1. Computer Vision – ILSVRC DATASET, RESNET

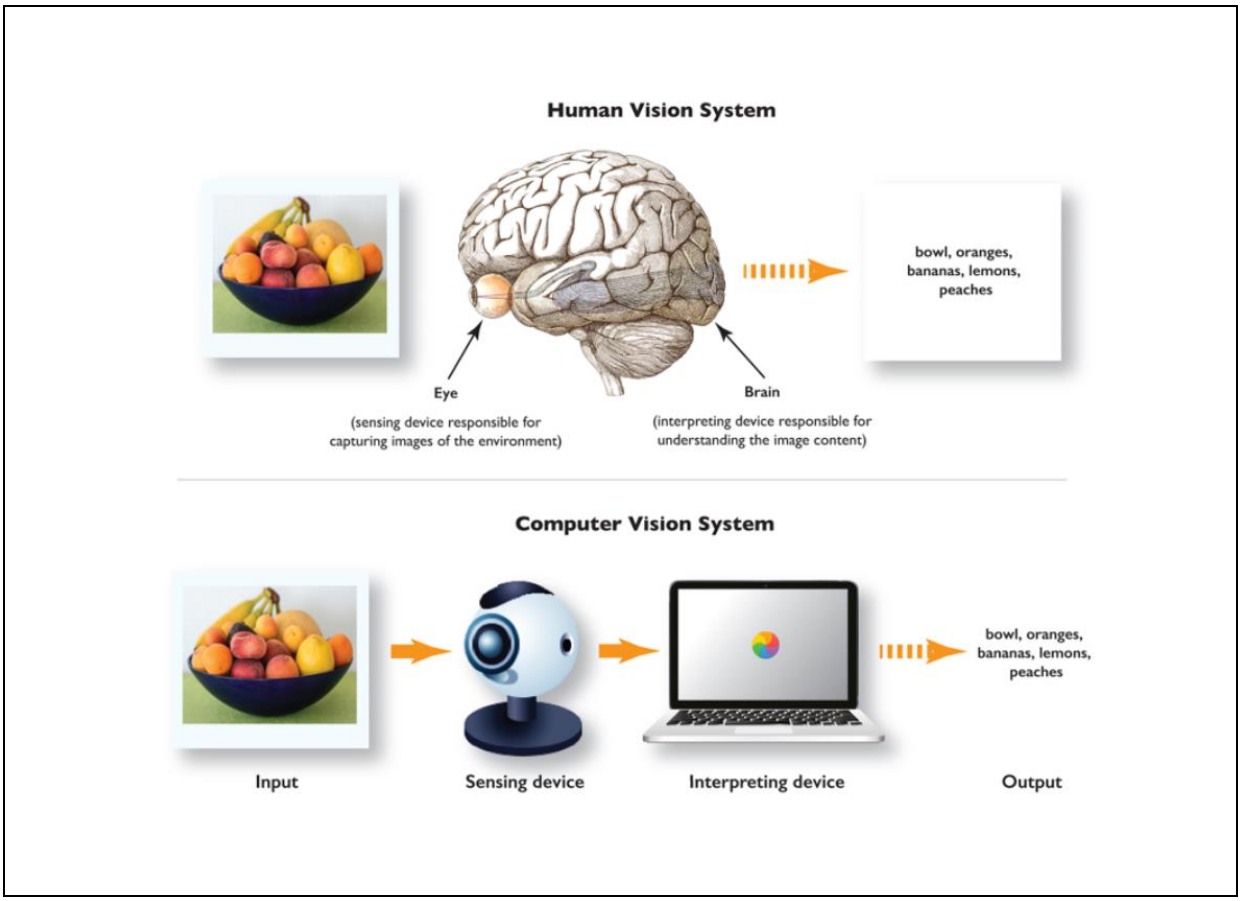
Contents

1. Computer Vision – ILSVRC DATASET, RESNET

1. Computer vision

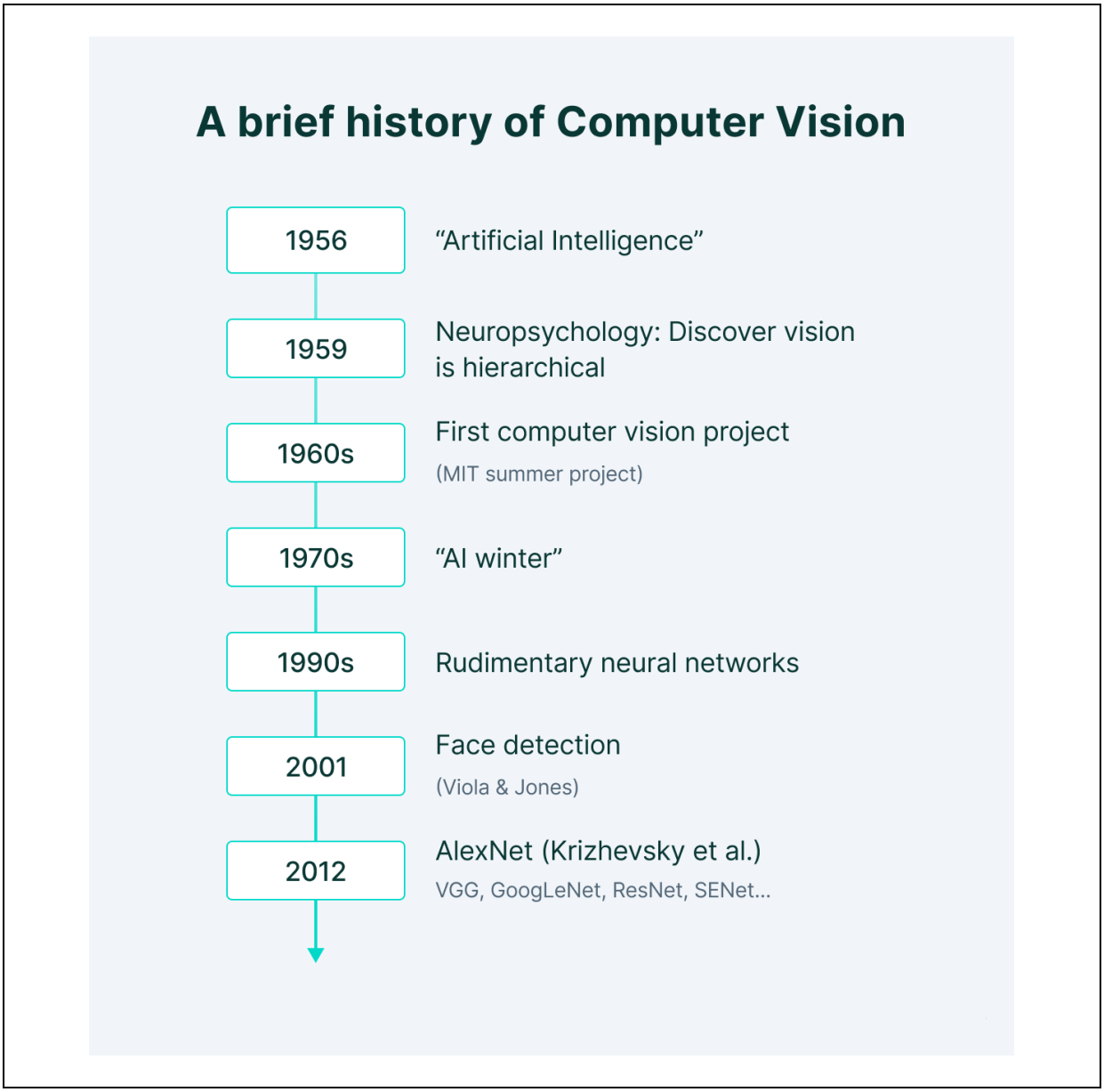
- ✓ Computer Vision is a subfield of Deep Learning and Artificial Intelligence.
- ✓ By using this human can teach computers to see and interpret the world around them.

2. Human & computer vision systems



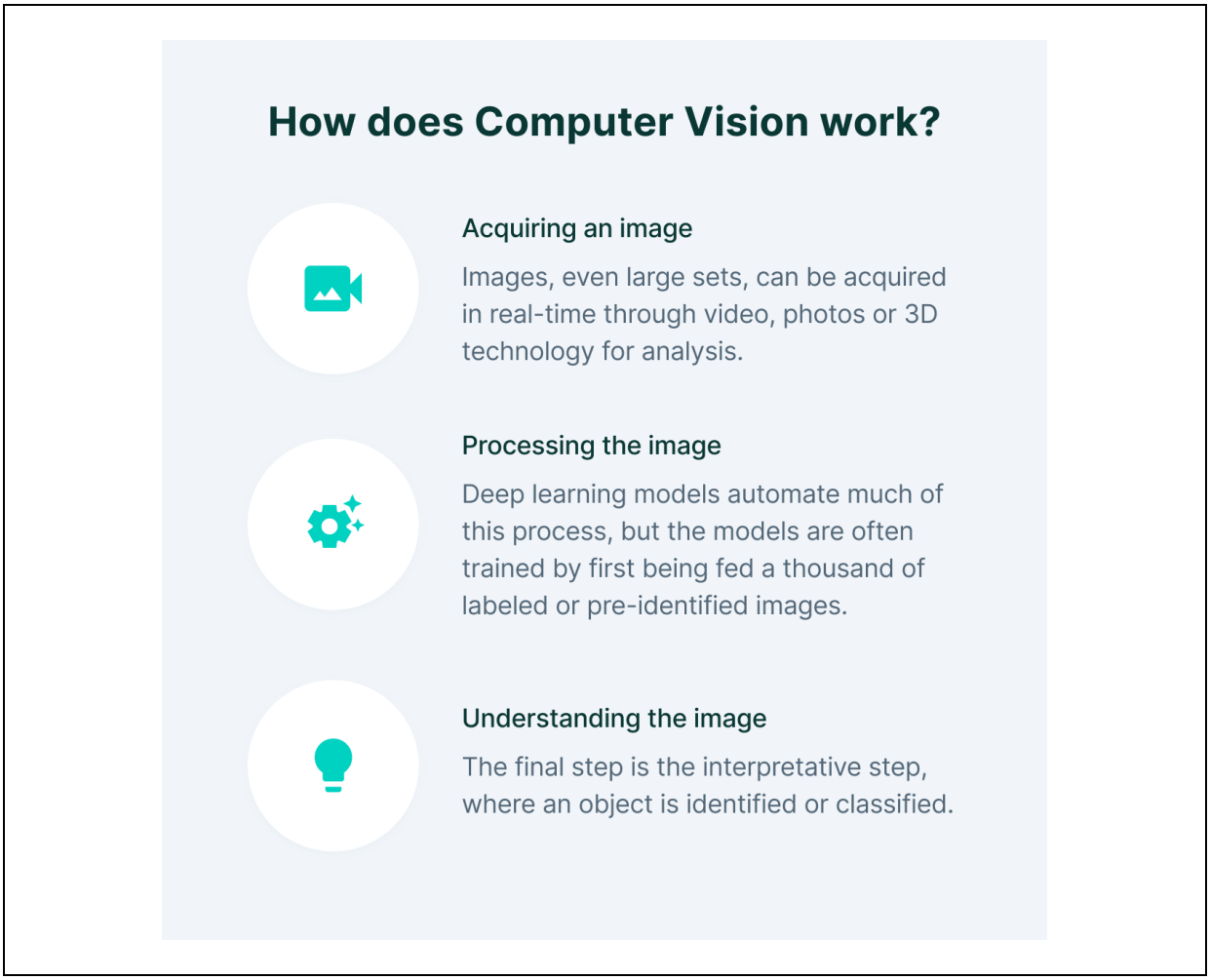
3. History of Computer Vision

- ✓ Machine vision, inspired by animal vision since 1959, evolved with advances in image technology.
- ✓ Key milestones: AI in the 1960s, OCR (Optical Character Recognition) in 1974, and complex topics by the 2000s.
  - Object identification
  - Facial recognition
  - Image Segmentation
  - Image Classification



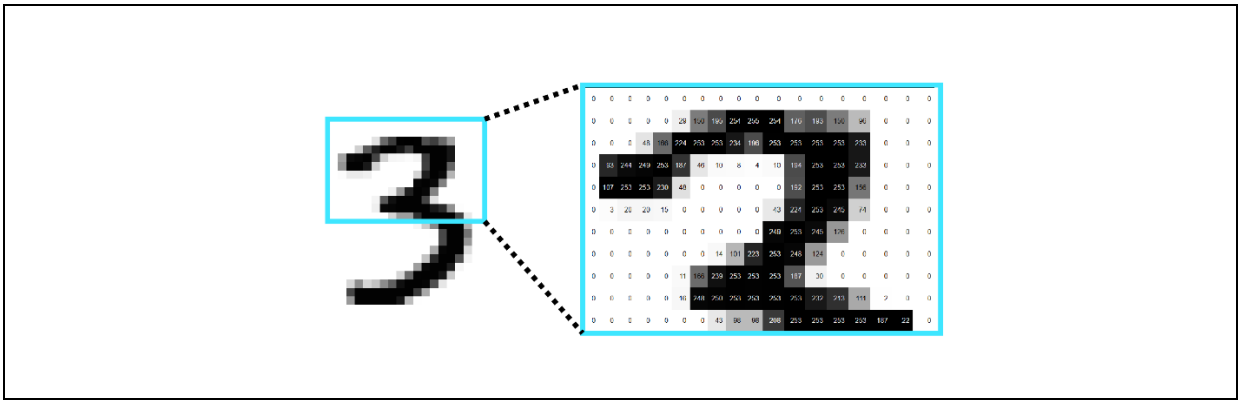
4. How Computer Vision works?

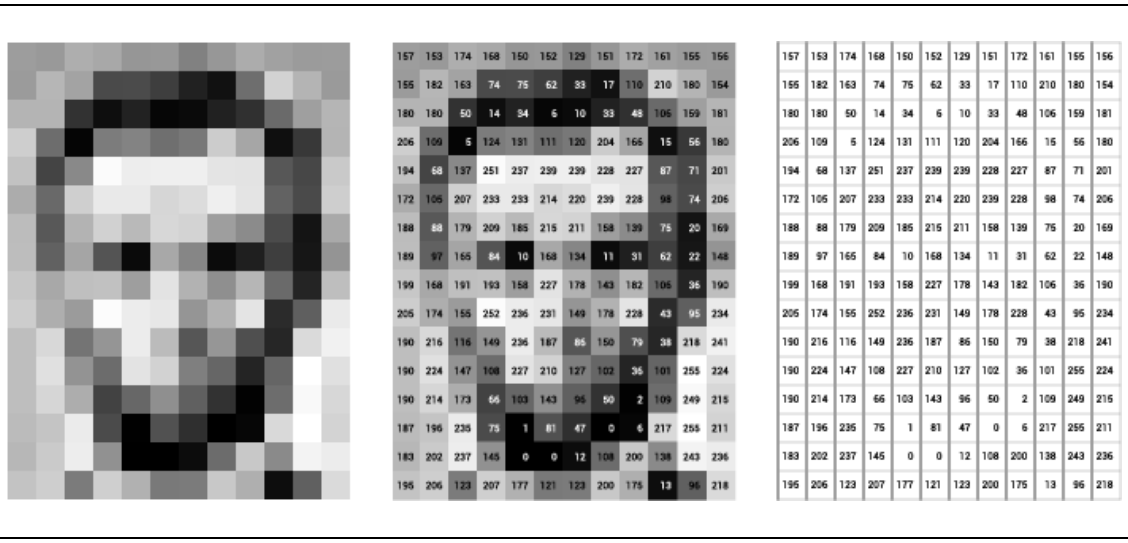
- ✓ There are mainly three steps,
  - Acquiring an image
  - Processing the image
  - Understanding the image



5. Image means pixel to computer

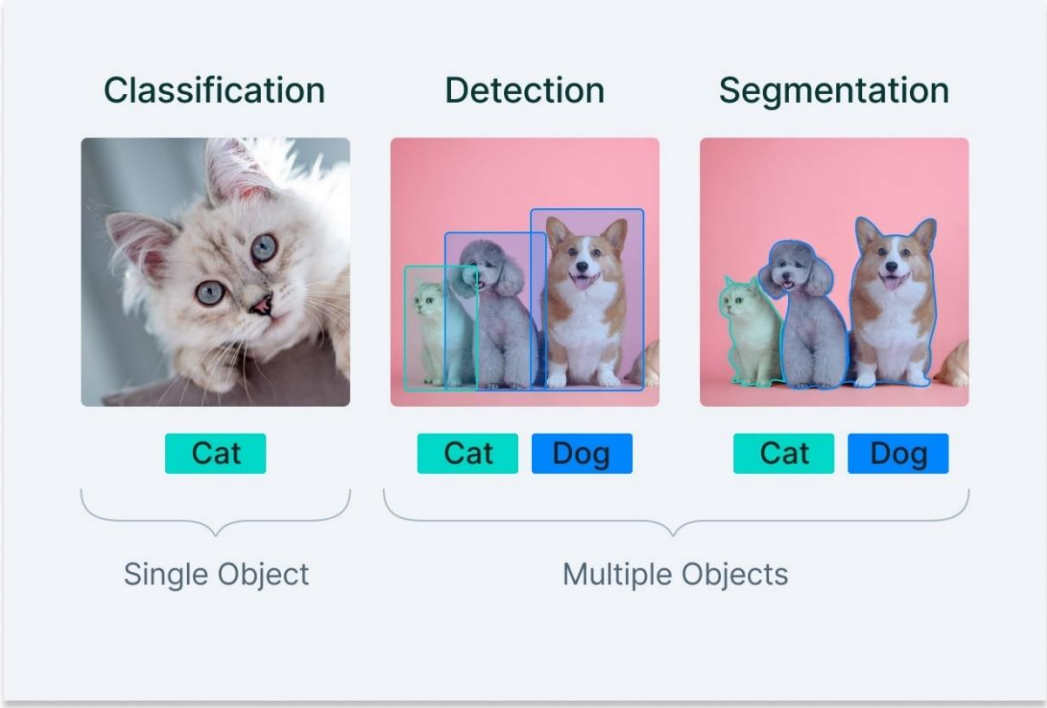
- ✓ An image is group of pixels.
- ✓ Computers process images as pixel arrays.
- ✓ Computer vision uses,
    - Linear algebra for simple tasks and convolutions with pooling for complex tasks.





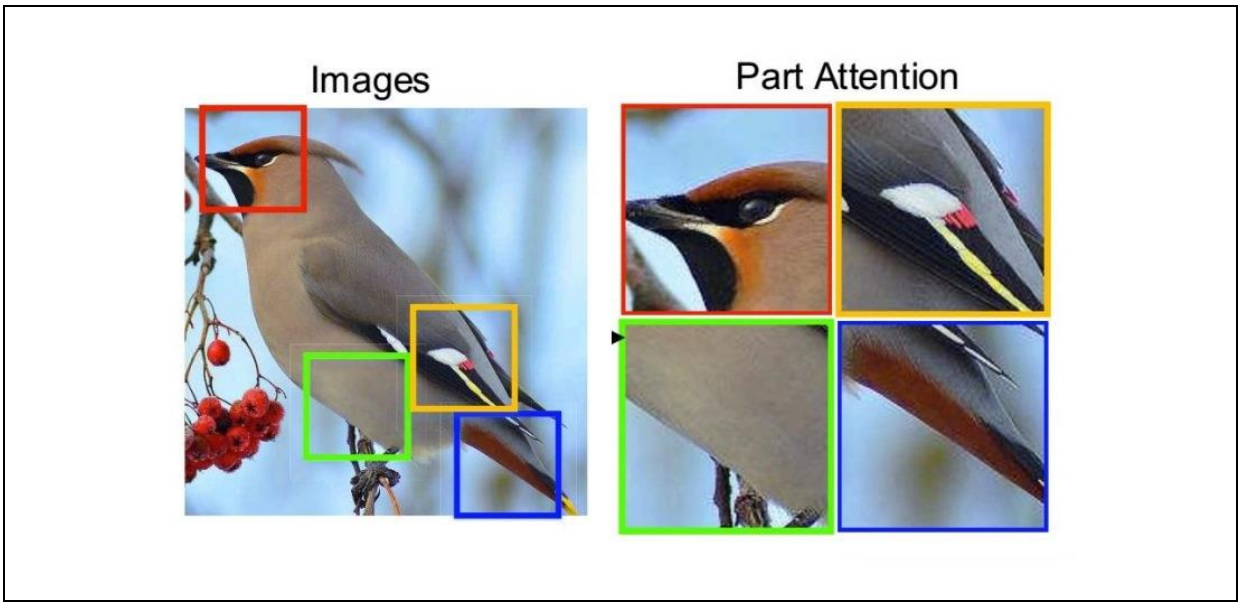
6. Common computer vision tasks

- ✓ Classification
- ✓ Detection
- ✓ Segmentation
- ✓ Face recognition
- ✓ Edge detection & etc



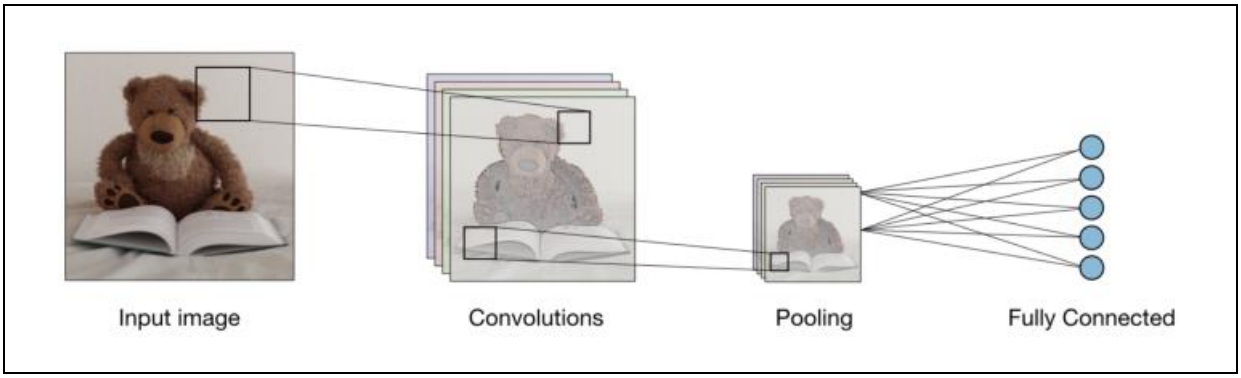
7. Human Image recognition

- ✓ Image recognition means, a computer or machine can see/observe the images.
- ✓ This is something like, **how our eye works.**
  - If we see an image then automatically it split into parts of image.
  - In next step our eyes can analyse sub-parts individually.
  - By assembling these parts, our eyes process and interpret the image.
- ✓ The same process was implemented in CNN or Convolutional Neural Network.



8. Short Introduction: Convolutional Neural Network

- ✓ CNN is a type of deep learning model.
- ✓ This is commonly used for image recognition and classification tasks.
- ✓ The main parts in CNN,
  - Convolution Layer
  - Pooling Layer
  - Fully-Connected Layer



8.1. Convolutional Layer

- ✓ In this layer, the input image is convolved (In maths: combine one function or series) with a set of learnable filters, also known as kernels.
- ✓ Every filter extracts different features from the input image by performing element-wise multiplication and summation operations.

8.2. Pooling layers

- ✓ Pooling layers reduce the size of the input by taking either the maximum or average value from small regions.
- ✓ This makes the network faster and helps prevent overfitting.

8.3. Fully Connected Layer

- ✓ Connects every neuron in the previous layer to every neuron in the next layer.
- ✓ This layer learns global features.

9. Image recognition task by CNN

- ✓ CNN having layers to recognize the image.
- ✓ Every layer can learn and delivery each task as,
    - The first layer can learn to detect basic edges and shapes.
    - The second layer can recognize these more detailed features.
    - The third layer can recognize parts of objects etc.
- ✓ This learning process helps the network to build a detailed understanding of the data.

10. ImageNet

10.1. The Birth of ImageNet

- ✓ Once upon a time, in the mid-2000s, a brilliant researcher named Fei-Fei Li.
  - https://profiles.stanford.edu/fei-fei-li
  - https://en.wikipedia.org/wiki/Fei-Fei_Li
- ✓ She had an idea like, just as humans learn by seeing countless objects and scenes throughout their lives.
- ✓ Her idea to make computers see and understand the world like humans.
- ✓ So, here needed a vast amount of data.
- ✓ She and her team aimed to create ImageNet.
- ✓ Started gathering images and created dataset
- ✓ The ImageNet is a very large collection of a dataset.
  - https://image-net.org/index.php

10.2. Dataset Information

- ✓ Gathering images, labelling them accurately was really a challenging task.
- ✓ Amazon Mechanical Turk thousands of team members manually verify and annotate the images.

10.3. Dataset Information

- ✓ This dataset contains,
  - More than **14** million images
  - More than **22** thousand groups or classes.
  - More than **1** million images that have bounding box annotations

10.4. The Birth of the Competition: ILSVRC

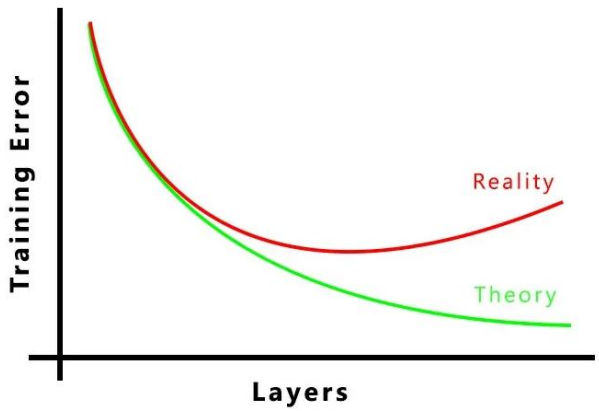
- ✓ ILSVRC the full form is,
  - ImageNet Large Scale Visual Recognition Challenge
- ✓ In 2010, Fei-Fei Li and her team introduced the ImageNet Large Scale Visual Recognition Challenge (ILSVRC).
- ✓ This annual competition invited researchers from around the world to develop algorithms that could classify and detect objects within the ImageNet dataset.
- ✓ It quickly became the benchmark for measuring progress in computer vision.

11. The problem with deep neural networks

- ✓ To get the better results, the architecture used to add more and more layers, but this can lead to vanishing gradient issues.

12. If we increase more layers then…

- ✓ If we increase more layers (deep networks) then we may get the problem of vanishing/exploding gradients.



13. Vanishing / Exploding Gradients

- ✓ If we add more layers to the neural network then one of the problems is like, gradients can become too small (vanishing) or too large (exploding).
- ✓ Due to this Neural Network cannot learn effectively.

14. How to solve Vanishing / Exploding Gradients problem?

- ✓ We can solve this problem by using **ResNet**.

15. History of ResNet

- ✓ ResNet was implemented by Kaiming He and his team, in 2015.
- ✓ This research team achieved top results for object detection and object detection with localization tasks.

15.1. ResNet Paper

- ✓ This research paper written for Deep Residual Learning for Image Recognition.
- ✓ Find the below link for research paper,
  - https://arxiv.org/abs/1512.03385

15.2. Winner of ILSVRC 2015

- ✓ ResNet won the champion for ILSVRC 2015 in,
  - Image classification, detection, and localization

16. ResNet Introduction

- ✓ ResNet stands for Residual Network.
- ✓ It's a type of deep learning model.
- ✓ The main goal of ResNet (Residual Network) is to address the vanishing gradient problem.
- ✓ ResNet proves that, it is easy to train very deep neural networks.

17. ResNet Example

Program Name: ResNet Example
demo1.py

```
import numpy as np

from keras.preprocessing import image
from keras.applications import ResNet50
from keras.applications.resnet50 import preprocess_input
from keras.applications.resnet50 import decode_predictions

model = ResNet50(weights = 'imagenet')

def load_and_preprocess_image(img_path):
        img = image.load_img(img_path, target_size = (224, 224))
        img_array = image.img_to_array(img)
        img_array = np.expand_dims(img_array, axis = 0)
        img_array = preprocess_input(img_array)

        return img_array


img_path = '1.jpeg'
img_array = load_and_preprocess_image(img_path)
predictions = model.predict(img_array)
decoded_predictions = decode_predictions(predictions, top =
1)[0]

for i, (imagenet_id, label, score) in
enumerate(decoded_predictions):
        print(f"{i + 1}: {label} ({score:.2f})")
```

Output

```
goldfish
```