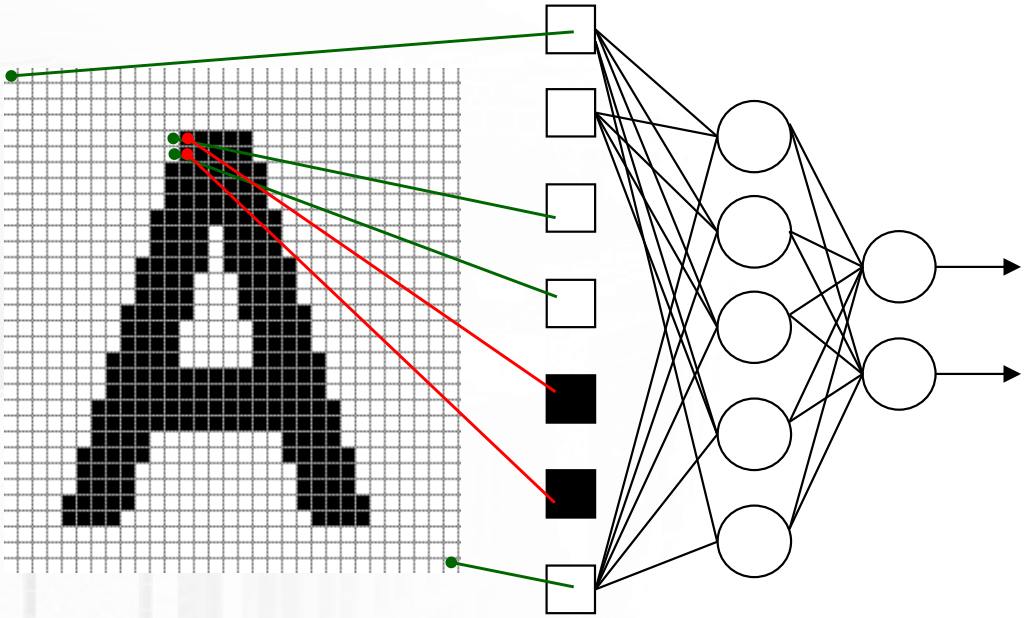
Drawbacks of previous neural networks

- the number of trainable parameters becomes extremely large



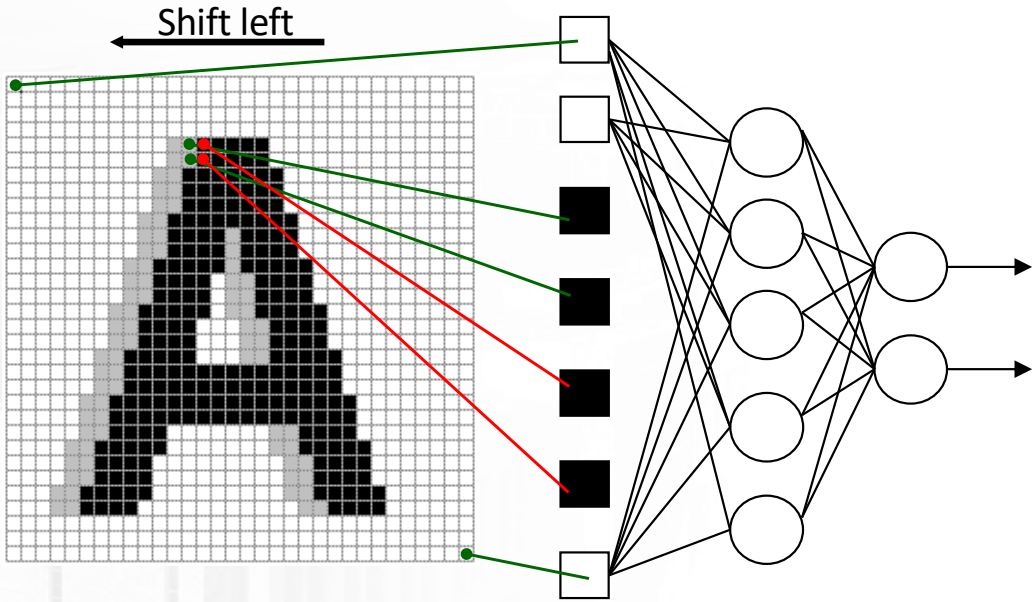
Drawbacks of previous neural networks

- Little or no invariance to shifting, scaling, and other forms of distortion



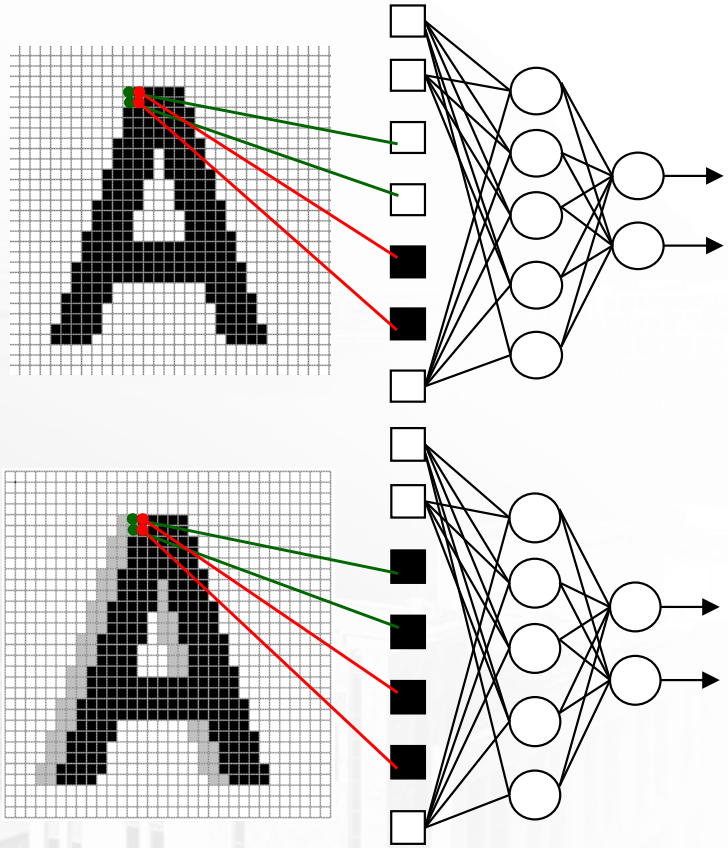
Drawbacks of previous neural networks

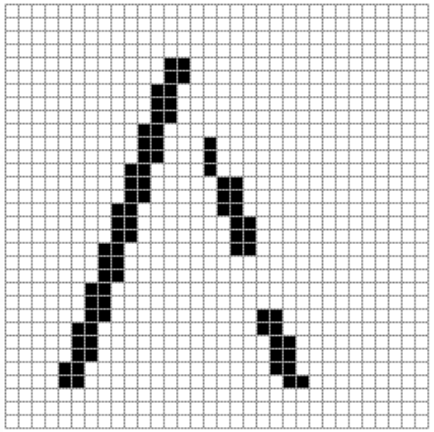
- Little or no invariance to shifting, scaling, and other forms of distortion



Drawbacks of previous neural networks

* topology : where does the data come from
- NN does not care for topology but we want it to

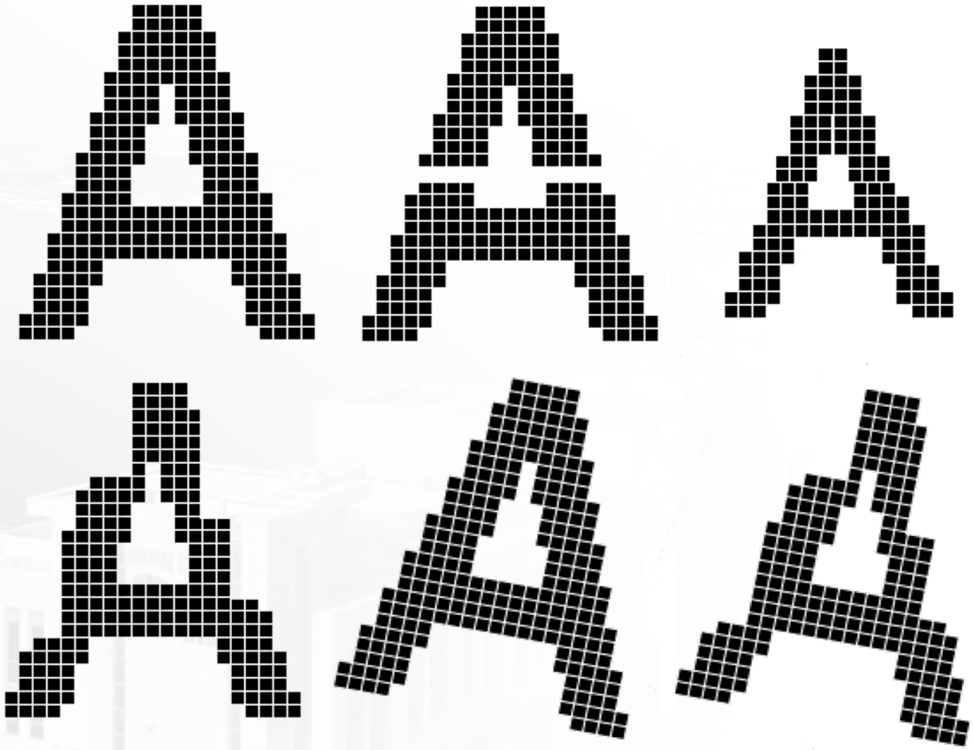




154 input change
from 2 shift left → true : if I did this the output should not change & NN should know this
77 : black to white
77 : white to black

Drawbacks of previous neural networks

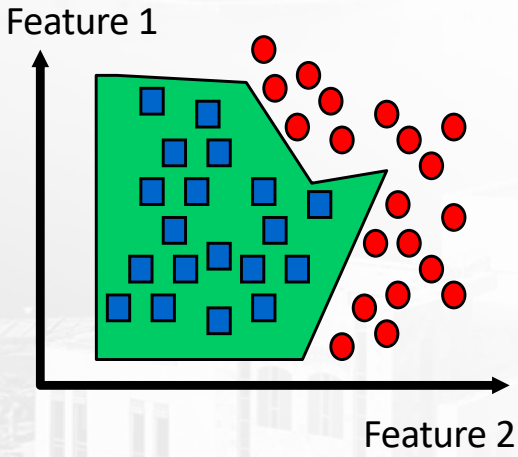
- scaling, and other forms of distortion

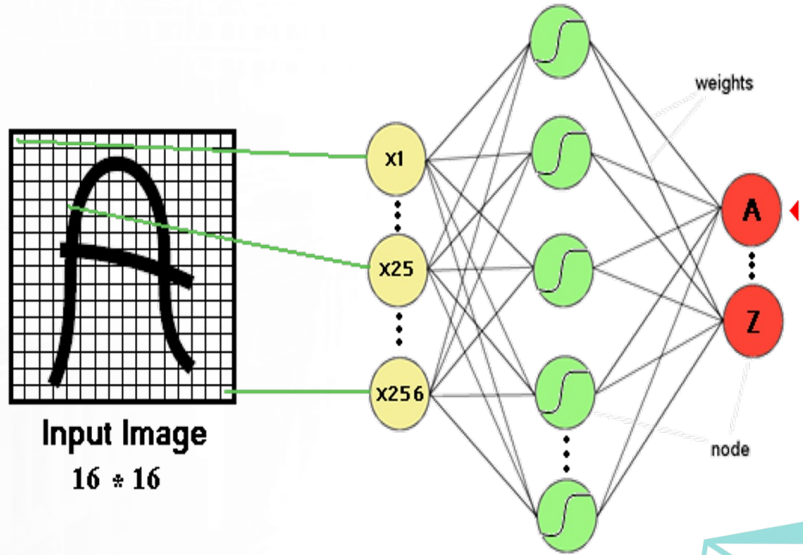


Drawbacks of previous neural networks

- the topology of the input data is completely ignored
- work with raw data.

If we look at data Pixel wise we would never recognize a thing. It's the pattern that it counts





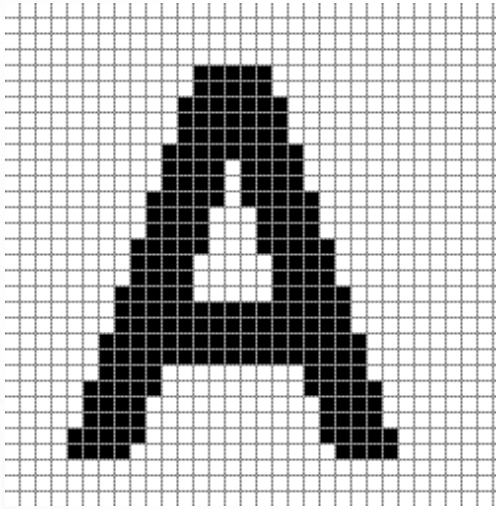
Drawbacks of previous neural networks

Black and white patterns:

$$2^{32*32} = 2^{1024}$$

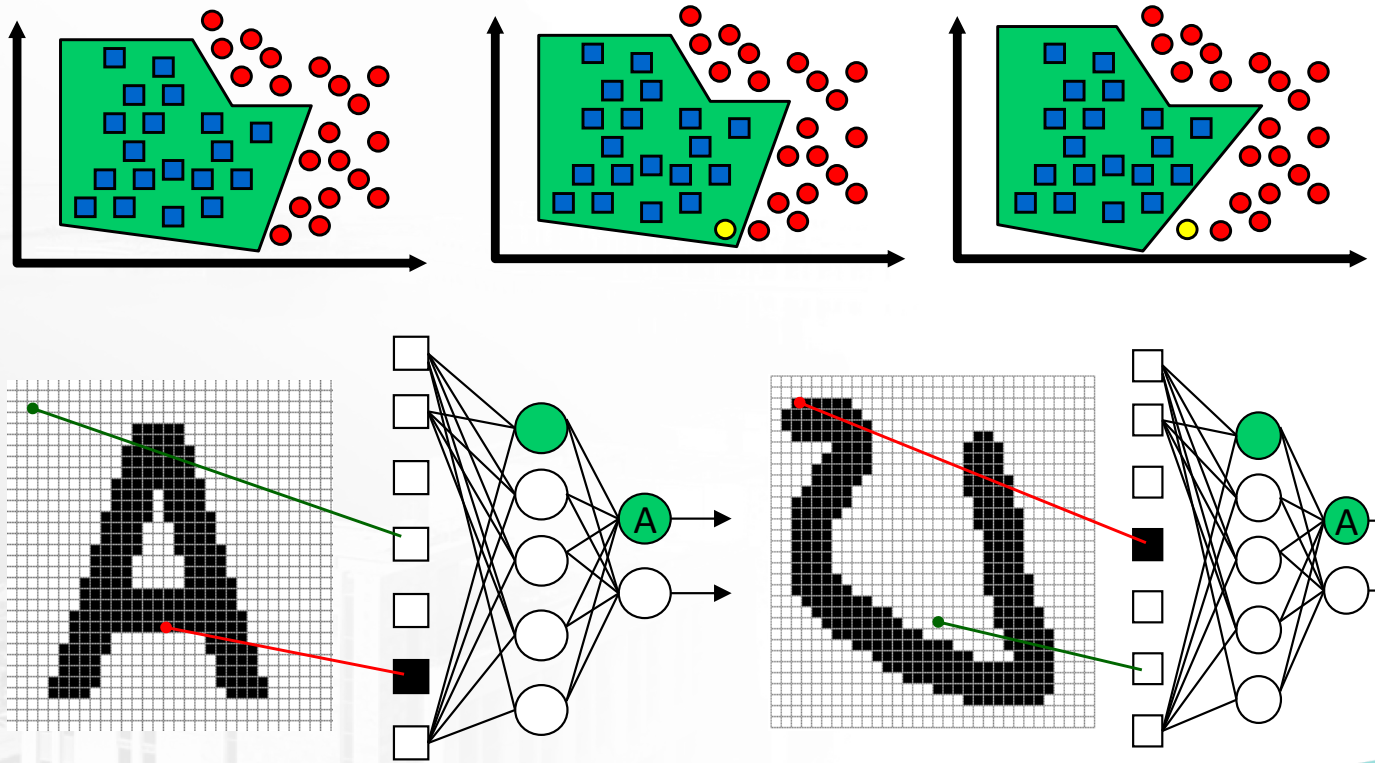
$$256^{32*32} = 256^{1024}$$

Gray scale patterns :



32 * 32 input image

Drawbacks of previous neural networks



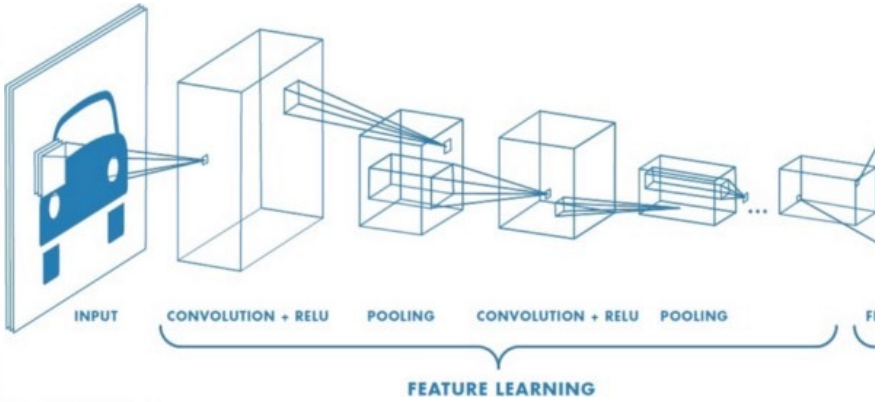
Improvement

- Fully connected network of sufficient size can produce outputs that are invariant with respect to such variations.
  - **Training time**
  - **Network size**
  - **Free parameters**

Convolutional Neural Network (CNN)

- Local connectivity
- Reuse parameters

Convolutional network



$$z_{ij} = W * x_{i,j} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} W_{ab} x_{(i+a)(j+b)}$$

Convolutional Neural Network

- How Convolution works
- Filters / Padding / Stride
- Activations
- Pooling Layers

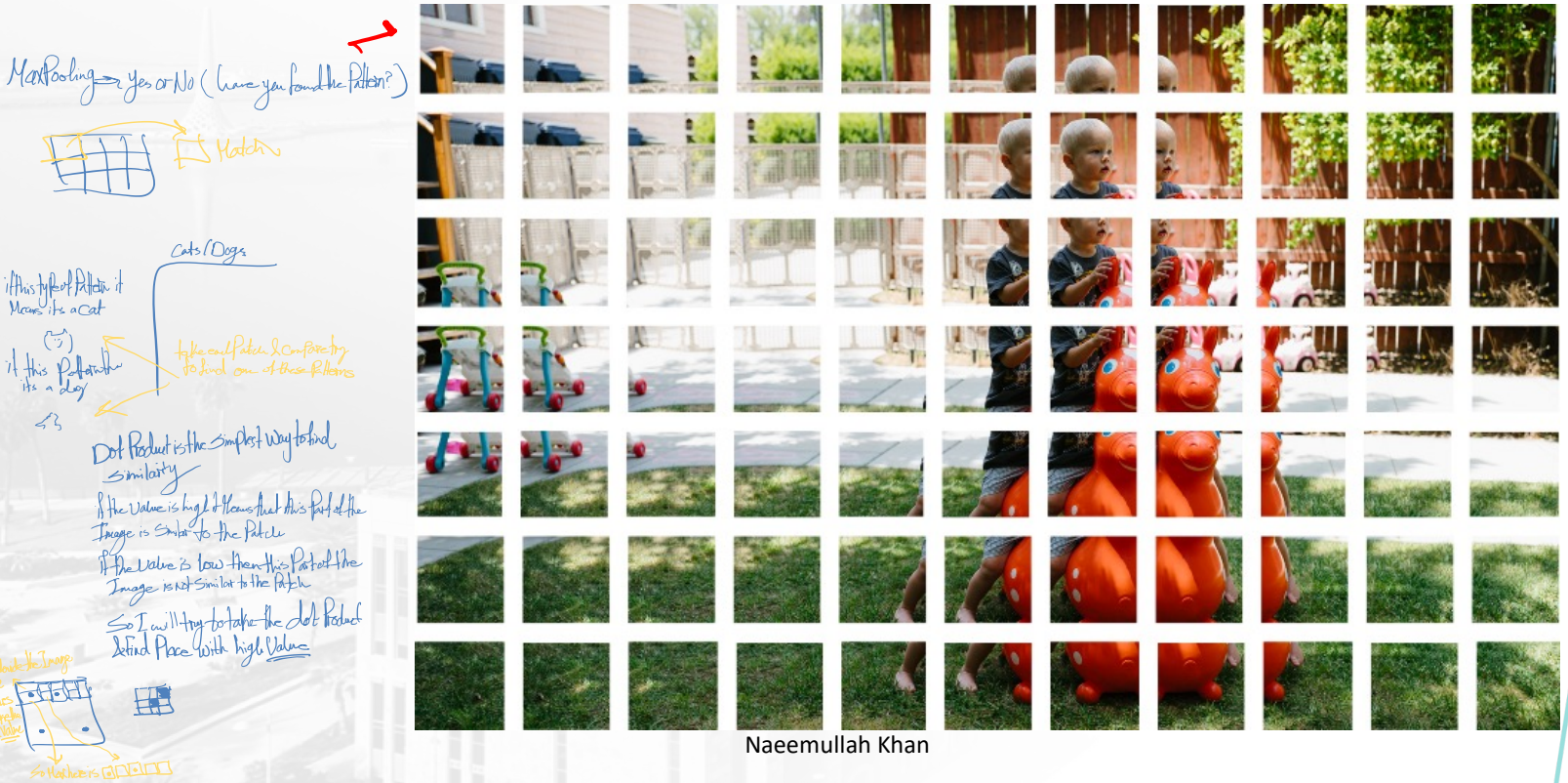
How Convolution works

- Convolution on an image

Naeemullah Khan

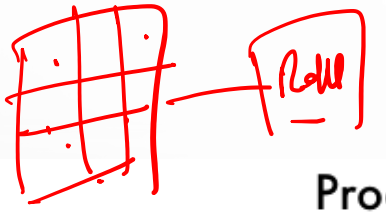
How Convolution works

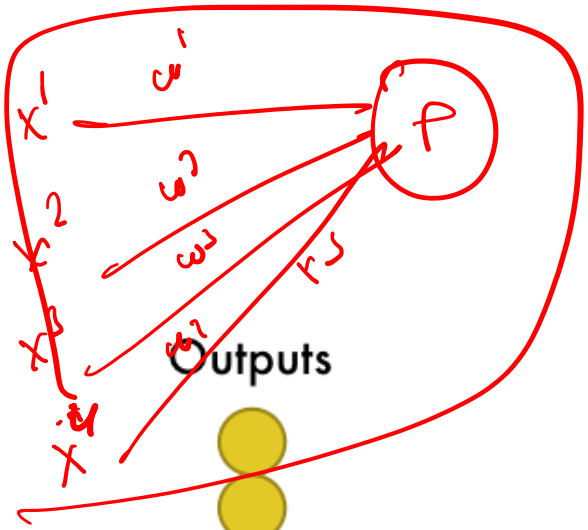
- Break the image in overlapping tiles



How Convolution works

- Feed each image tile into small NN



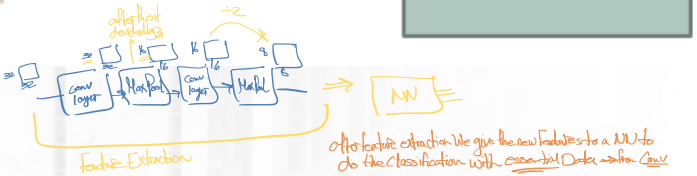


Processing a single tile

Input Tile

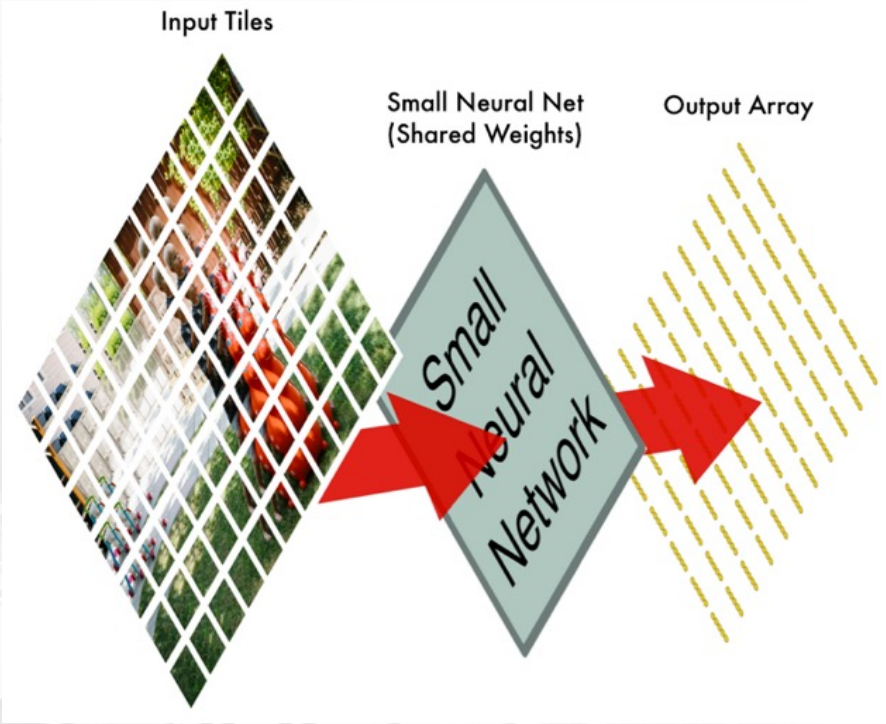
Small Neural Network

Outputs



How Convolution works

- Store results from each tile into a new array



$$z_{ij} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} W_{ab} x_{(i+a)(j+b)}$$

Filter/kernel

CNN – Filters

Basically How Conv changes the Image:

if filter is fs x fs you will lose $\frac{fs-1}{2}$ from the boundry

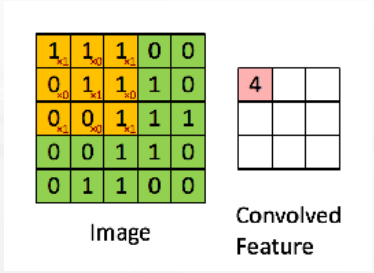
if fs 3x3 $\Rightarrow \frac{3-1}{2} = 1$, 5x5 $\Rightarrow \frac{5-1}{2} = 2$

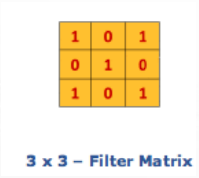
here we will lose 2 row, column from the boundary of the Image

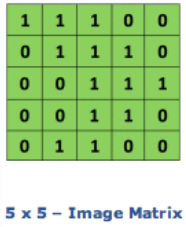
$$z_{ij} = W * x_{i,j}$$

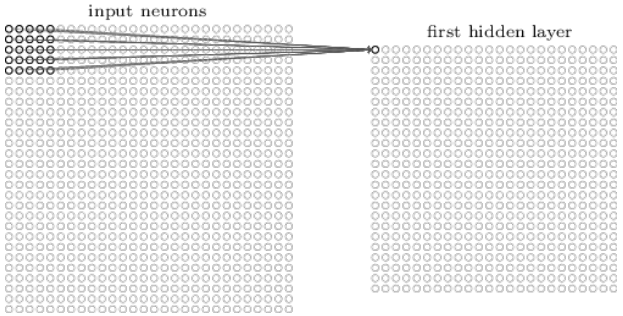
- Filters and convolution

* You can add Padding if you want to keep the output size = your Image







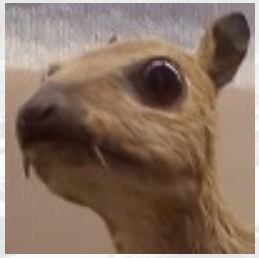


highlighting the more enhanced Parts of the Img

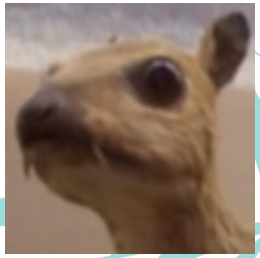
Smoothing

Multiple filters

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

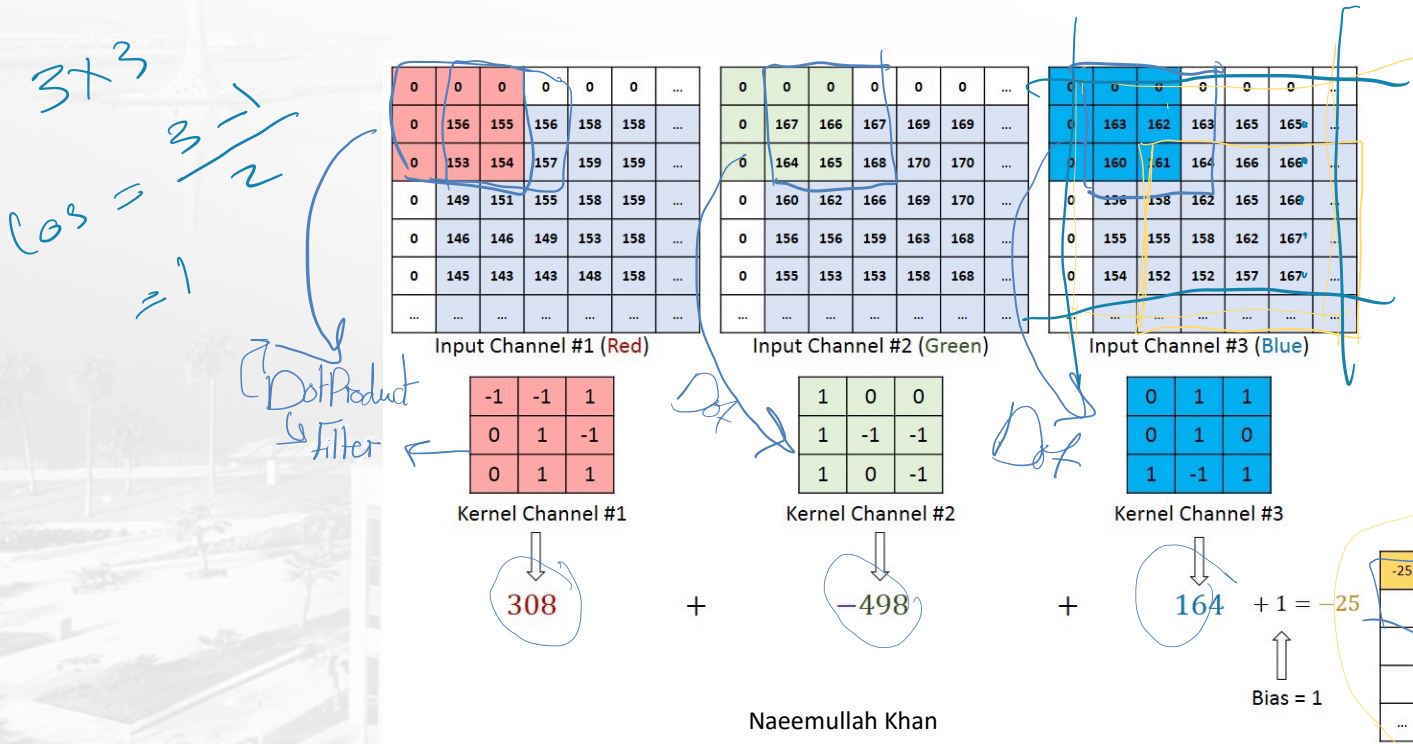




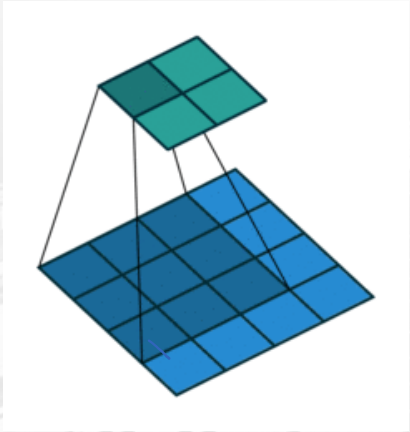
CNN – Filters

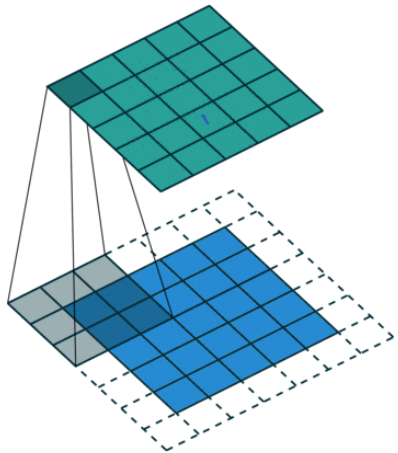
- Multiple channels
  - Ex: RGB color



CNN - Padding

- What happen to the borders?
  - Reduce size of the output
  - Pad the picture with zeros (zero-padding)

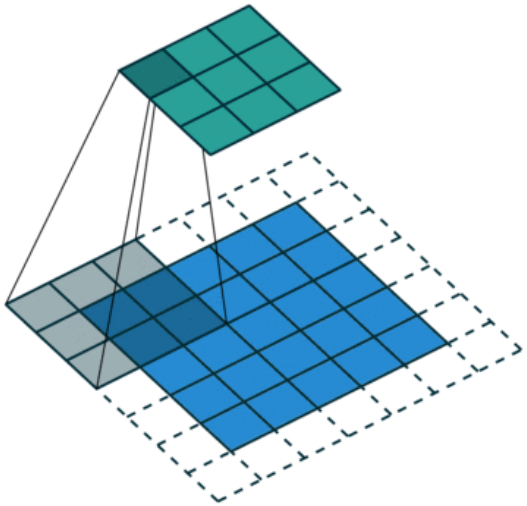




More on: http://deeplearning.net/software/theano/tutorial/conv_arithmetic.html

CNN – Strides

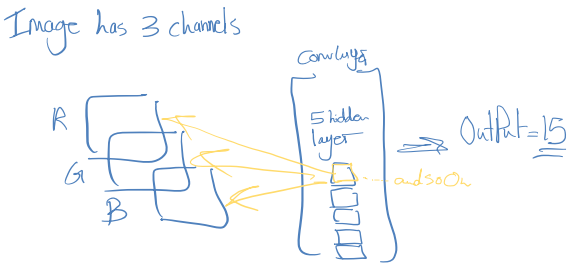
- If not need to keep all output
  - Skip some pixels
  - Stride: number of skipped pixels
  - Output with smaller resolution



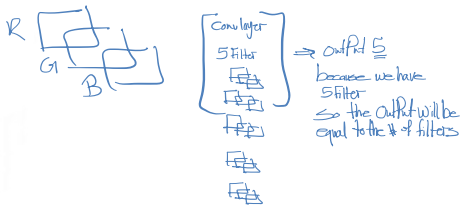
More on: http://deeplearning.net/software/theano/tutorial/conv_arithmetic.html

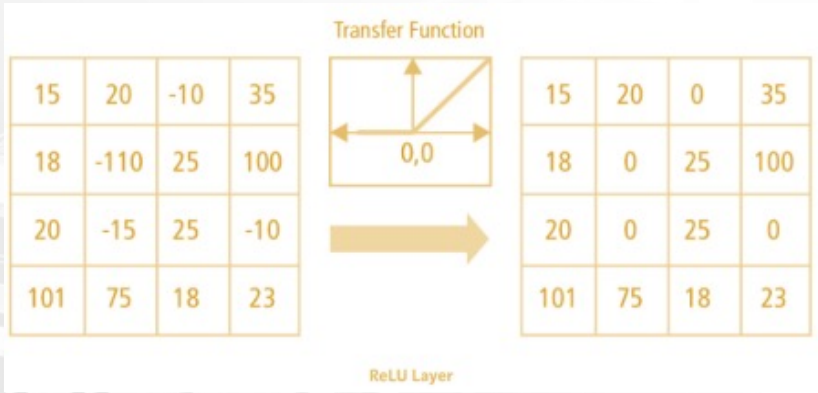
CNN – Activations

- Any previous activation fits
- Breaks linearity
- Ex: Rectified Linear Unit (ReLU)



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



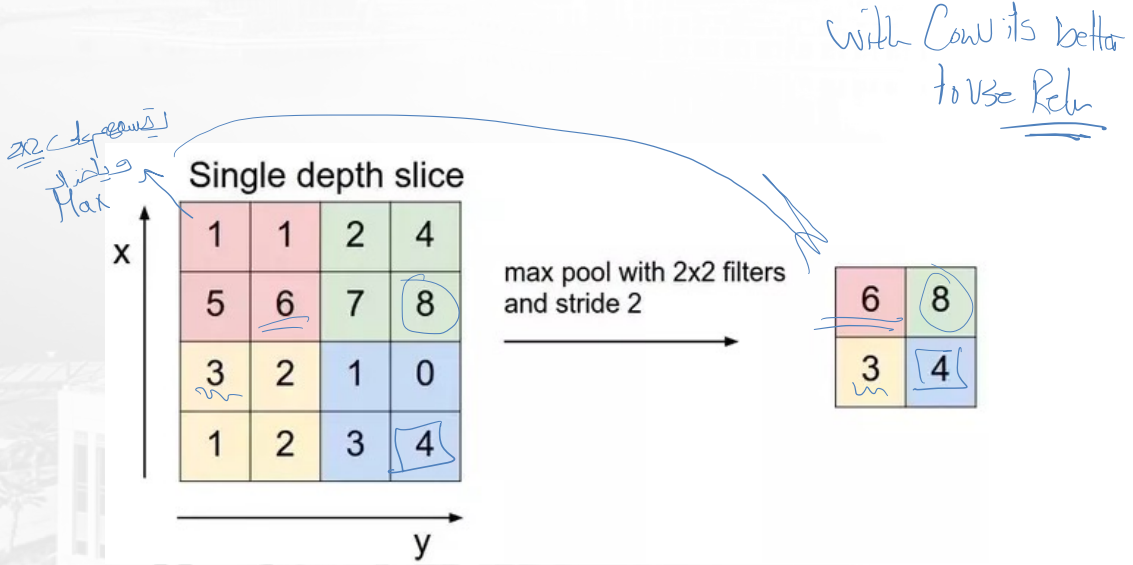


15	20	-10	35
18	-110	25	100
20	-15	25	-10
101	75	18	23

15	20	0	35
18	0	25	100
20	0	25	0
101	75	18	23

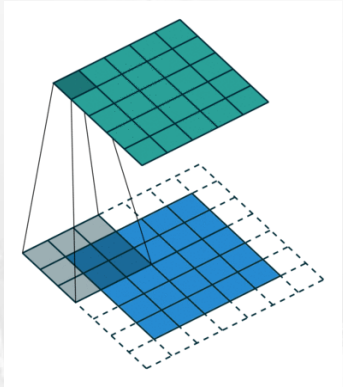
CNN – Pooling Layer

- Max / Mean / Sum Pooling
- Reduce the resolution



A convolution layer

Convolution

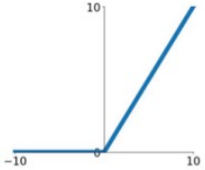


Filter
Padding
Stride

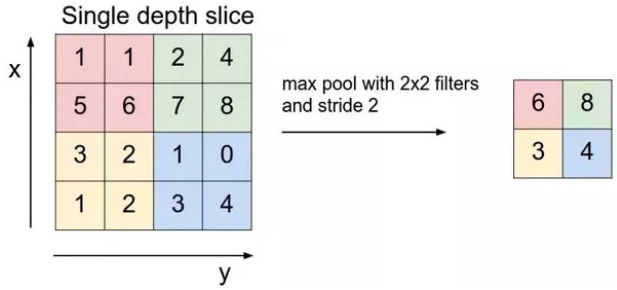
Activation

ReLU

$\max(0, x)$

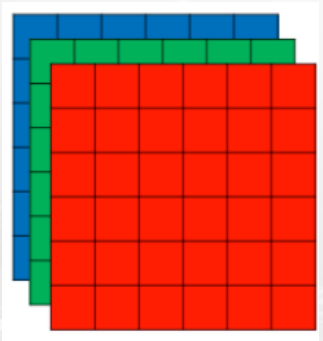


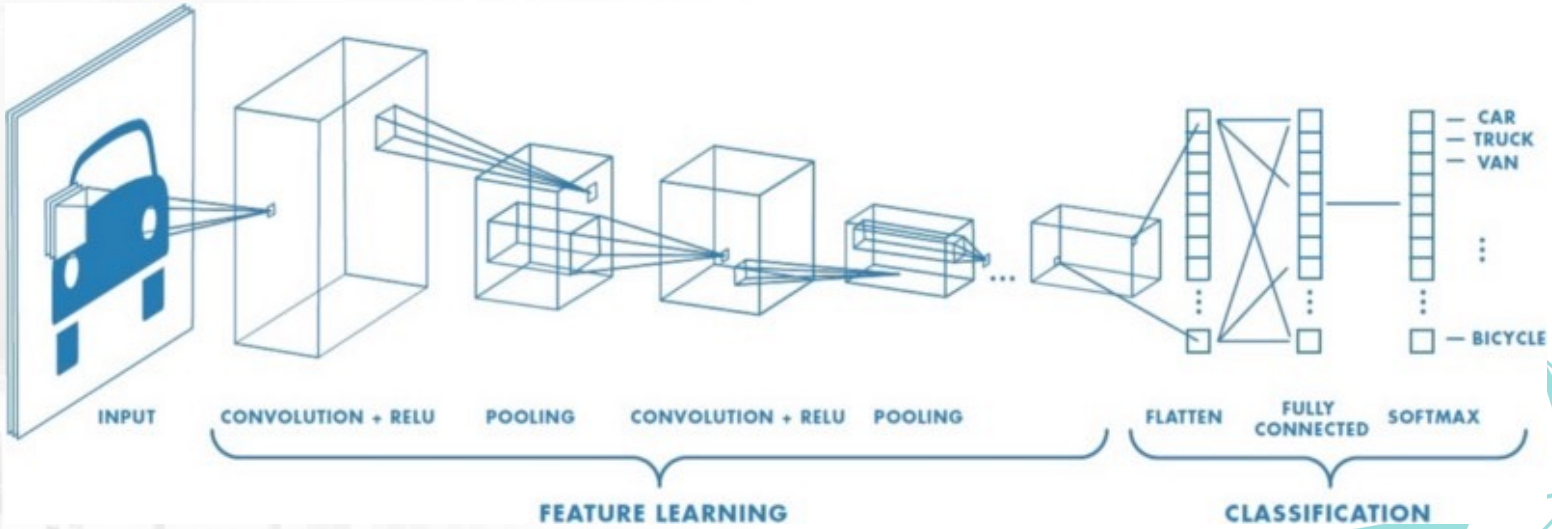
Pooling



Deep Convolutional network

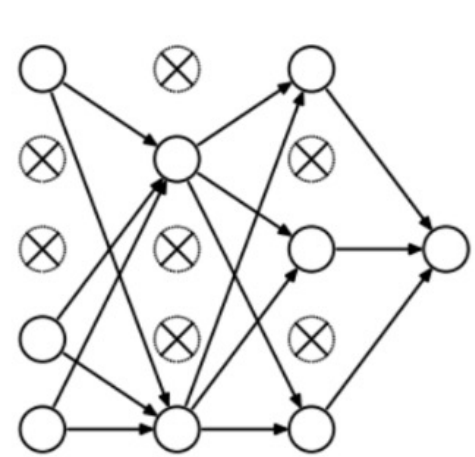
- Multiple convolutional layers



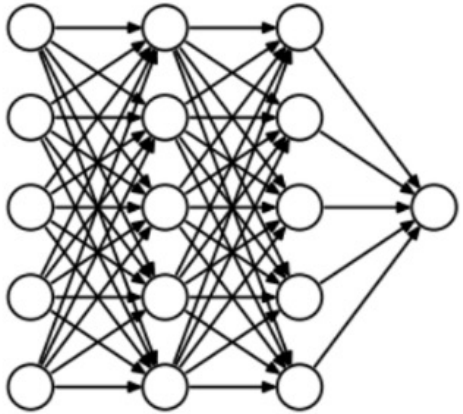


Deep Convolutional network

- Add redundancy to the network
- A way of regularization



Training

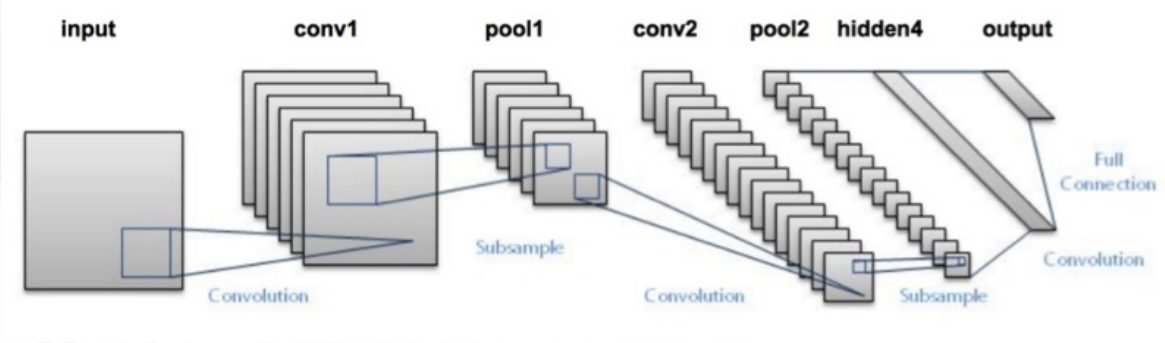


Testing

CNN – Most notable network

- LeNet-5 (1998)
  - 2 convolutional layers, 2 pooling layers,2 fully connected layers
  - Used for handwritten numbers on checks

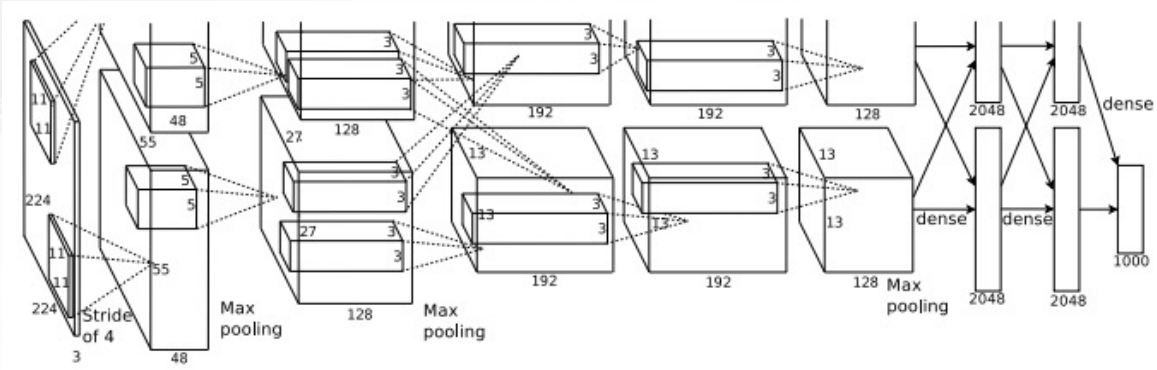
[PDF] LeNet - Yann LeCun
yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf
by Y LeCun - 1998 - Cited by 17209 - Related articles
OF THE IEEE, NOVEMBER 1998. 1 ... Real-life document recognition systems are composed .... C onvolutional Neural Network c alled L eNet - 5 des c rib ed in.



CNN – Most notable network

[PDF]
ImageNet Classification with Deep Convolutional ... - NIPS Proceedings
https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neu...
by A Krizhevsky - 2012 - Cited by 36889 - Related articles
The specific contributions of this **paper** are as follows: we trained one of the largest convolutional neural networks to date on the subsets of ImageNet used in the ...

- AlexNet (2012)
  - First big improvement in image classification
  - CNN / MaxPool / Dropout / ReLU
  - 5 convolutional layers, followed by max-pooling layers; the last three are fully connected layers

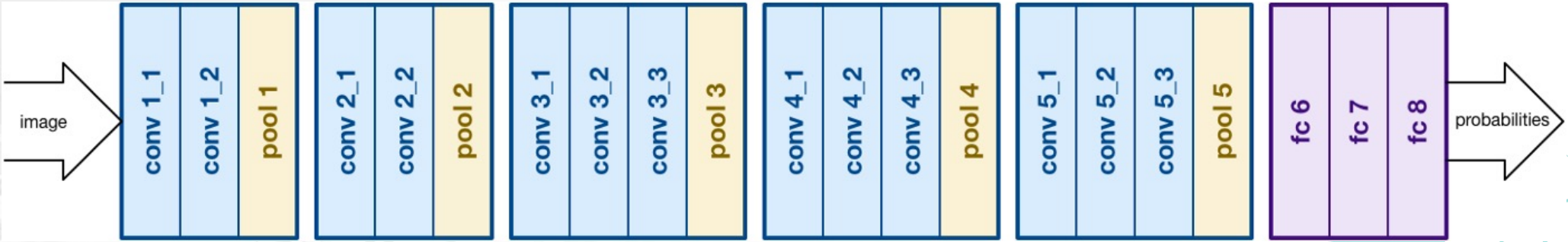


Naeemullah Khan

CNN – Most notable network

- VGGNet-16 (2014)
  - 13 convolutional layers, 3 fully-connected la

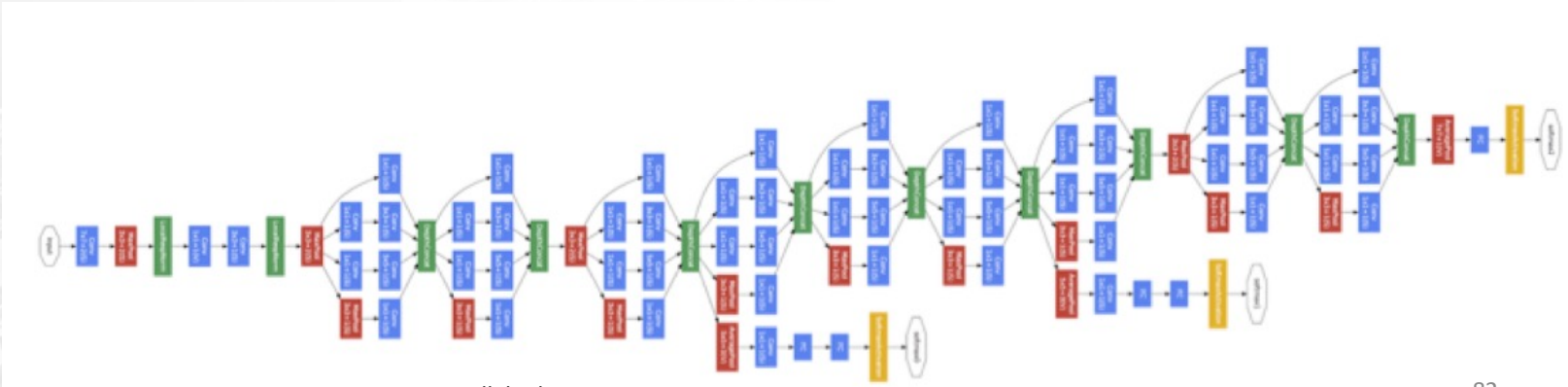
[PDF] Very Deep Convolutional Networks for Large-Scale Image Recognition
https://arxiv.org/pdf/1409.1556
by K Simonyan - 2014 - Cited by 20568 - Related articles
Apr 10, 2015 - CV] 10 Apr 2015. Published as a conference **paper** at ICLR 2015 ... Google DeepMind. 1http://www.robots.ox.ac.uk/~**vgg**/research/very_deep/. 1 ...



CNN – Most notable network

- GoogleNet/Inception (2014)
  - Going Deep: 22 layers
  - Very small convolutions
  - Drastically reduce the number of parameters
    - From 138M (VGG) to 4M parameters

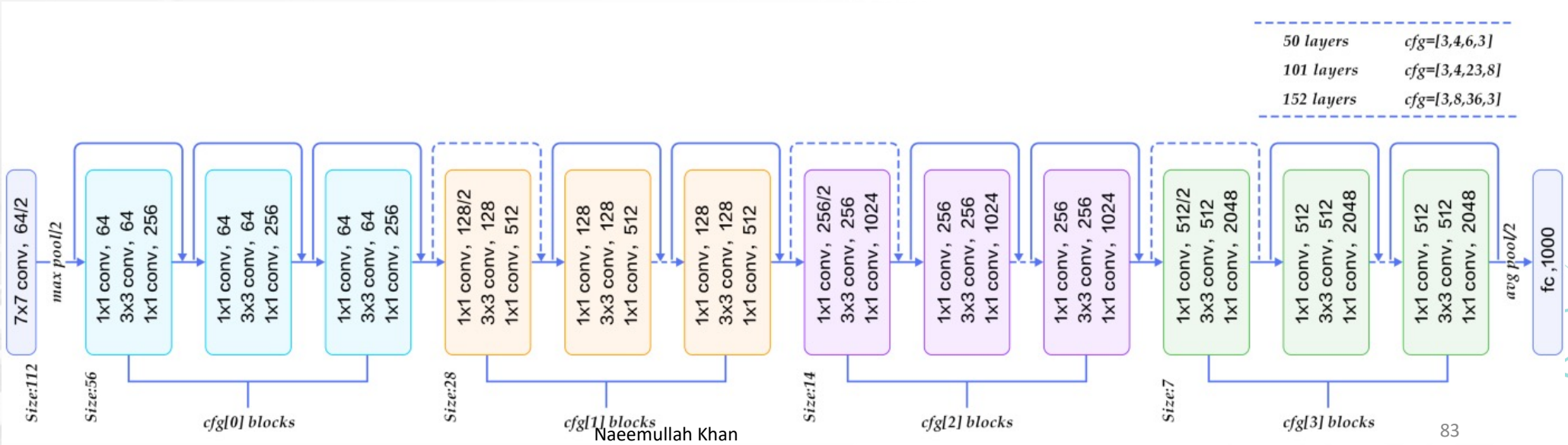
Going Deeper with Convolutions
https://arxiv.org › cs ▾
by C Szegedy - 2014 - Cited by 12157 - Related articles
Sep 17, 2014 - Abstract: We propose a deep convolutional neural network architecture codenamed "**Inception**", which was responsible for setting the new state ...



CNN – Most notable network

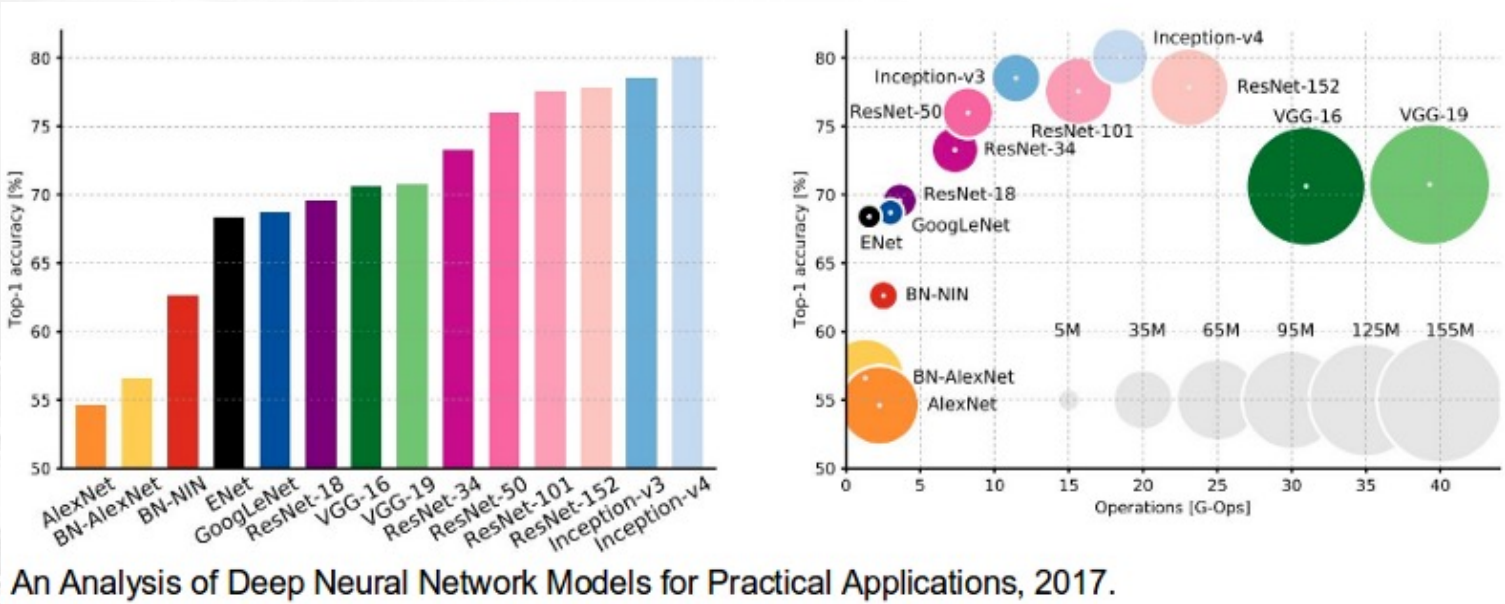
- ResNet (2015)
  - Residual layer
  - Up to 152 layers (50-101-152)
  - Requires huge resources

ResNet - arXiv
www.arxiv.org/abs/1512.03385
by K He - 2015 - Cited by 19344 - Related articles
No information is available for this page.
Learn why



CNN – Most notable network

- Totally more than 100 K citations

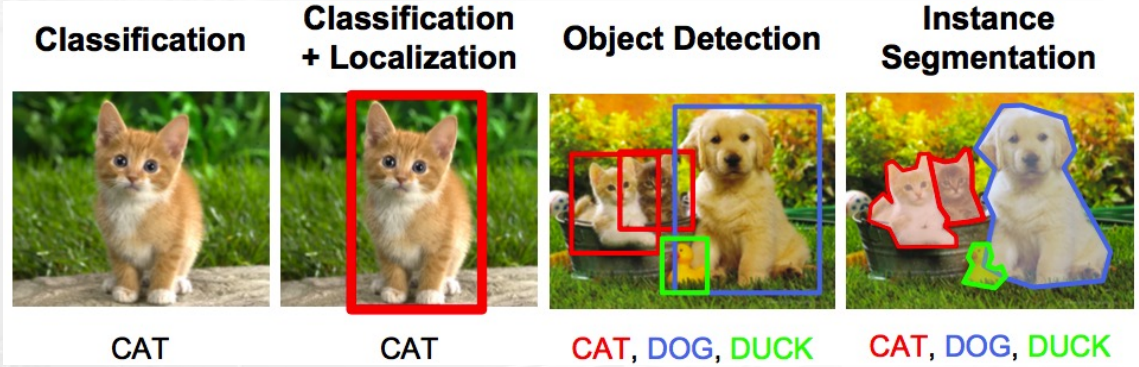


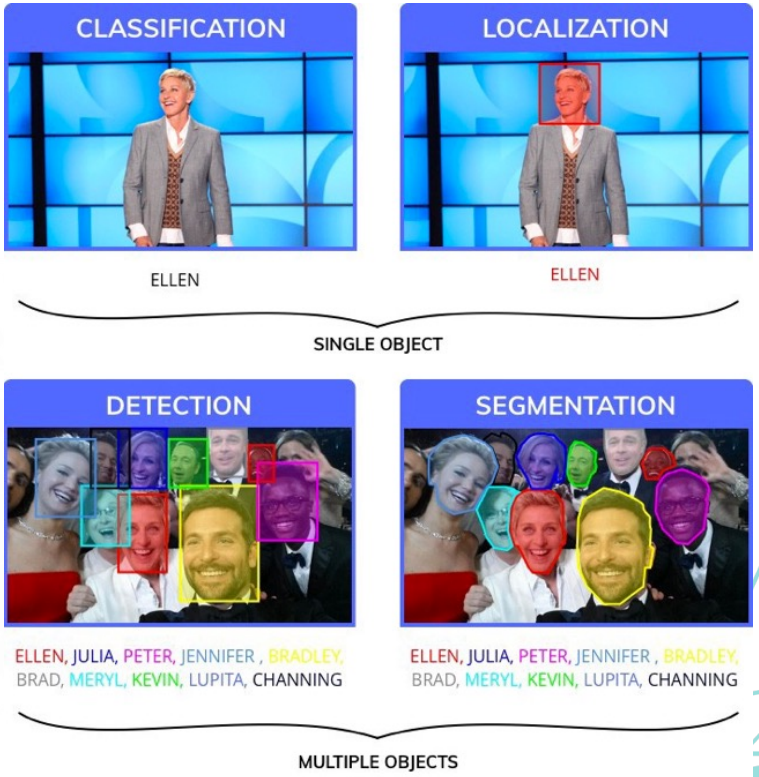
An Analysis of Deep Neural Network Models for Practical Applications, 2017.

Outline

- From machine learning to deep learning
- Deep neural network
- Train a deep neural network
- Convolutional neural networks
- Application on computer vision

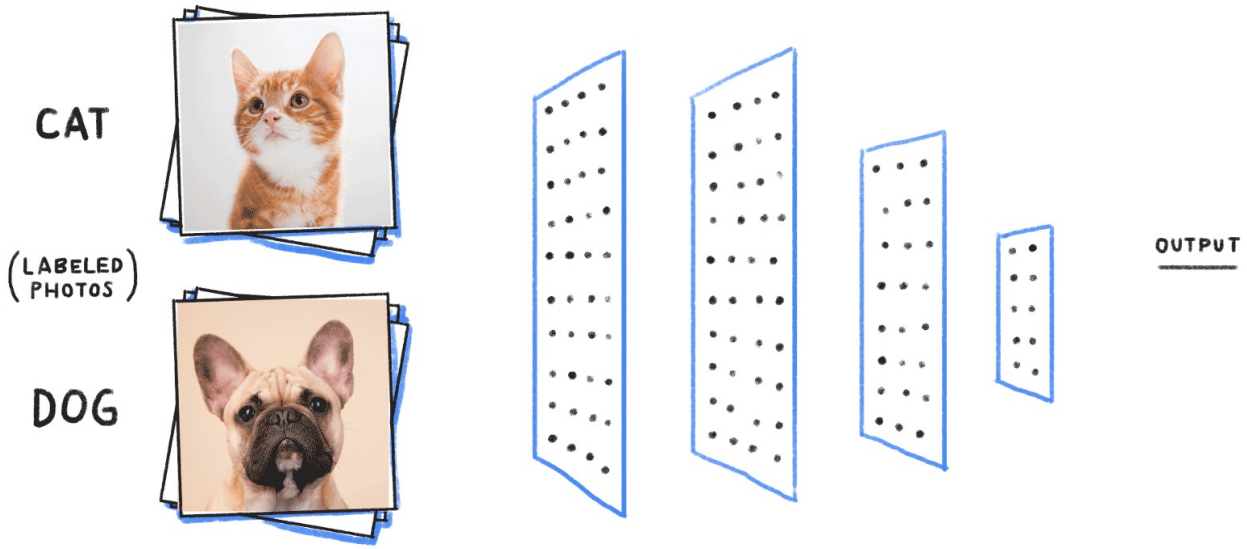
Application on Computer Vision





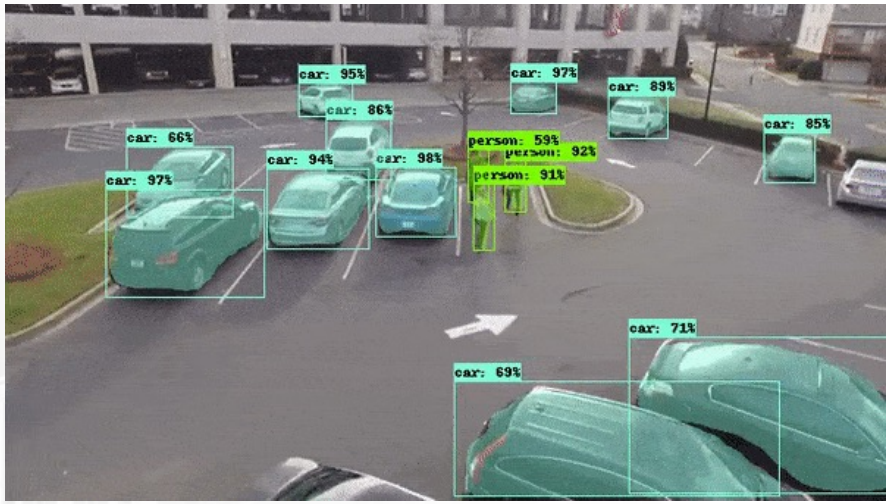
Application on Computer Vision

Classification



Application on Computer Vision

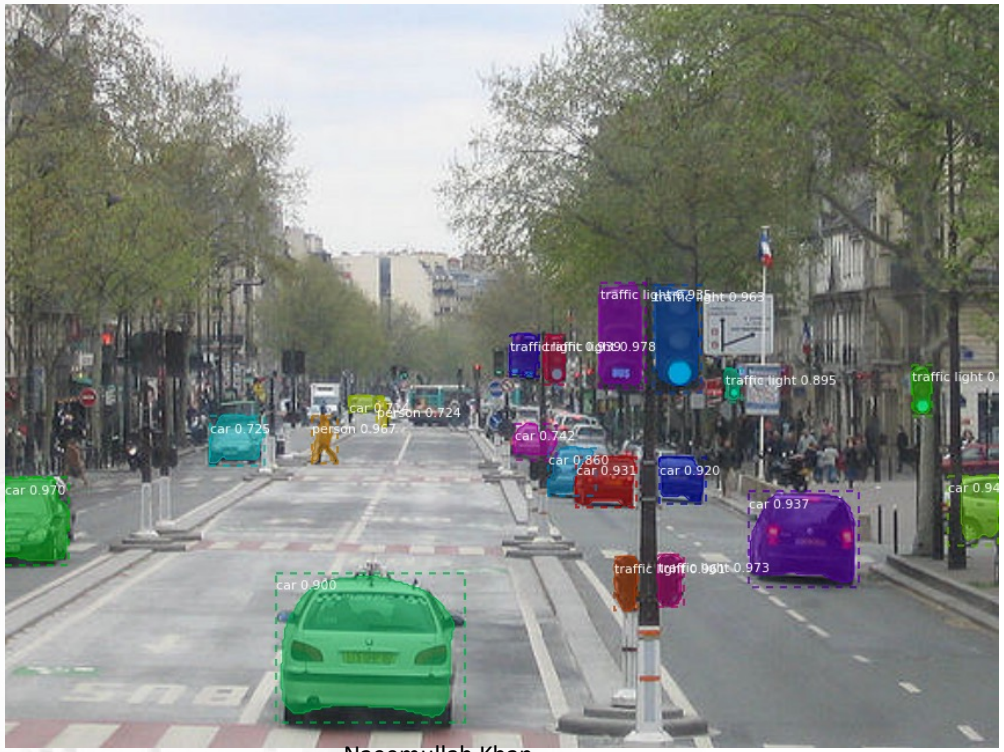
Object detection



Application on Computer Vision

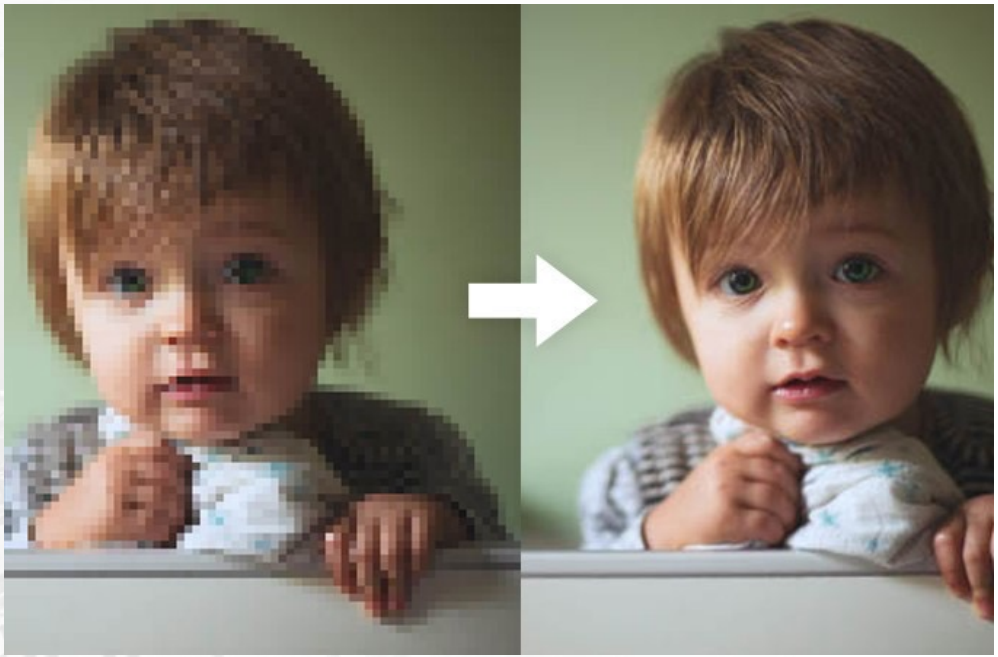
Semantic segmentation

Instance segmentation



Application on Computer Vision

Image super-resolution



Dataset and Challenges

- MNIST
  - Digit Classification (handwritten)
  - Grayscale intensity from 0 to 255
  - 28 x 28 = 768 pixels
  - 10 classes
  - 10,000 images

Classification

CAT

Dataset and Challenges

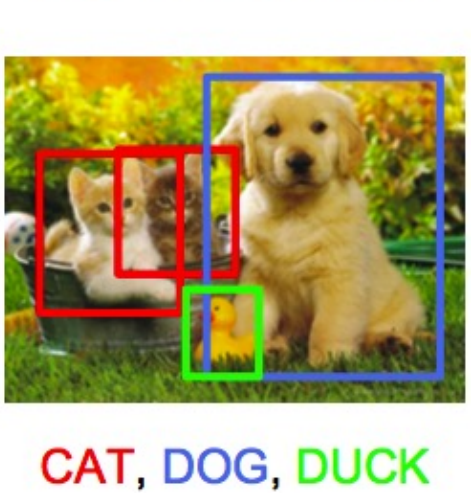
- CIFAR-10
  - Object Classification
  - 3 RGB channels from 0 to 255
  - 32 x 32 = 1024 pixels
  - 10 classes
  - 60,000 images

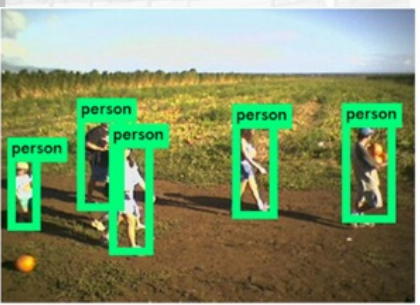
Dataset and Challenges

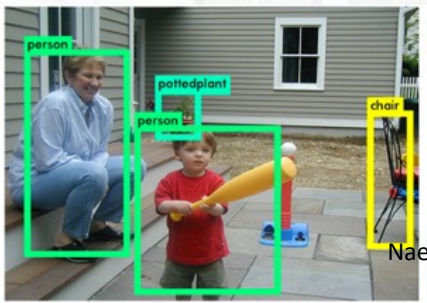
- ImageNet
  - Object Classification
  - 3 RGB channels from 0 to 255
  - Any resolution
  - 1000 classes
  - 14,197,122 images

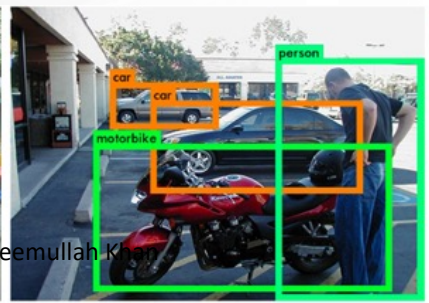
Dataset and Challenges

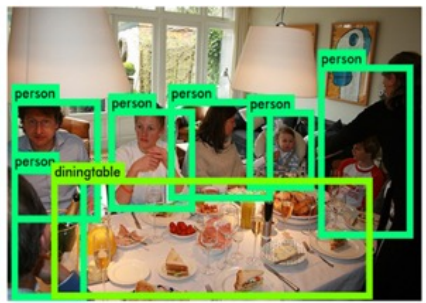
- Pascal VOC (2007)
  - Object Detection
  - 9,963 images
  - 24,640 annotated objects

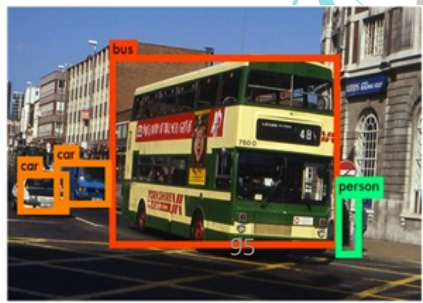












Dataset and Challenges

- Pascal VOC (2012)
  - Object Segmentation
  - 11,530 images
  - 27,450 objects
  - 6,929 segmentations

Instance Segmentation

CAT, DOG, DUCK

Image Classification – Applications

- Image/Photo search (Google Lens)

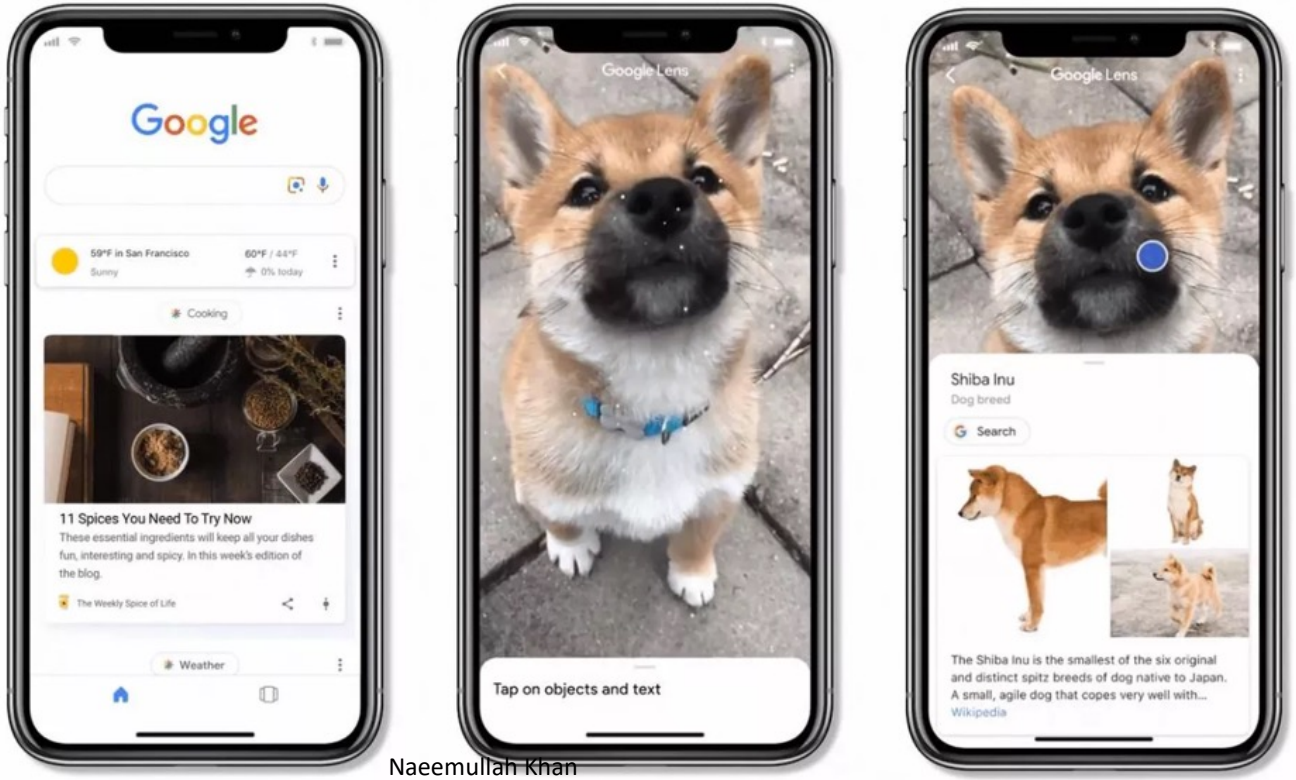


Image Classification – Applications

- Agriculture

PlantNet Plant Identification

Download on the App Store

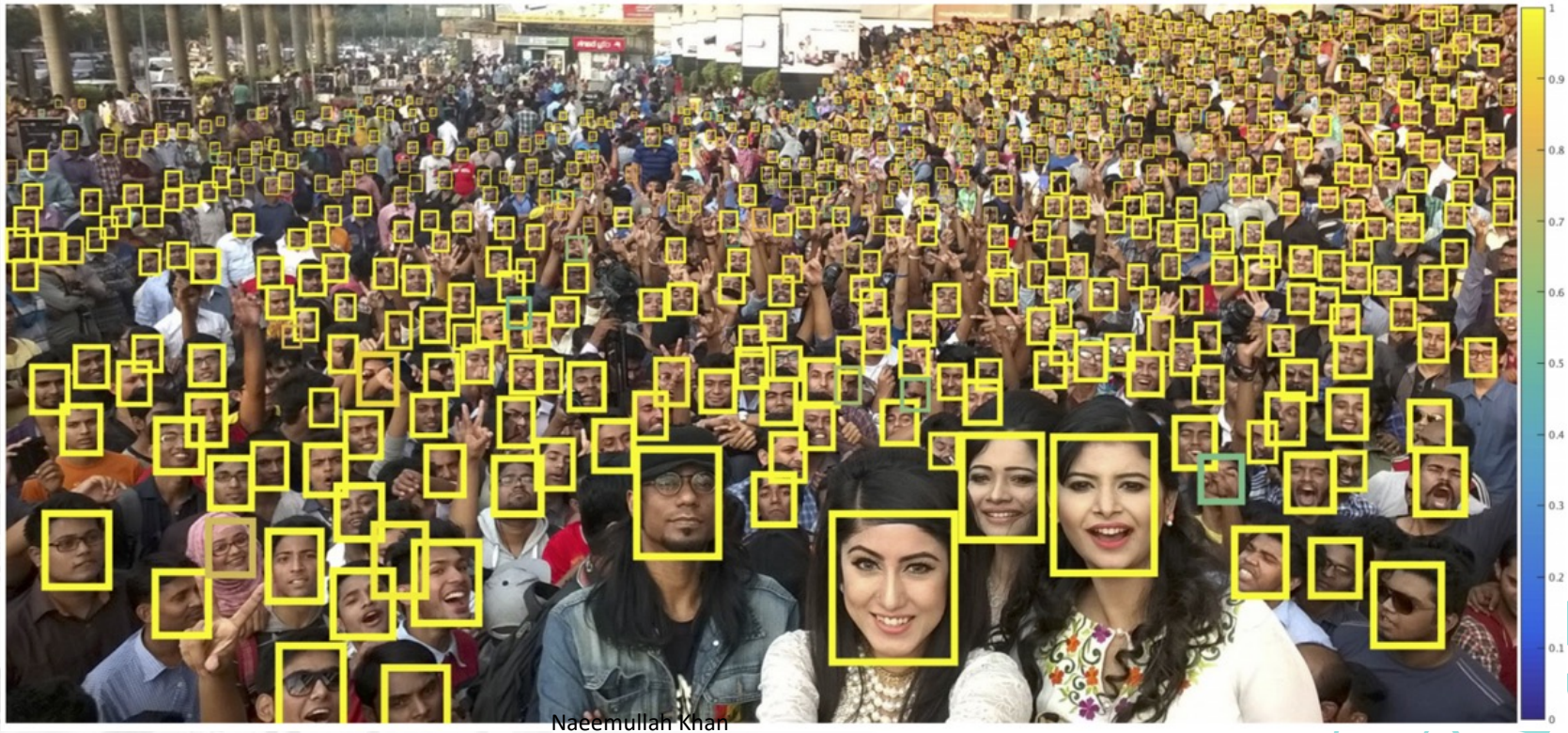
ANDROID APP ON Google play

Object Localization

- Detection as a classification
  - Sliding window

Object Localization - Application

- Finding tiny faces in the wild



Object Detection – Application

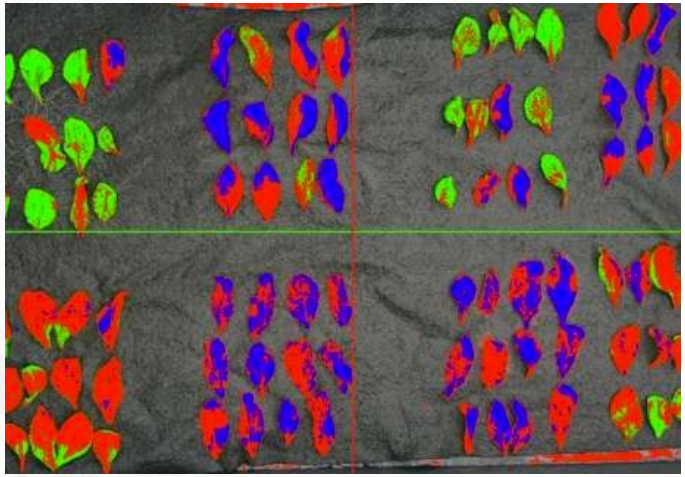
- Drone Application
  - Parking Lot Vehicle Detection Using DL

Naeemullah Khan

More on: https://medium.com/geoai/parking-lot-vehicle-detection-using-deep-learning-49597917bc4a

Object Detection – Application

- Agriculture:
  - Early plant disease detection
-



Naeemullah Khan

More on: https://phys.org/news/2017-04-drones-early-disease-crops.html#jCp

Image Segmentation

- SegNet

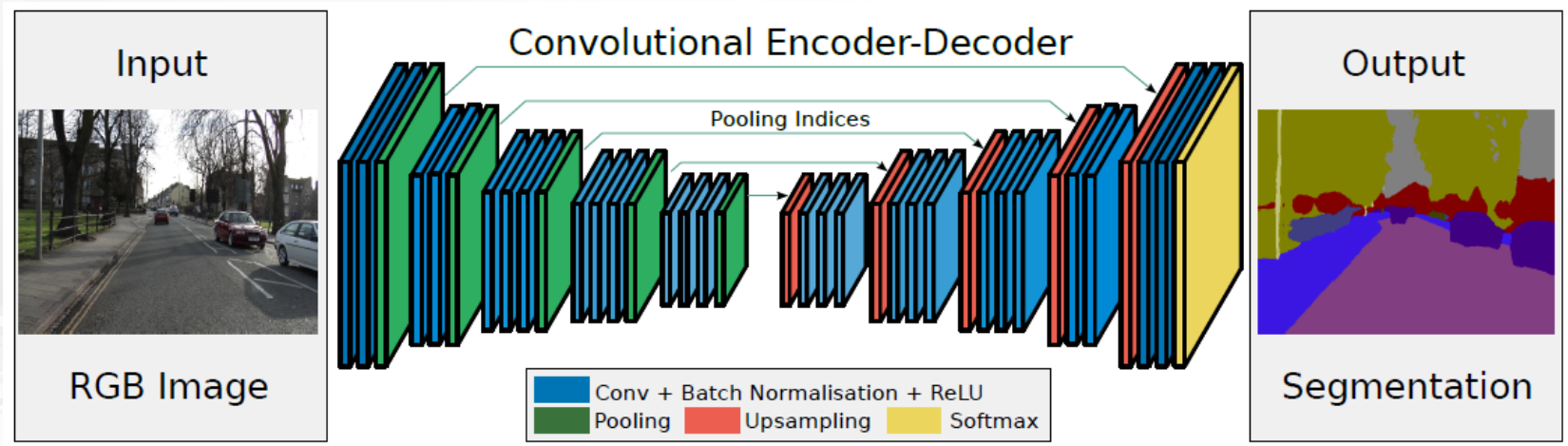
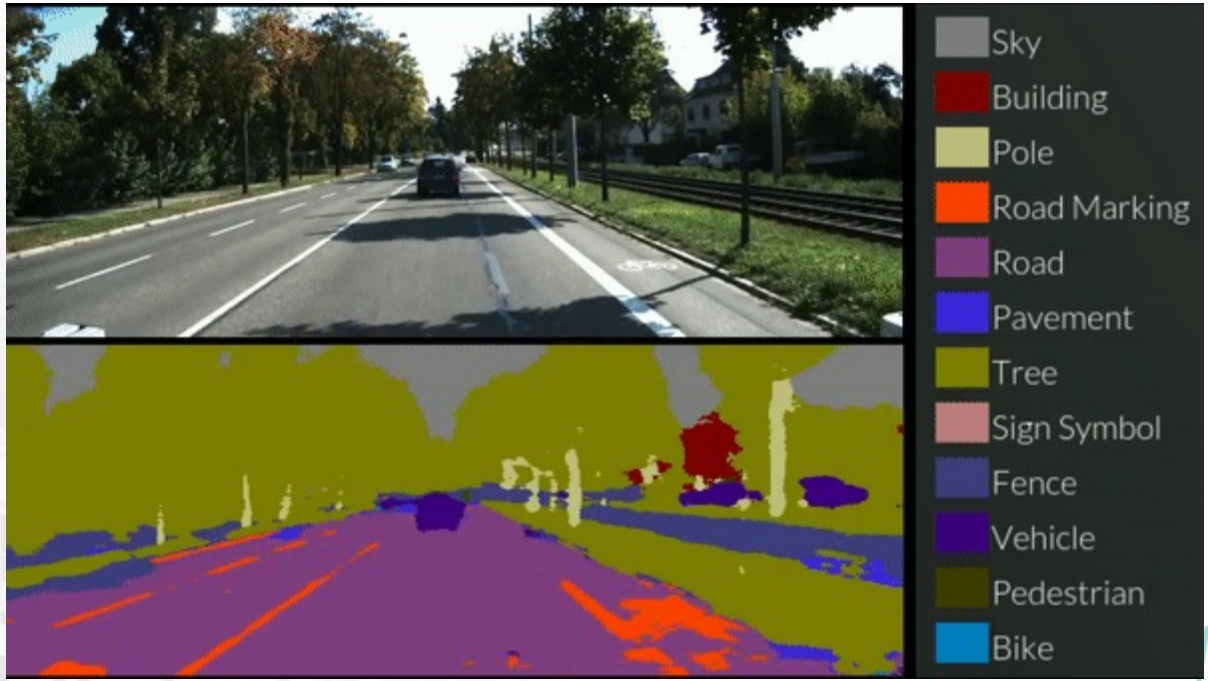


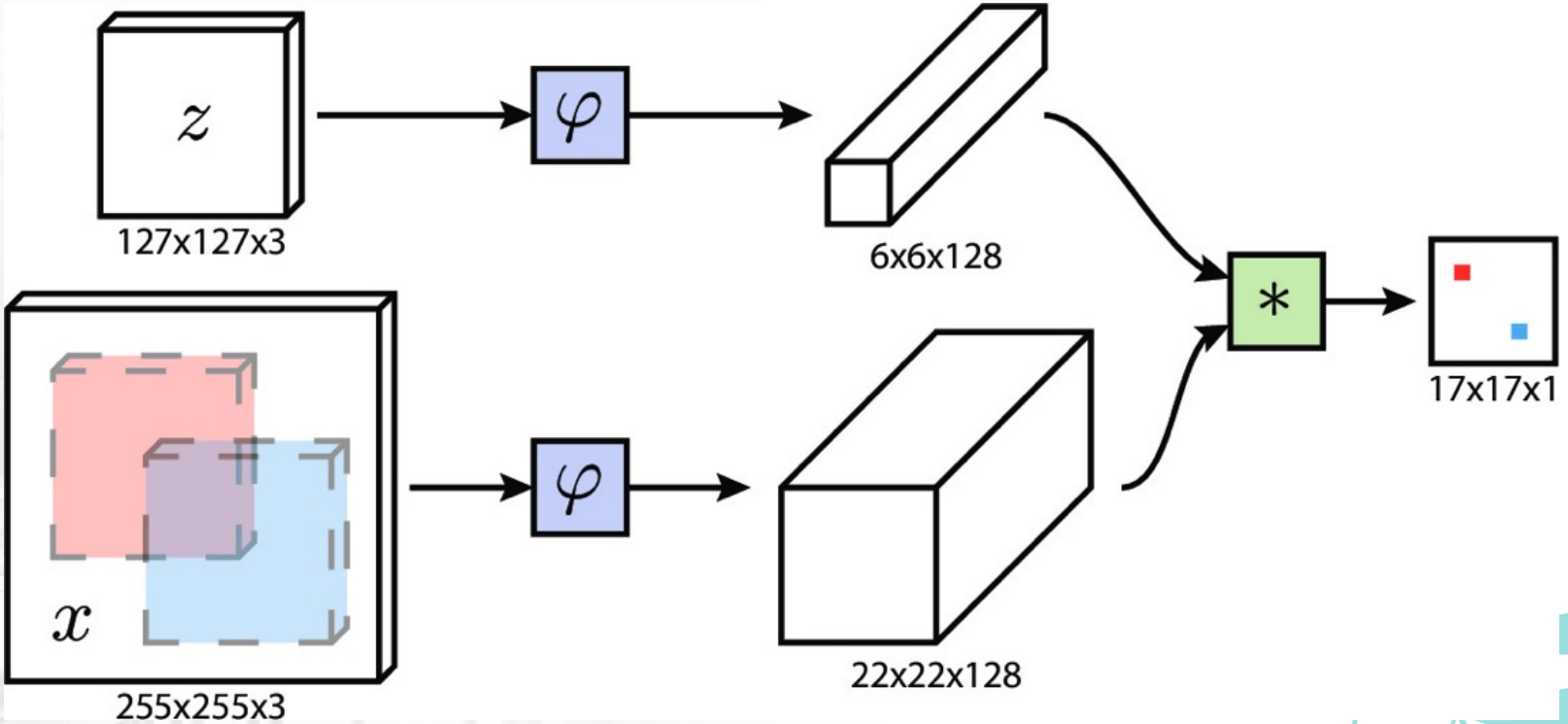
Image Segmentation – Application

- SegNet



Object Tracking

- Tracking

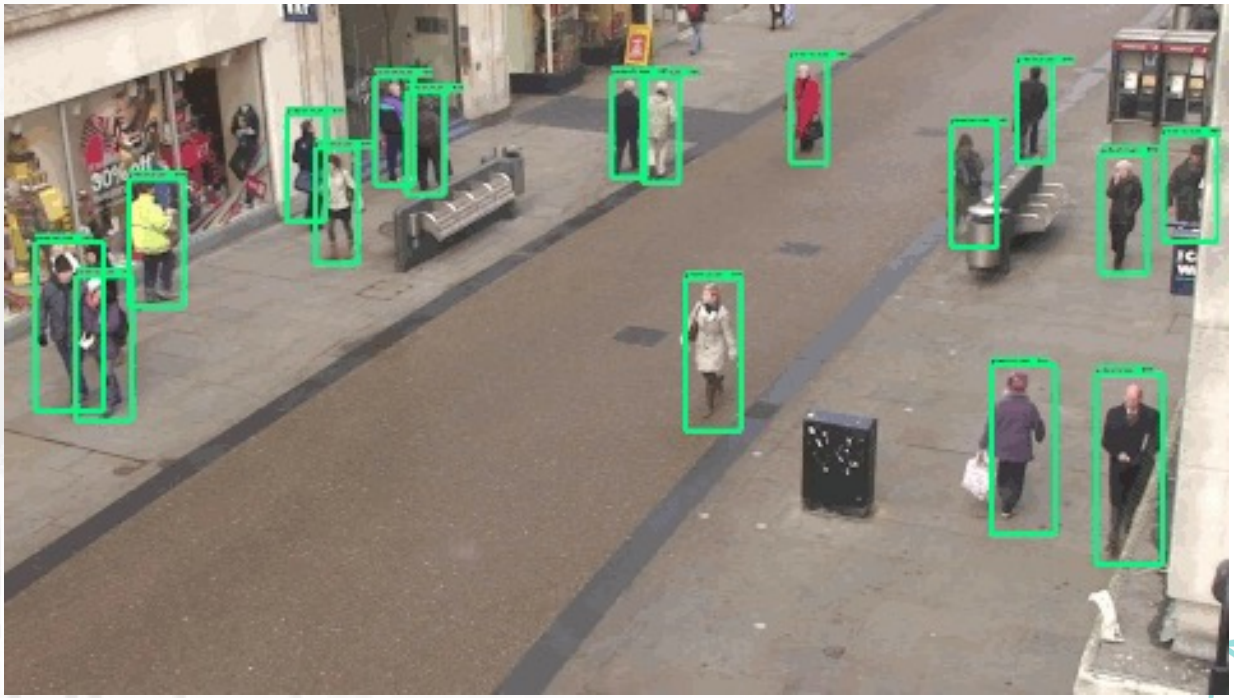


Object Tracking – Application

- Tracking

Object Tracking – Application

- People Tracking

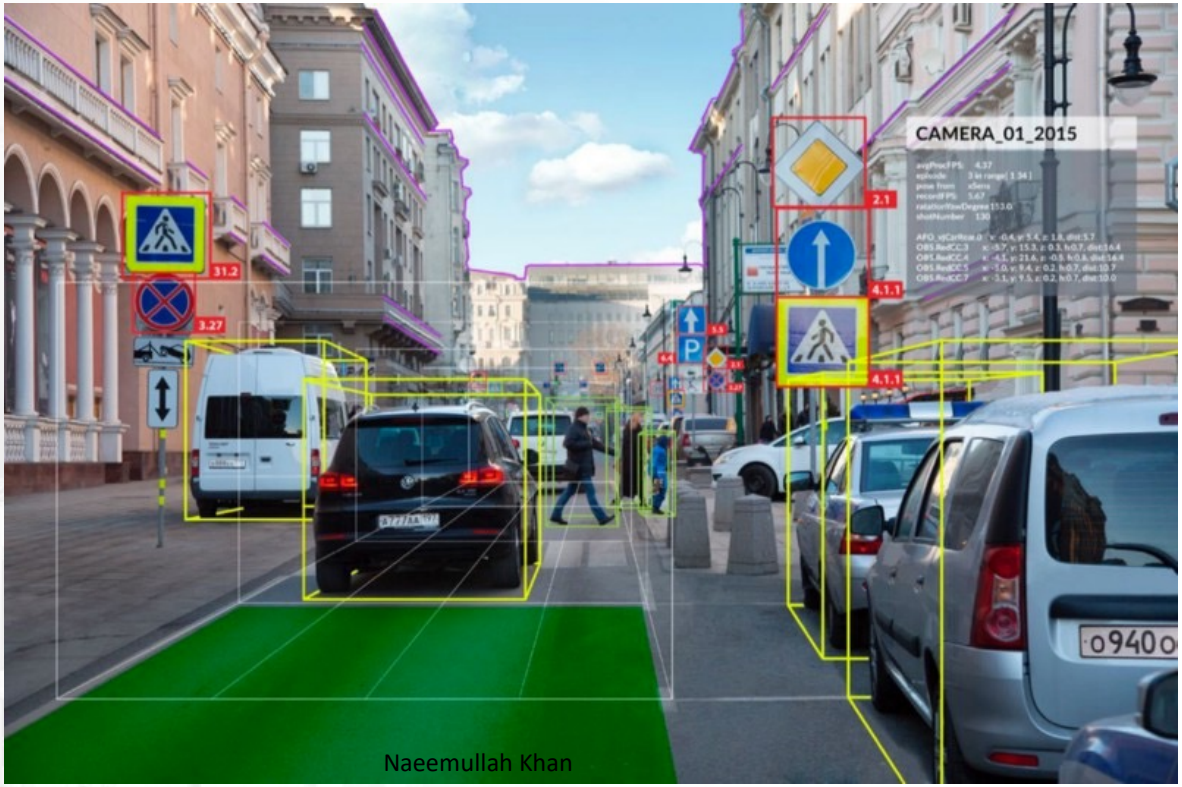


Object Tracking – Application

- UAV tracking

Computer Vision – Application

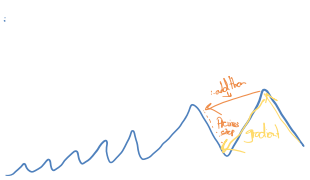
- Putting everything together:



ML and AI tools Usage Guide

- Supervised Learning
- Unstructured (Tabular) data Neural Networks
- Grid Data CNNs
- Time-series data (Transformers, RNNs, LSTMs)
- Graph Data (Molecule structure, Social Networks) Graph Neural Networks

Practice Guide



- Use avialable models architectures
- Data Augmentation
- Fine-Tuning
- Data Collection better than Architecture search