# Intelligent Consumer Technologies

Prof. Paolo Napoletano
a.a. 2023/2024

Signal, image, and natural language processing in Consumer Technologies

# Computer Vision in Consumer Technologies

Topics: Computer Vision, CV in Consumer Technologies, Face Detection, Face Recognition

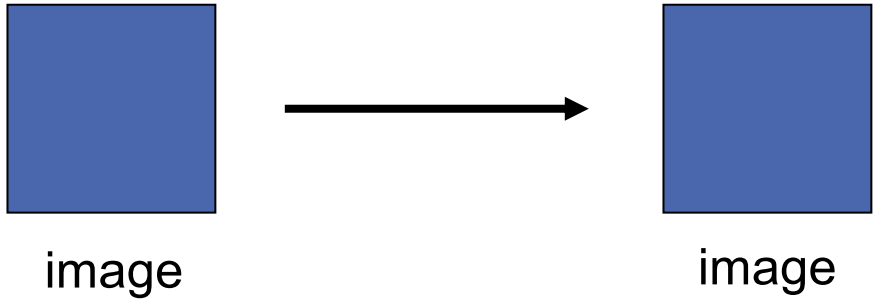
Learning Objectives

- Being aware of the variety of CV tasks
- Being able to define CV modules in CT applications
- Being able to define a pipe for face detection and recognition
- Being able to define the modules of a photo content management system

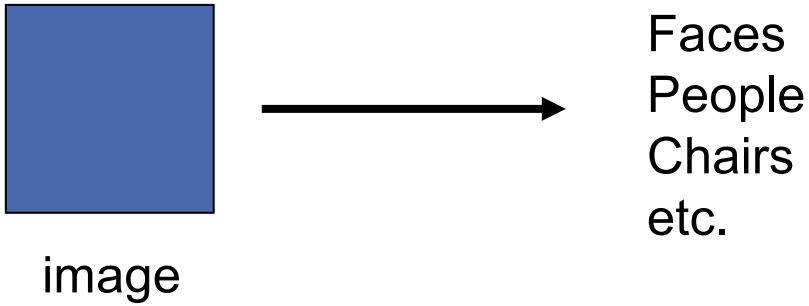
# Image Processing and Computer Vision

## definitions

- **Image Processing**
  - Research area within electrical engineering/signal processing
  - Focus on syntax, low level features

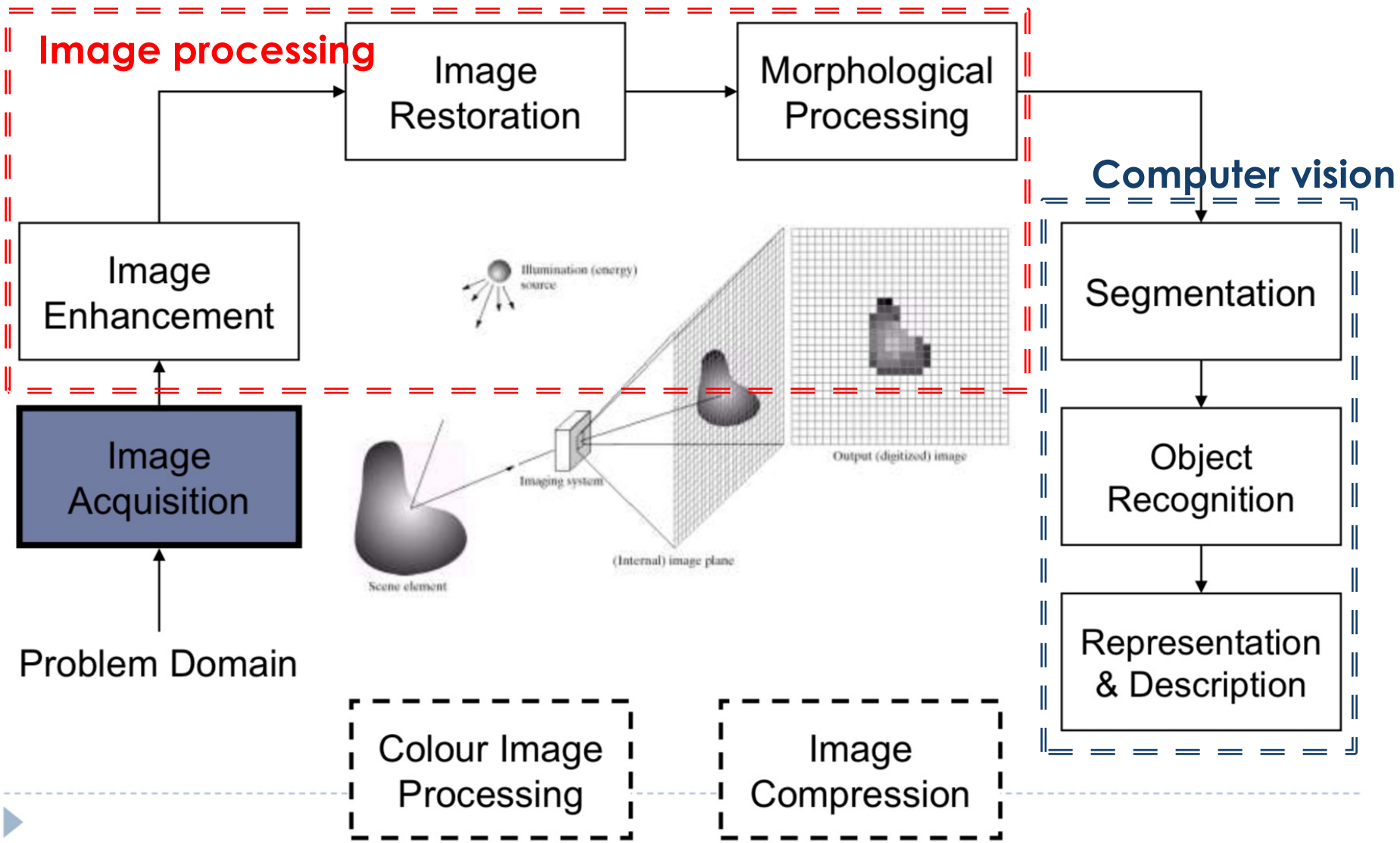



- **Computer Vision**
  - Research area within computer science/artificial intelligence
  - Focus on semantics, symbolic or geometric descriptions



# Image Processing and Computer Vision

## definitions



# Computer Vision

## definitions

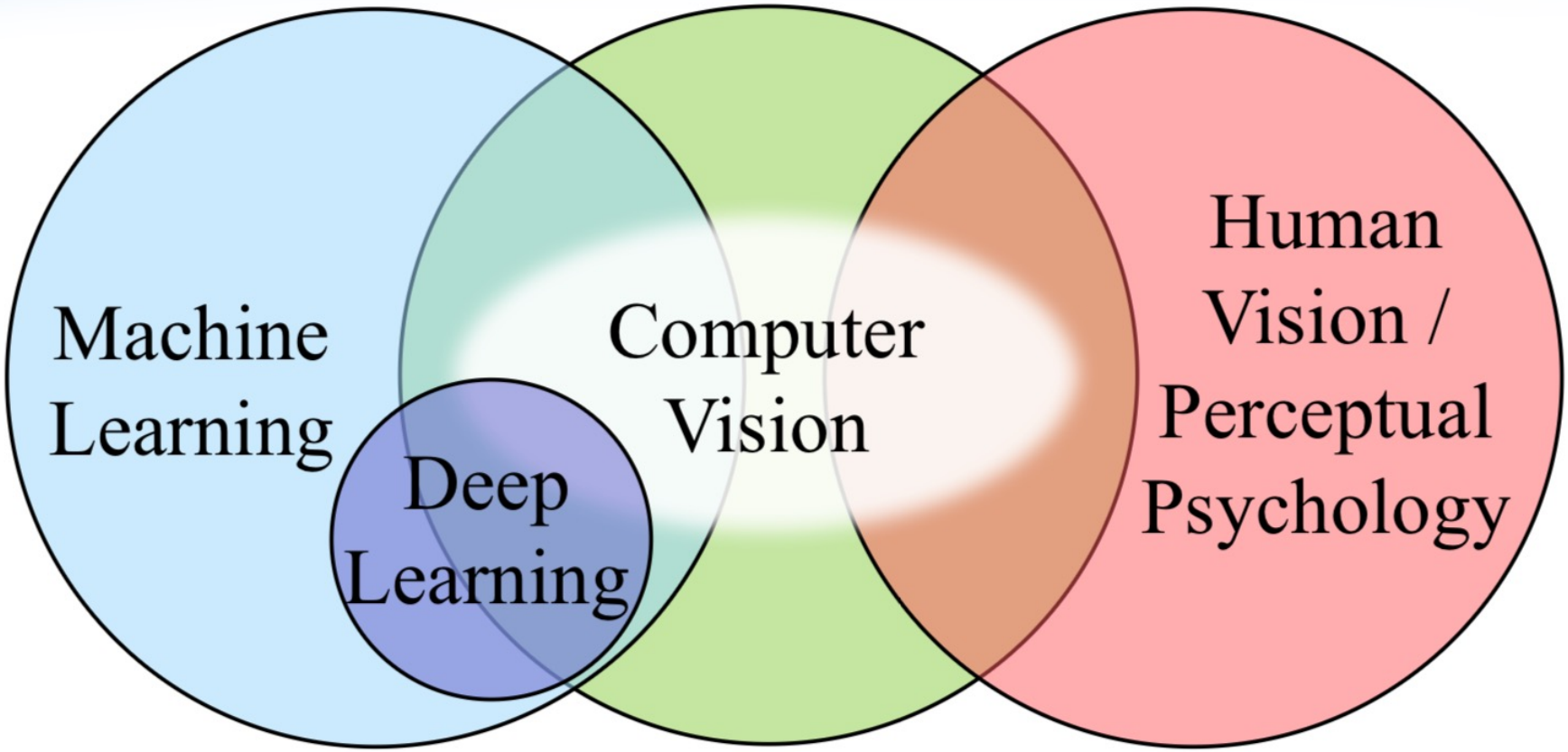
- An interdisciplinary field that deals with how **computers** can be made to gain **understanding** from **digital images** or **video**.

- From the perspective of engineering, it seeks to **automate** tasks that the **human visual system** can do."


* More details on https://developer.ibm.com/articles/introduction-computer-vision/

# Computer Vision

## definitions

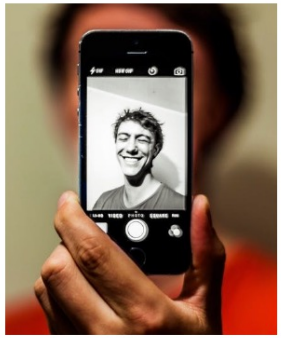



* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

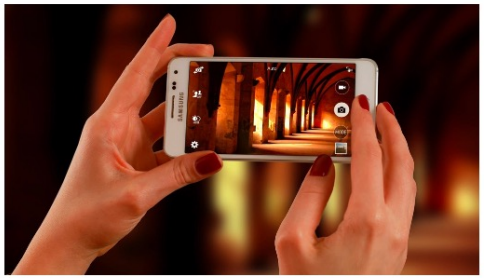
# Computer Vision is Everywhere

## Applications



Left to right:
Image by Roger H Goun is licensed under CC BY 2.0
Image is CC0 1.0 public domain
Image is CC0 1.0 public domain
Image is CC0 1.0 public domain




Left to right:
Image is free to use
Image is CC0 1.0 public domain
Image by NASA is licensed under CC BY 2.0
Image is CC0 1.0 public domain



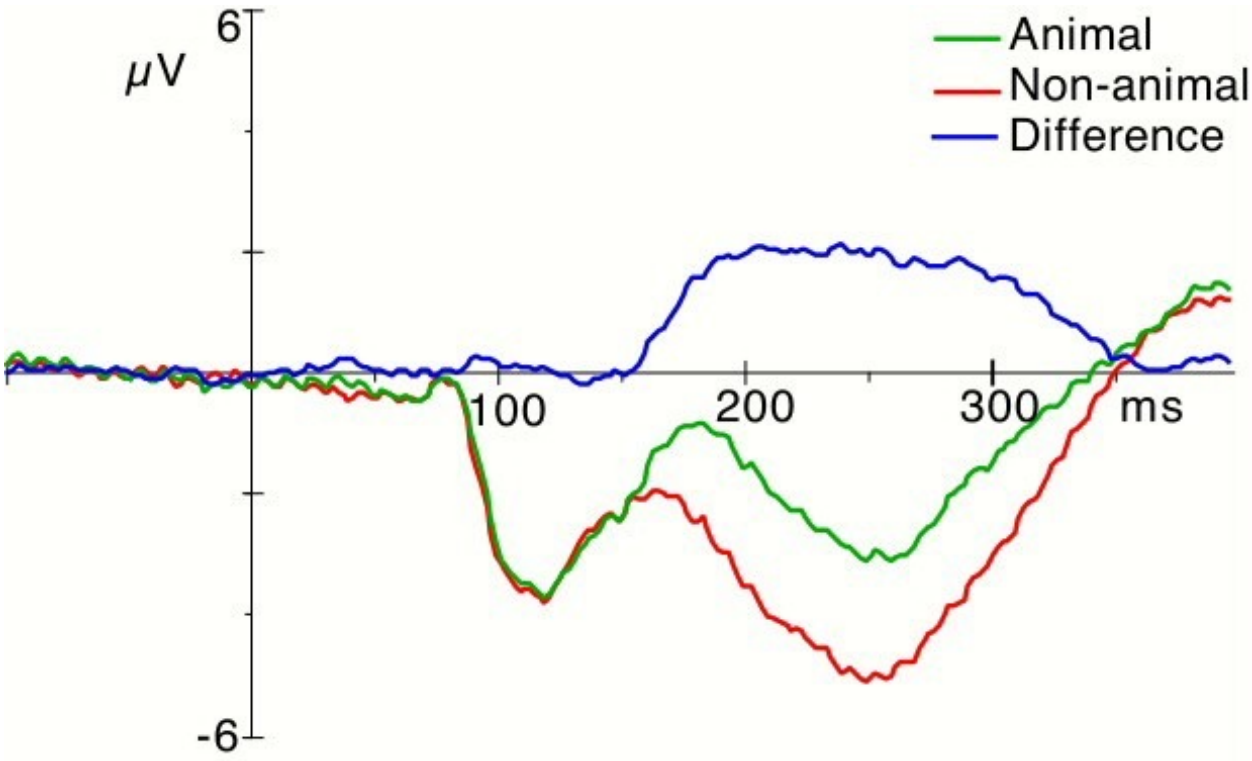

Bottom row, left to right
Image is CC0 1.0 public domain
Image by Derek Keats is licensed under CC BY 2.0; changes made
Image is public domain
Image is licensed under CC-BY 2.0; changes made

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

Studies have shown that humans recognize a real-world scene at a single glance.

# Human Vision - perception

## Speed of visual processing in our brain




**150 ms !!**

Thorpe, et al. *Nature*, 1996


* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

Look at the next three images
and
guess the subject of each image!

How smart are machine-learning-based solutions expert in photo understanding?

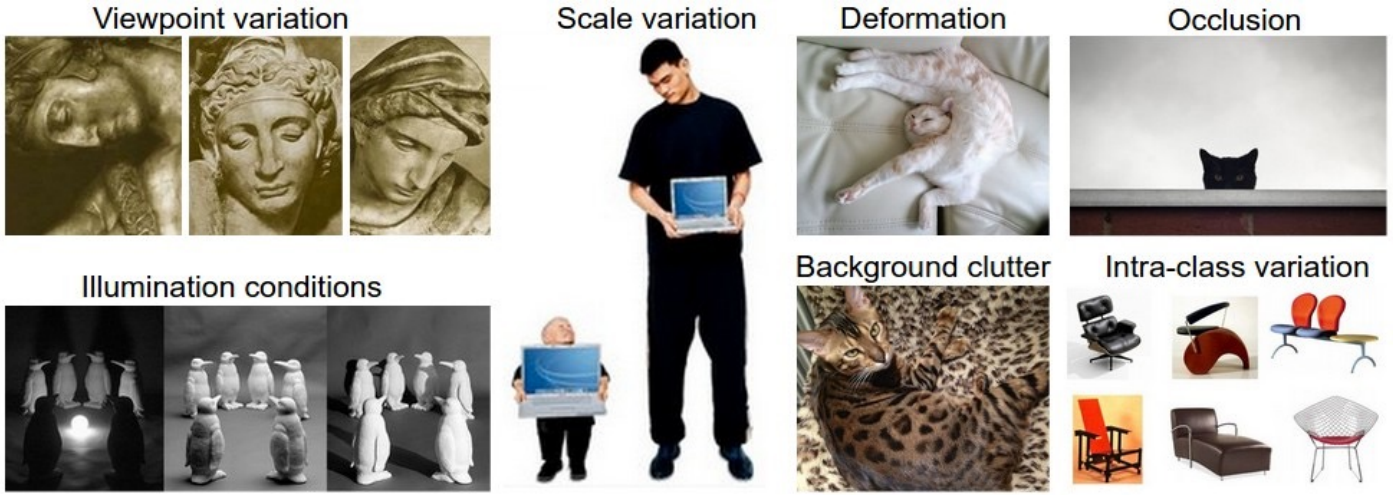
# Dog, Pig or Loaf of Bread?

a bit of funny ☺

from the Netflix movie **The Mitchells vs. The Machines**

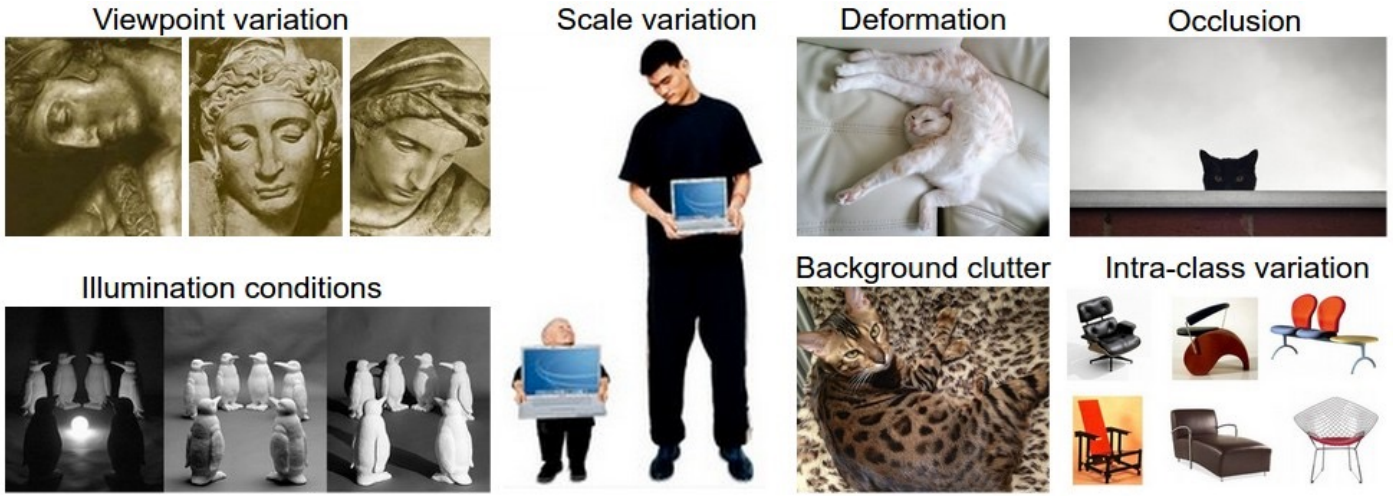
# Challenges in Image Recognition

- **Viewpoint variation**. A single instance of an object can be oriented in many ways with respect to the camera.
- **Scale variation**. Visual classes often exhibit variation in their size (size in the real world, not only in terms of their extent in the image).
- **Deformation**. Many objects of interest are not rigid bodies and can be deformed in extreme ways.
- **Occlusion**. The objects of interest can be occluded. Sometimes only a small portion of an object (as little as few pixels) could be visible.



# Challenges in Image Recognition

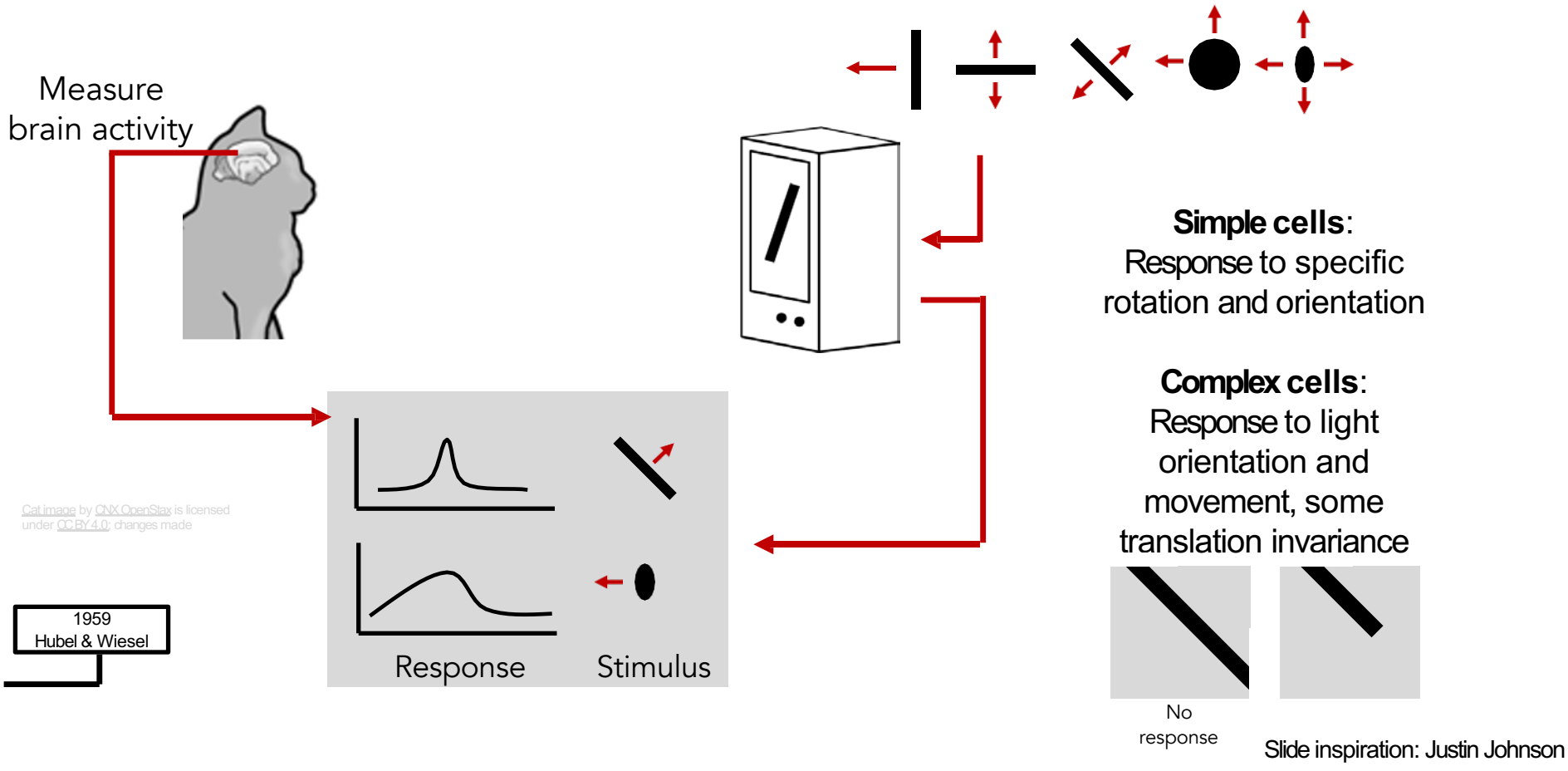
- **Illumination conditions**. The effects of illumination are drastic on the pixel level.
- **Background clutter**. The objects of interest may *blend* into their environment, making them hard to identify.
- **Intra-class variation**. The classes of interest can often be relatively broad, such as *chair*. There are many different types of these objects, each with their own appearance.



# Computer Vision from 50’s to nowadays

# Hubel and Wiesel, 1959

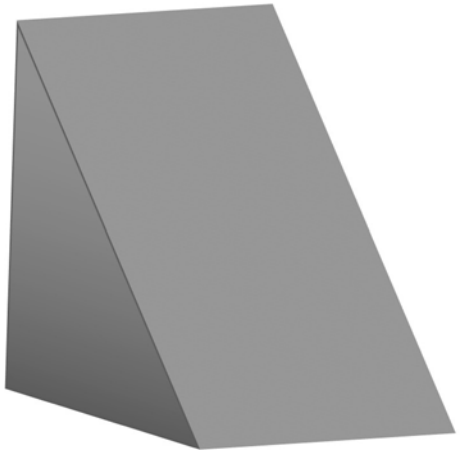
## From 50's to nowadays



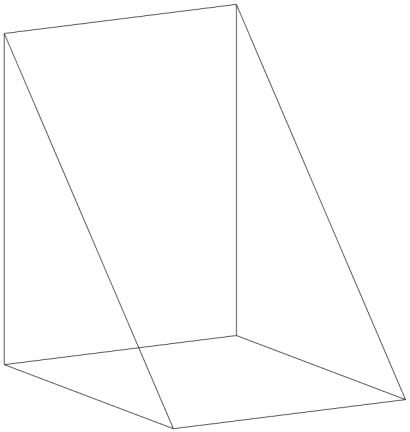

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# Larry Roberts, 1963

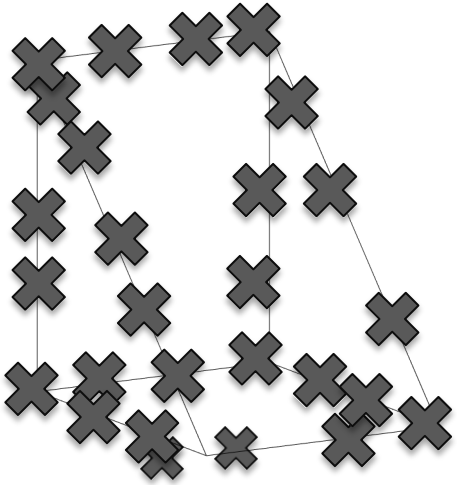
## From 50's to nowadays



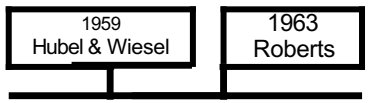
(a) Original picture



(b) Differentiated picture



(c) Feature points selected




Lawrence Gilman Roberts, "Machine Perception of Three-Dimensional Solids", 1963

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# Seymour Papert, 1966

## From 50's to nowadays

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

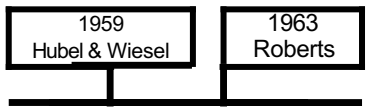
Artificial Intelligence Group — July 7, 1966
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

https://dspace.mit.edu/handle/1721.1/6125

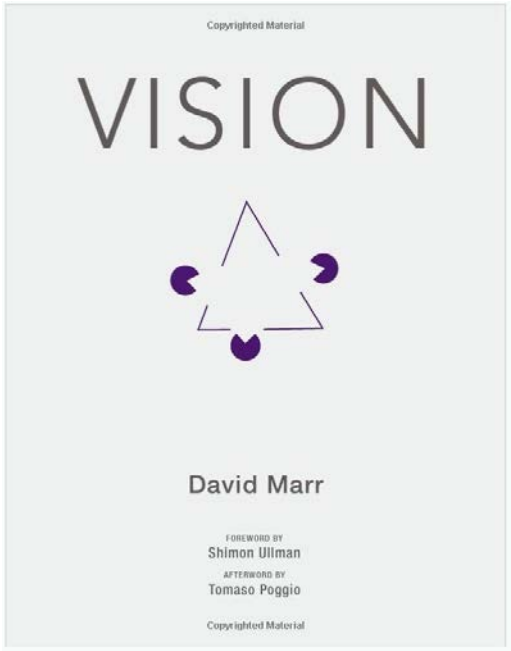


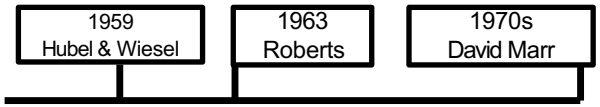

Slide inspiration: Justin Johnson

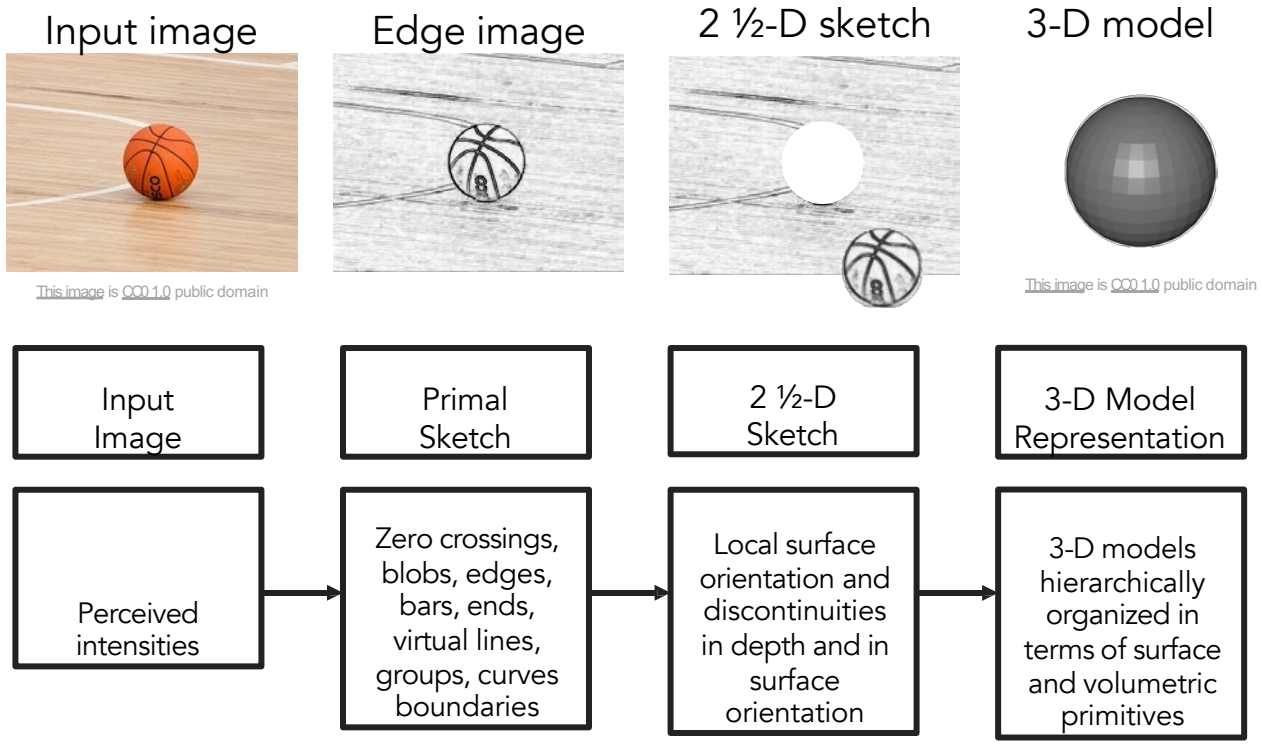
* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li, further reading https://zbigatron.com/the-early-history-of-computer-vision/

# David Marr, 1970

## From 50's to nowadays



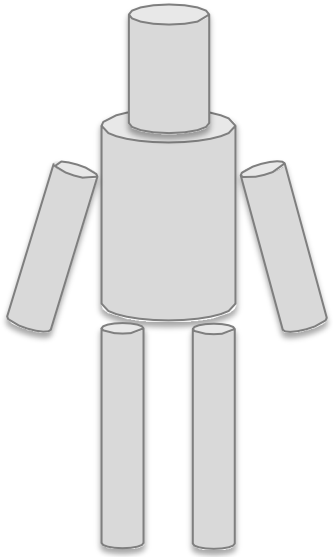




Stages of Visual Representation, David Marr, 1970s

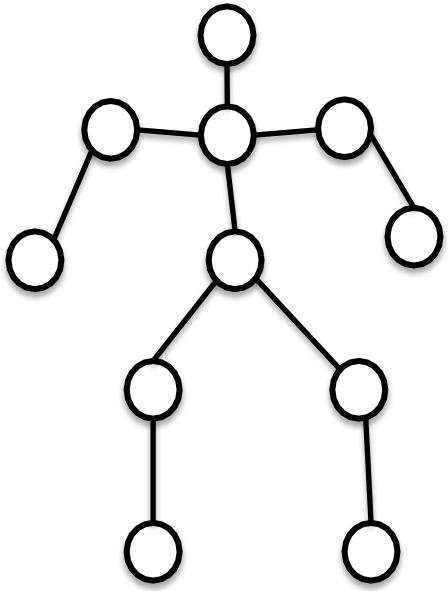
* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# Recognition via Parts, 1970s

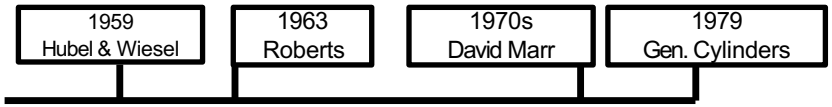
## From 50's to nowadays



Generalized Cylinders,
Brooks and Binford,
1979



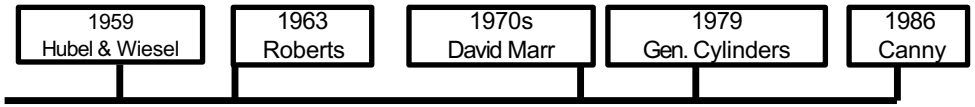
Pictorial Structures,
Fischler and Elshlager, 1973




* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# Recognition via Edge detection, 1980s

## From 50's to nowadays




John Canny, 1986
David Lowe, 1987



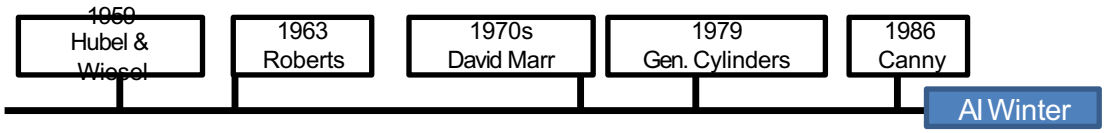
Slide inspiration: Justin Johnson

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# Arriving at an «AI winter»

## From 50's to nowadays

- Enthusiasm (and funding!) for AI research dwindled
- "Expert Systems" failed to deliver on their promises
- But subfields of AI continues to grow
  - Computer vision, NLP, robotics, compbio, etc.



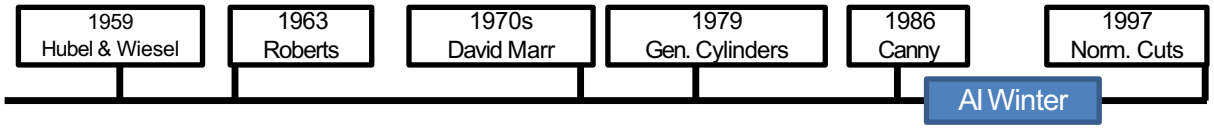



Slide inspiration: Justin Johnson

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# Recognition via Grouping, 1990s

## From 50's to nowadays




Normalized Cuts, Shi and Malik, 1997



Slide inspiration: Justin Johnson

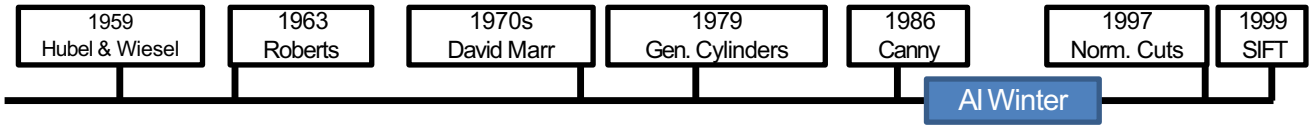
* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# Recognition via Matching, 2000s

## From 50's to nowadays


Image is public domain

Image is public domain




SIFT, David Lowe, 1999

Slide inspiration: Justin Johnson

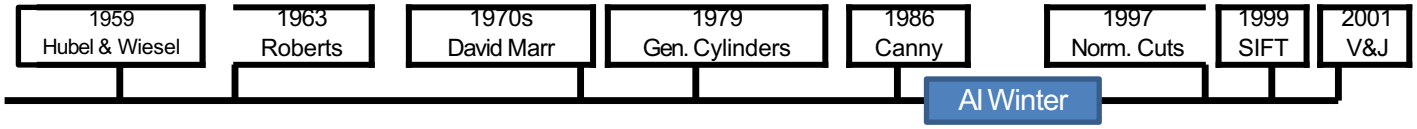
* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# Face Detection, 2000s

## From 50's to nowadays

Viola and Jones, 2001

One of the first successful applications of machine learning to vision

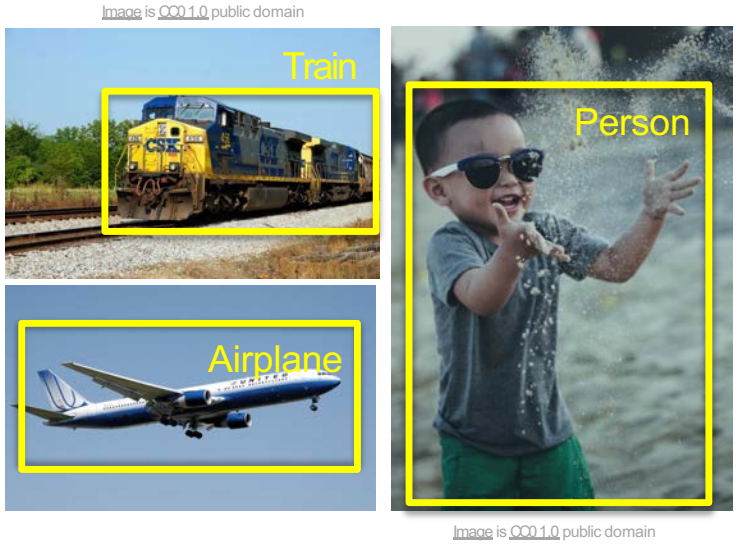


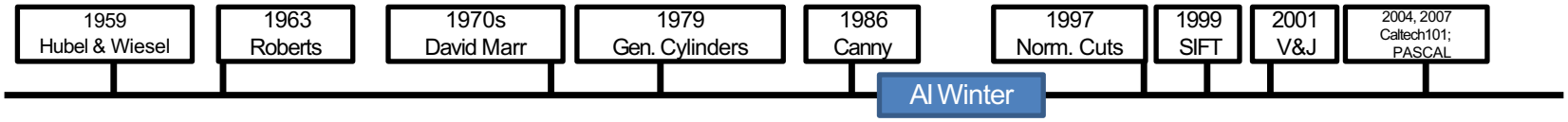

Slide inspiration: Justin Johnson

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# PASCAL Visual Object Challenge, late 2000s

## From 50's to nowadays

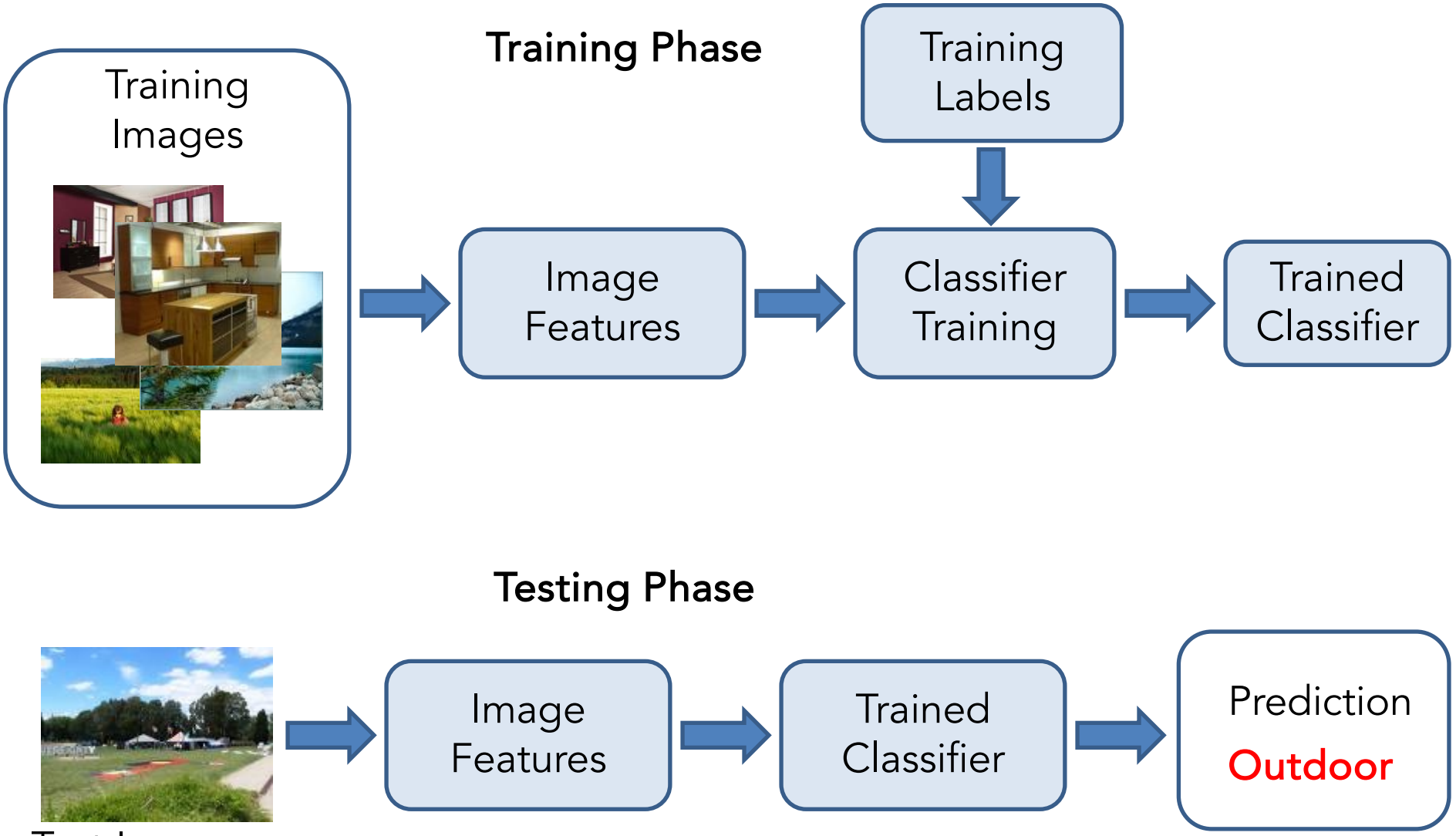




Slide inspiration: Justin Johnson

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

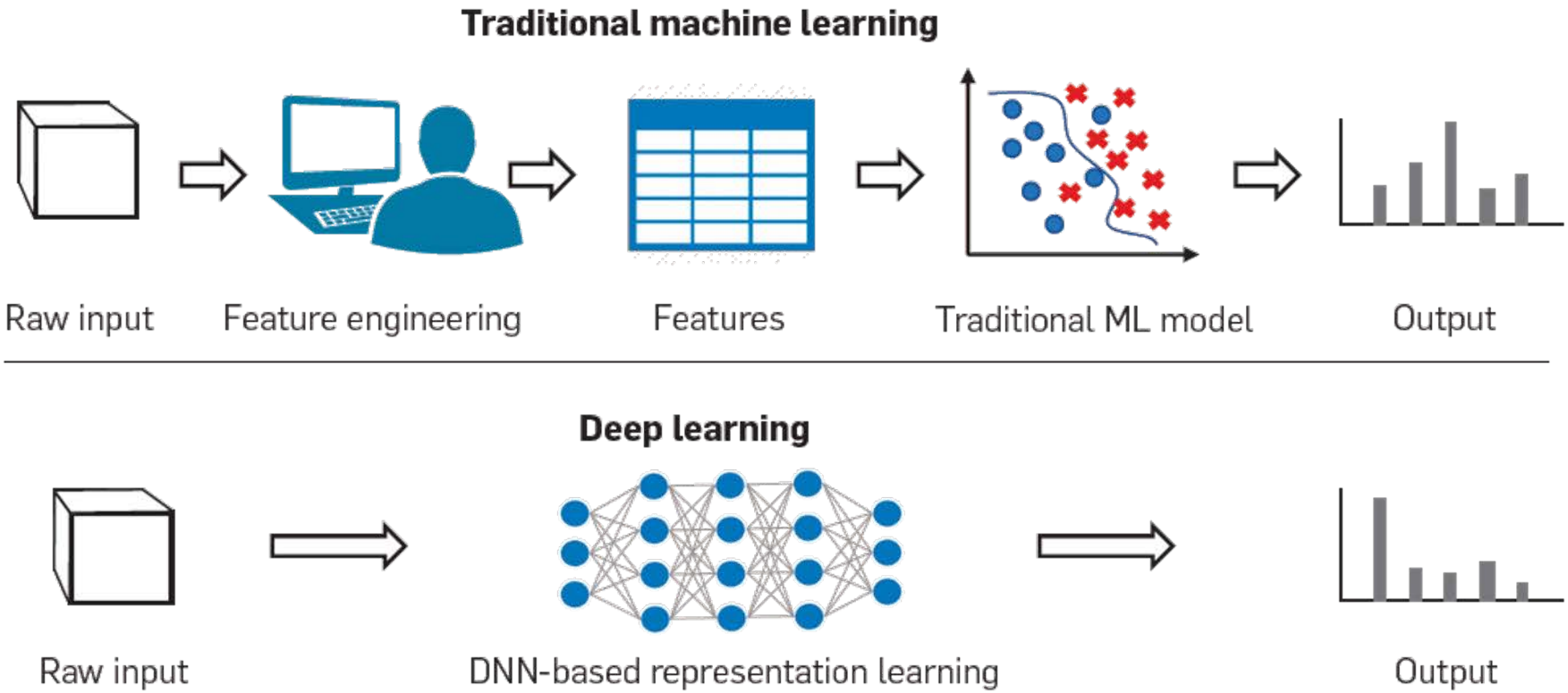
# Traditional pipeline – Learning framework

From 50's to nowadays



# Traditional pipeline vs Deep Learning

## From 50's to nowadays



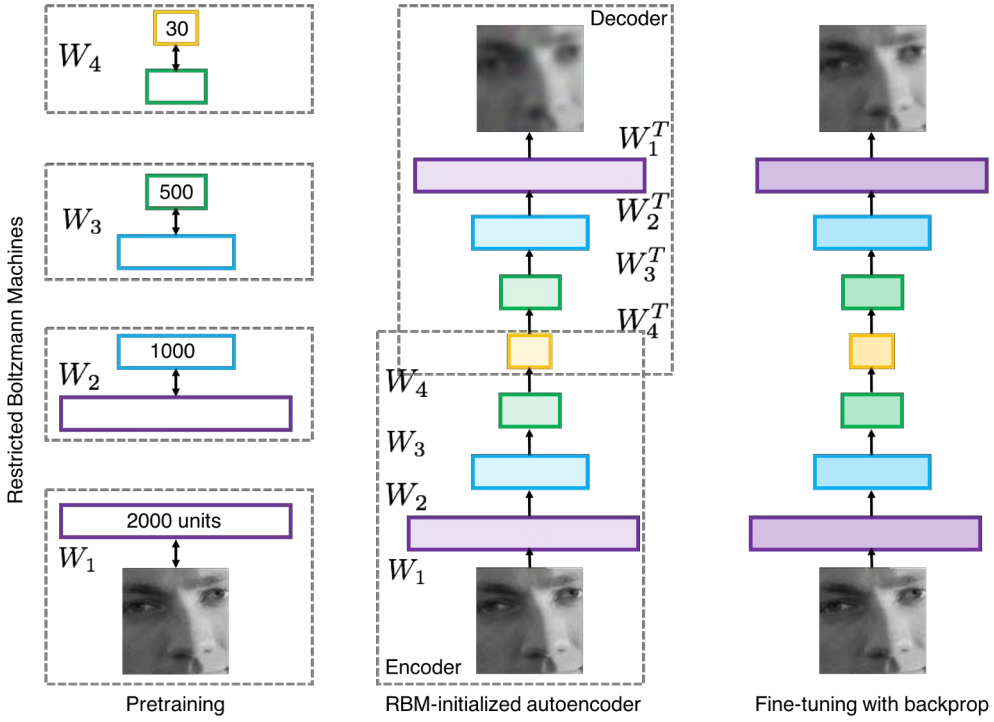
# Deep Learning (first attempts), late 2000s

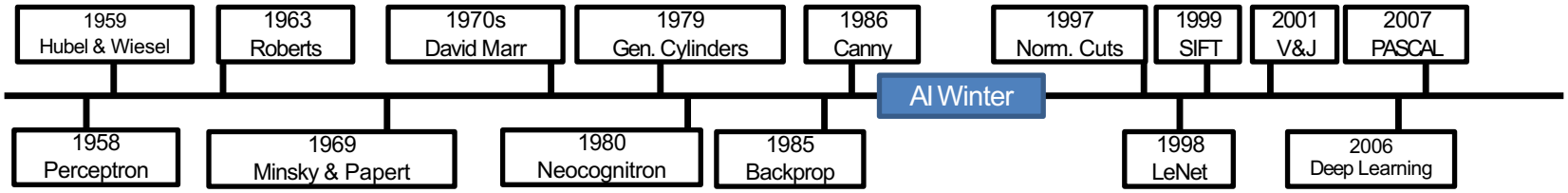
## From 50's to nowadays

- People tried to train neural networks that were deeper and deeper
- Not a mainstream research topic at this time

No good dataset to work on

Hinton and Salakhutdinov, 2006
Bengio et al, 2007
Lee et al, 2009
Glorot and Bengio, 2010




Slide inspiration: Justin Johnson

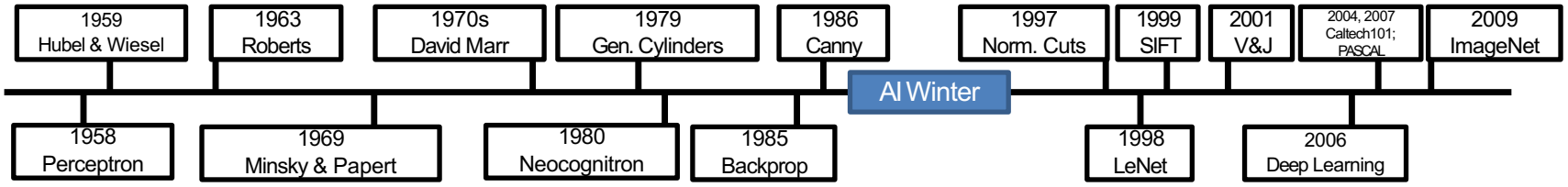
* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:
1,000 object classes
1,431,167 images

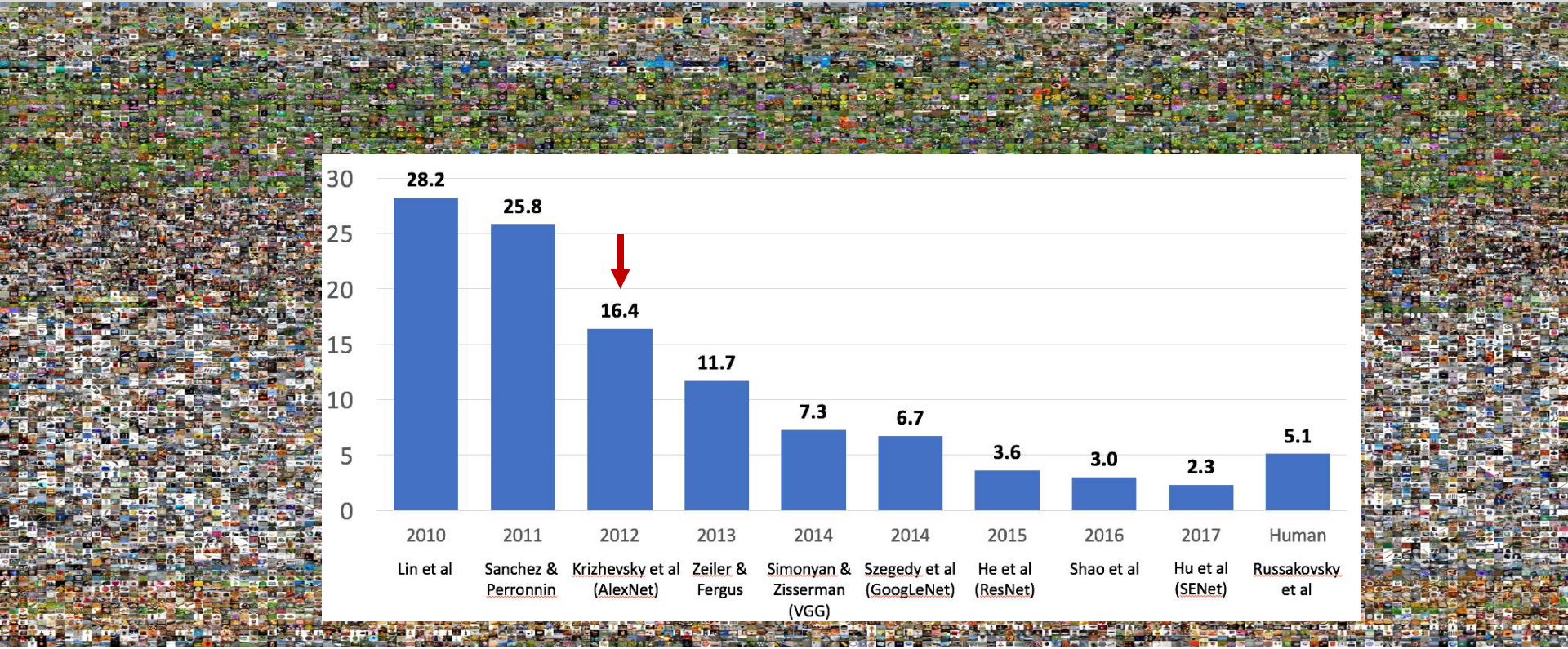
Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle

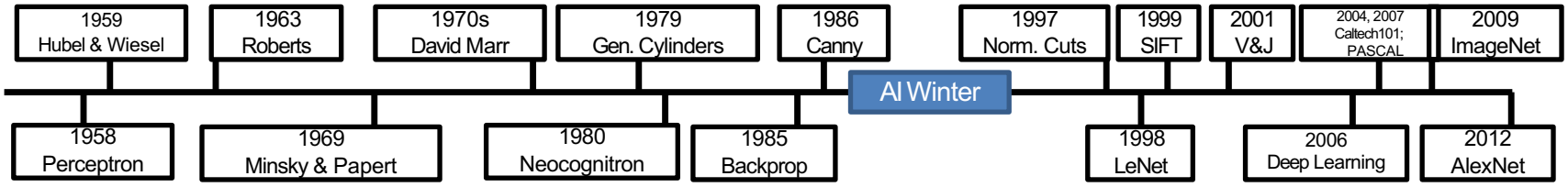
Deng et al, 2009
Russakovsky et al. IJCV 2015




* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# IMAGENET Large Scale Visual Recognition Challenge




AlexNet, 2012

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# Deep Learning is Everywhere (from 2012)

## From 50's to nowadays

### Image Classification


### Image Retrieval

Figures copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

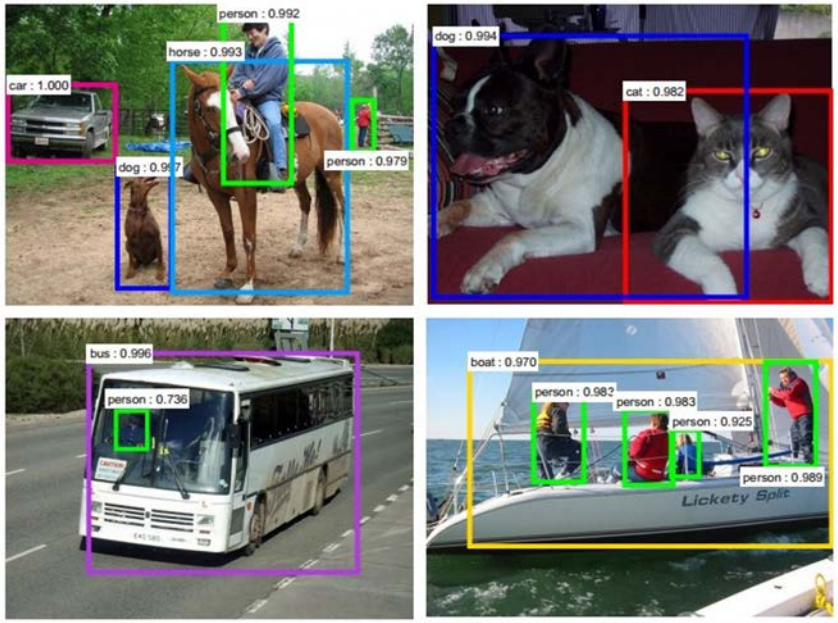
Slide inspiration: Justin Johnson

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# Deep Learning is Everywhere (from 2012)

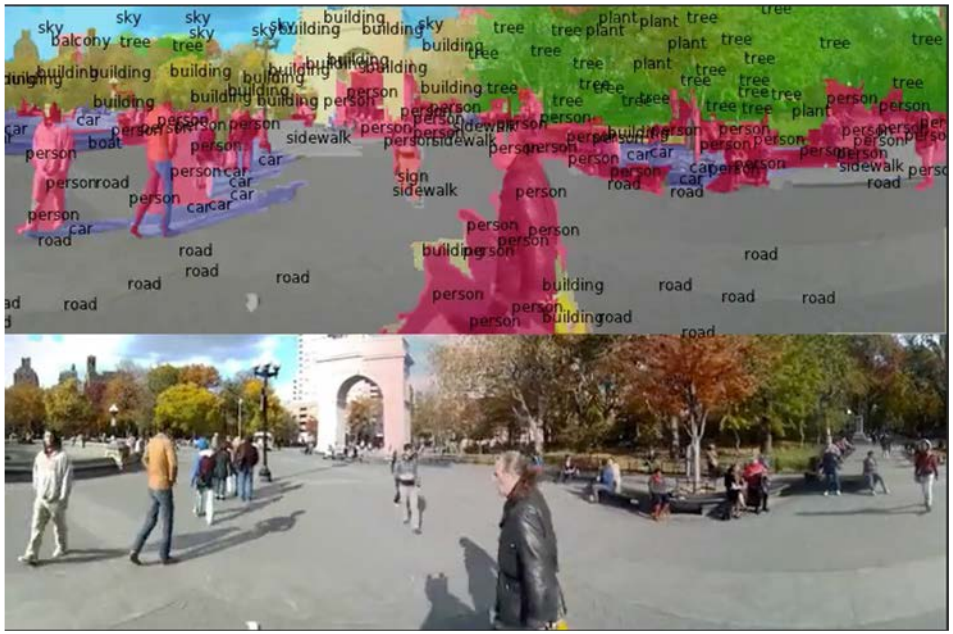
## From 50's to nowadays

Object Detection




Ren, He, Girshick, and Sun, 2015

Image Segmentation



Fabaret et al, 2012

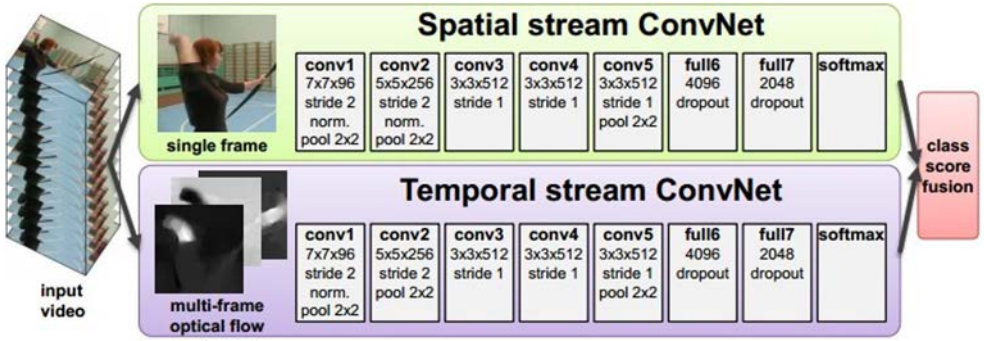
Slide inspiration: Justin Johnson

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# Deep Learning is Everywhere (from 2012)

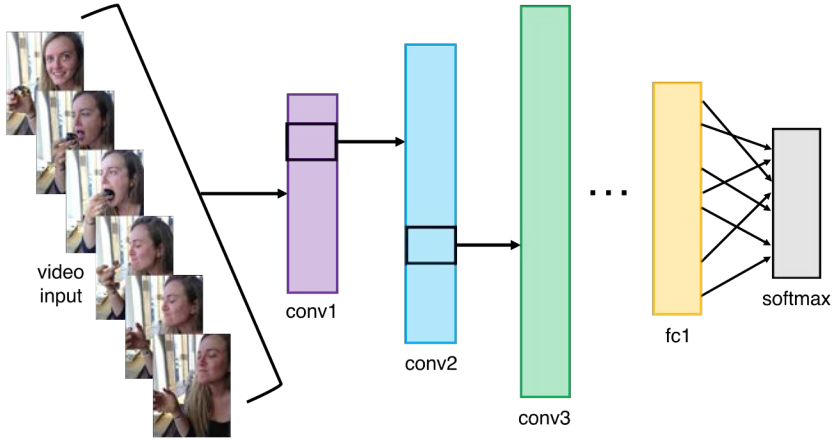
## From 50's to nowadays

Video Classification




Simonyan et al, 2014

Activity Recognition




Slide inspiration: Justin Johnson

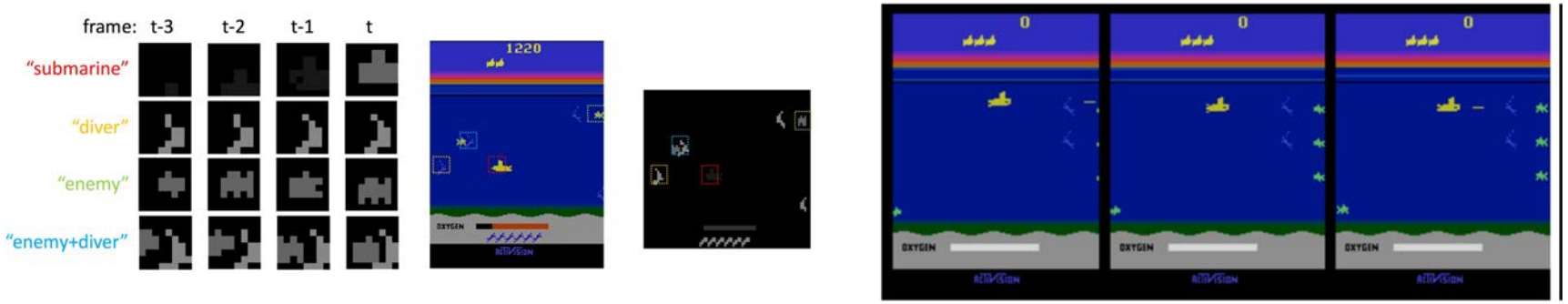
* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# Deep Learning is Everywhere (from 2012)

## From 50's to nowadays

Pose Recognition (Toshev and Szegedy, 2014)

Playing Atari games (Guo et al, 2014)



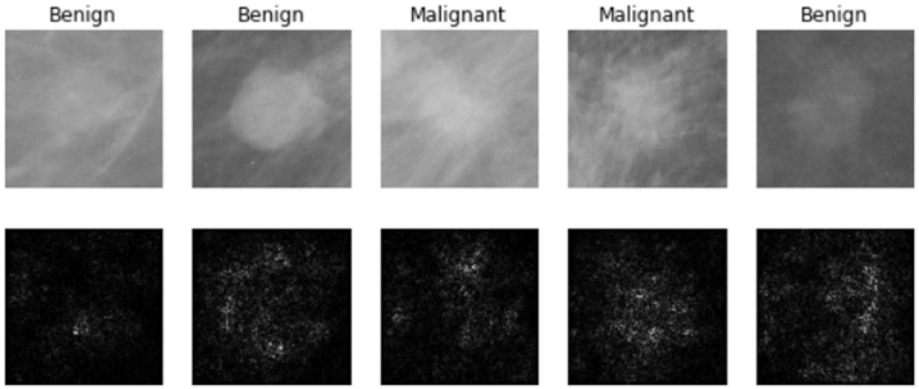

Slide inspiration: Justin Johnson

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# Deep Learning is Everywhere (from 2012)

## From 50's to nowadays

Medical Imaging




Levy et al, 2016 Figure reproduced with permission

Galaxy Classification

Dieleman et al, 2014

From left to right: public domain by NASA, usage permitted by ESA/Hubble, public domain by NASA, and public domain.

Whale recognition

Kaggle Challenge

This image by Christin Khan is in the public domain and originally came from the U.S. NOAA.

Slide inspiration: Justin Johnson

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# Deep Learning is Everywhere (from 2012)

## From 50's to nowadays

Image Captioning

Vinyals et al, 2015
Karpathy and Fei-Fei, 2015

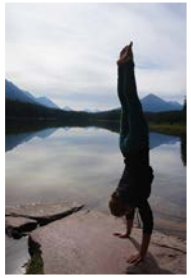
*A white teddy bear sitting in the grass*

*A man in a baseball uniform throwing a ball*

*A woman is holding a cat in her hand*

*A man riding a wave on top of a surfboard*

*A cat sitting on a suitcase on the floor*



*A woman standing on a beach holding a surfboard*



Captions generated by Justin Johnson using Neuraltalk2

Slide inspiration: Justin Johnson

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# 2012 to Present: Deep Learning is Everywhere

Slide inspiration: Justin Johnson

Karras et al, "Progressive Growing of GANs for Improved Quality, Stability, and Variation", ICLR 2018

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

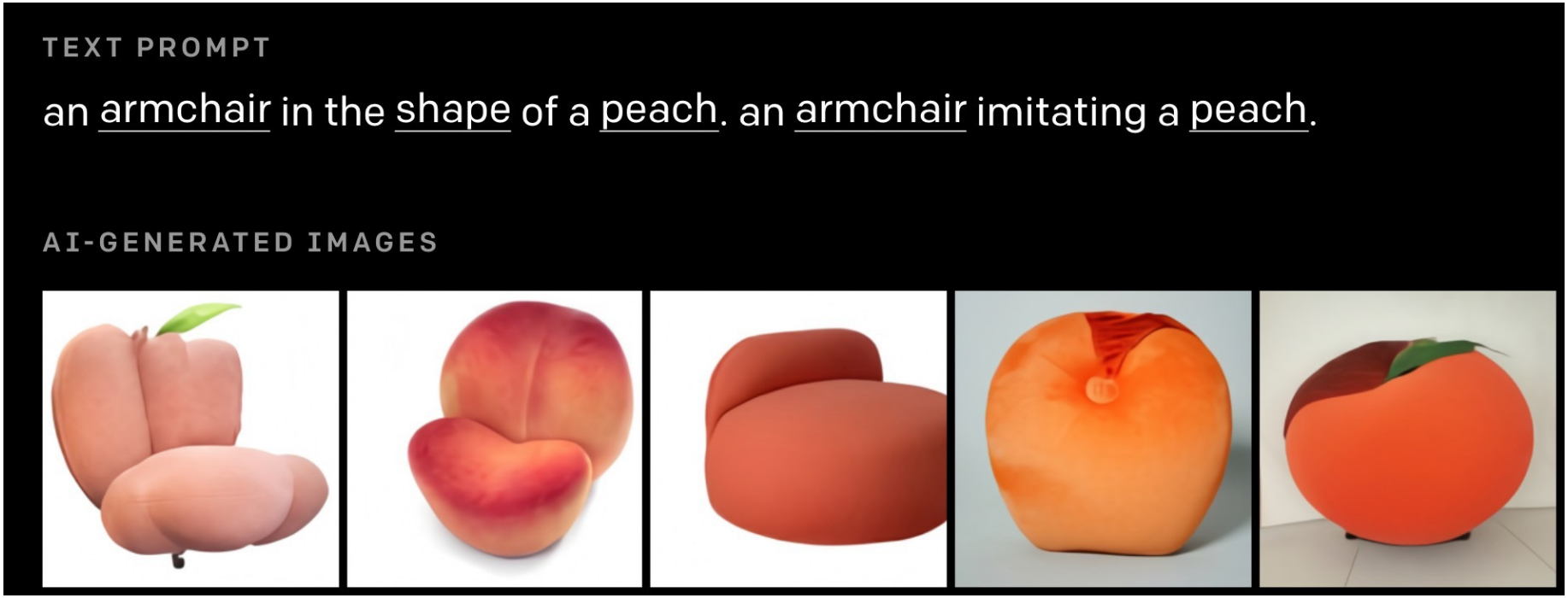
# 2012 to Present: Deep Learning is Everywhere


Ramesh et al, "DALL·E: Creating Images from Text", 2021. https://openai.com/blog/dall-e/

Slide inspiration: Justin Johnson

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# 2012 to Present: Deep Learning is Everywhere



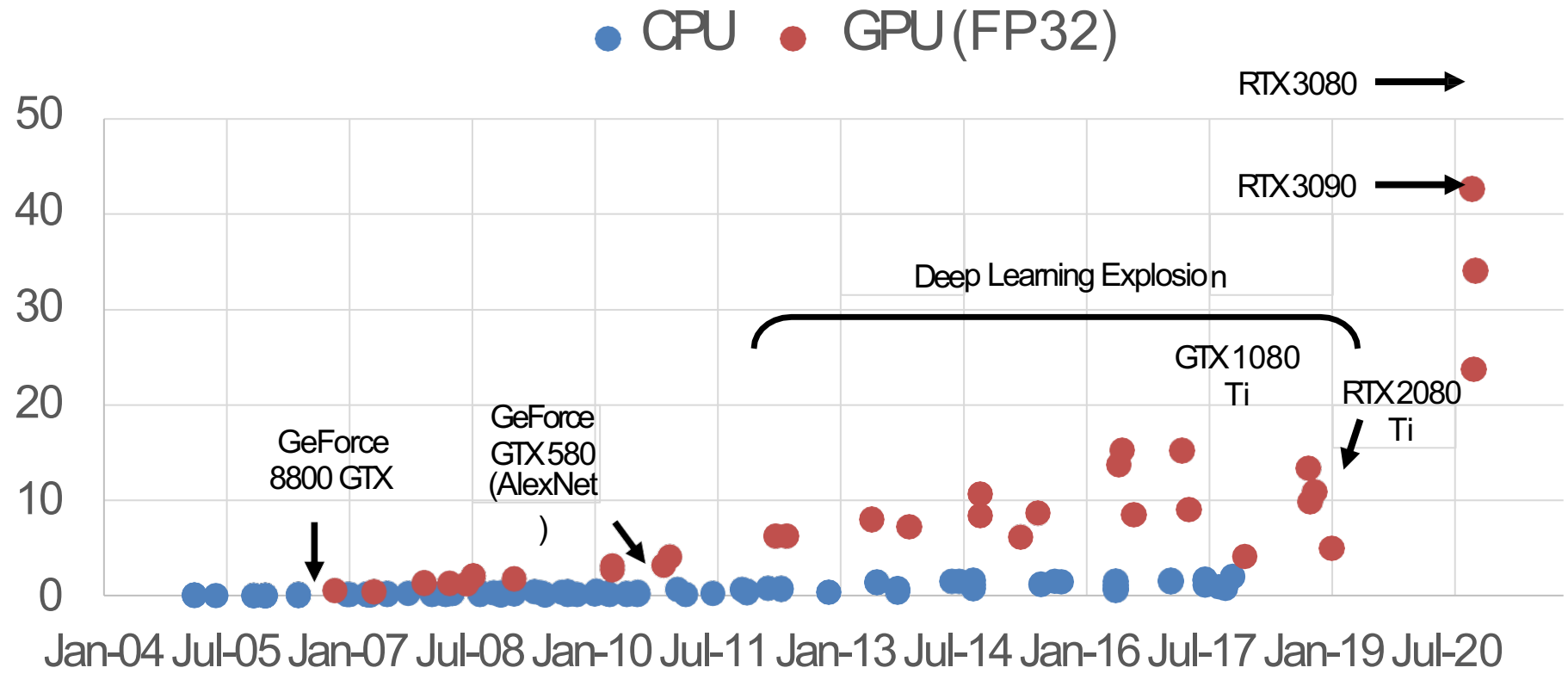

Ramesh et al, "DALL·E: Creating Images from Text", 2021. https://openai.com/blog/dall-e/

Slide inspiration: Justin Johnson

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# GFLOP per $

## From 50's to nowadays

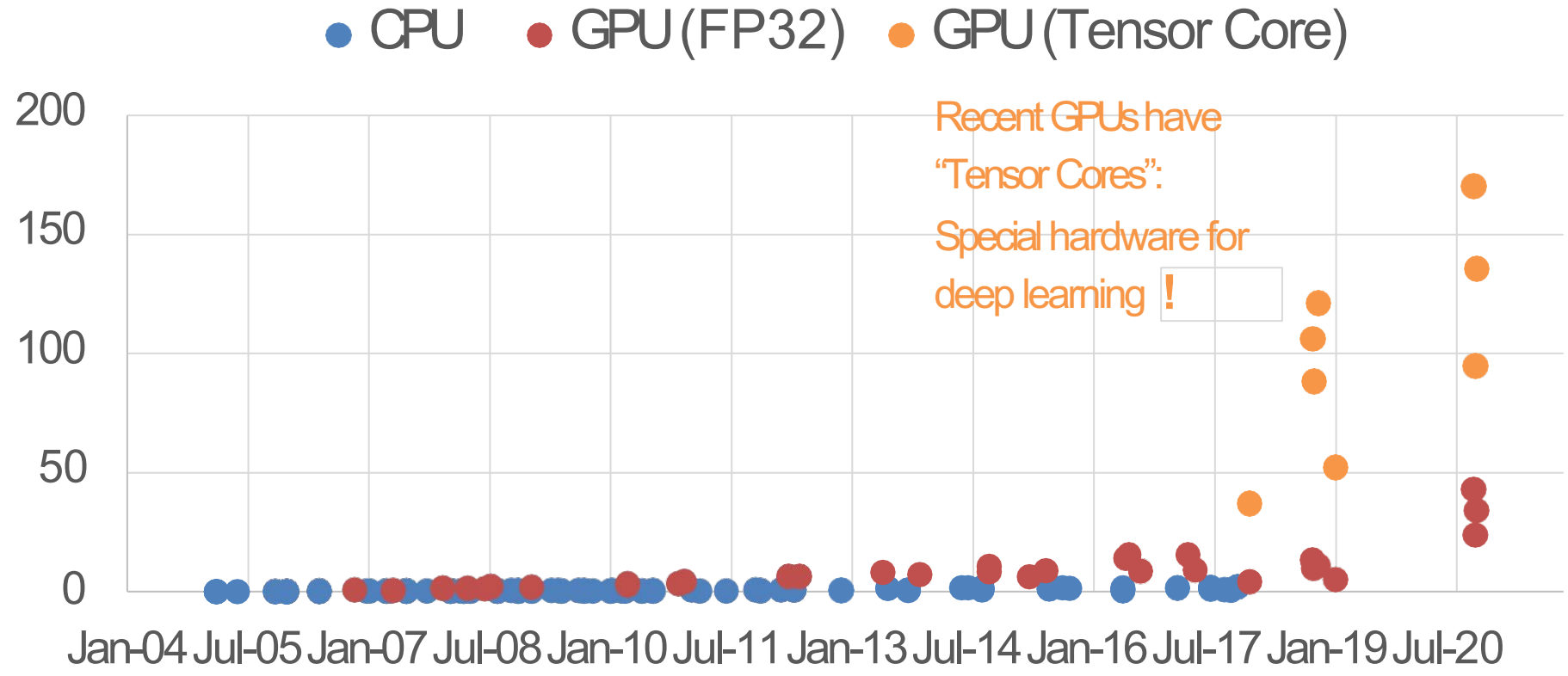



Slide inspiration: Justin Johnson

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# GFLOP per $

## From 50's to nowadays

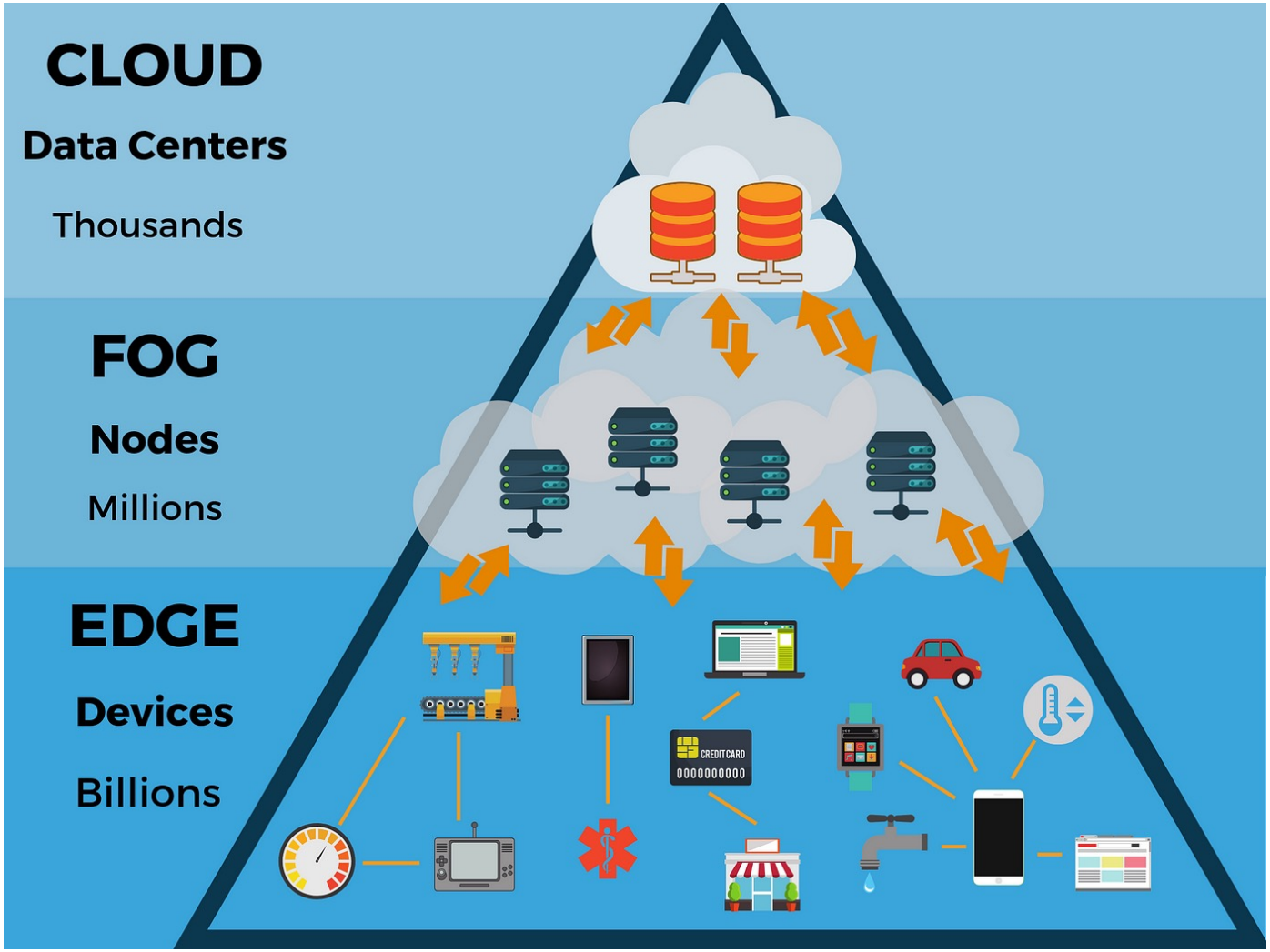



Slide inspiration: Justin Johnson

* Slide taken from Fei-Fei Li & Ruohan Gao & Yunzhu Li

# Where is better to do computations?

## Cloud-fog-edge



# Computer Vision in CT

# Computer Vision Applications in CT

## Examples

Mobile phones

Intelligent security cameras

Smart Home Appliances

Autonomous Vehicle

AR/VR headsets

Healthcare-Fitness devices

Drones

Consumer Robots

…

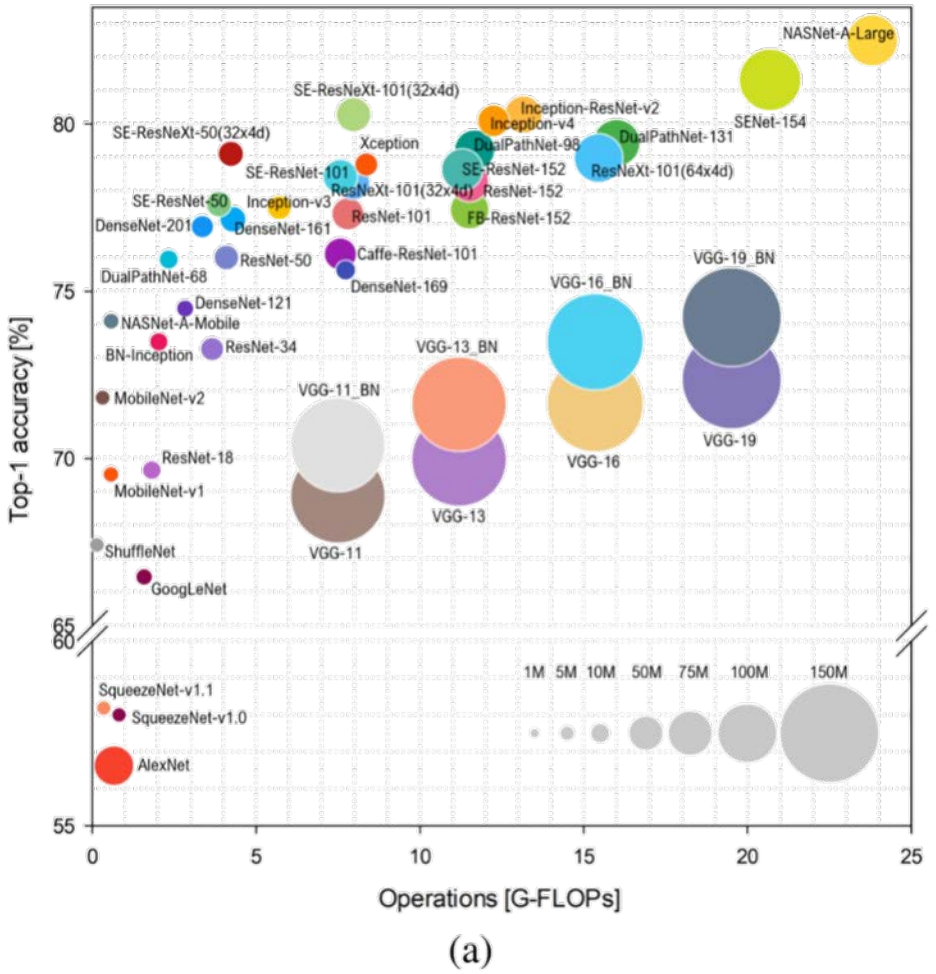
# CV in Mobile Phones

# DNNs comparison

## Computational Resources

**DL architectures**

The fundamental component of both the CONV and FC layers are the multiply-and-accumulate (MAC) operations, which can be easily parallelized using CPU and especially GPU



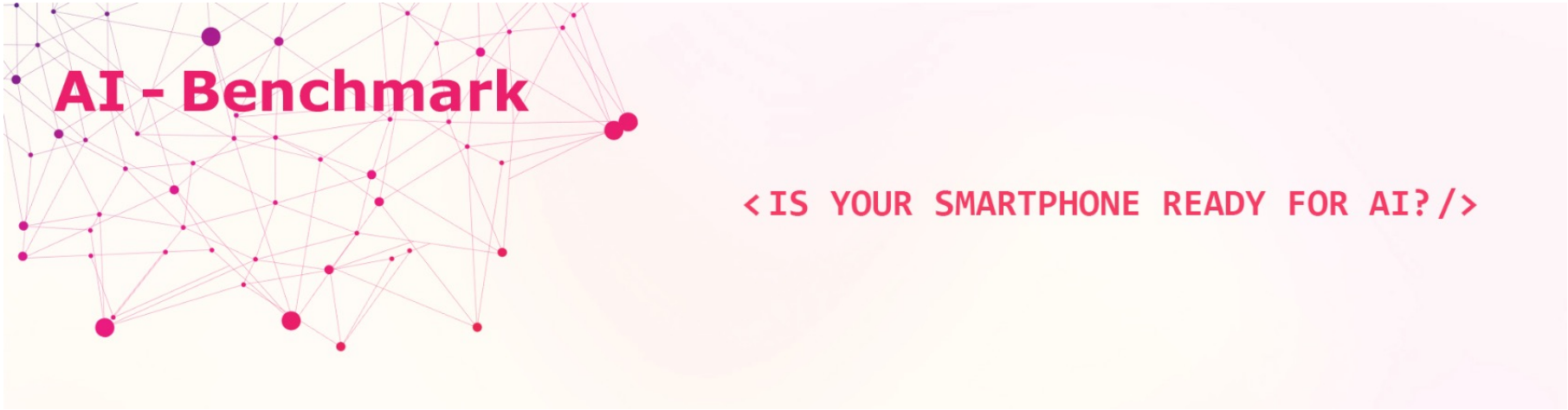
(a)

* See **additional materials**
on Bianco, S., Cadene, R., Celona, L., & Napoletano, P. (2018). Benchmark analysis of representative deep neural network architectures. IEEE Access, 6, 64270-64277.

# Is your smartphone ready for AI?

## tasks

- Is your **smartphone** capable of running the latest Deep Neural Networks to perform these **AI-based tasks**? Is it fast enough?
- Run AI Benchmark to test several key AI tasks on your phone and professionally evaluate its performance!

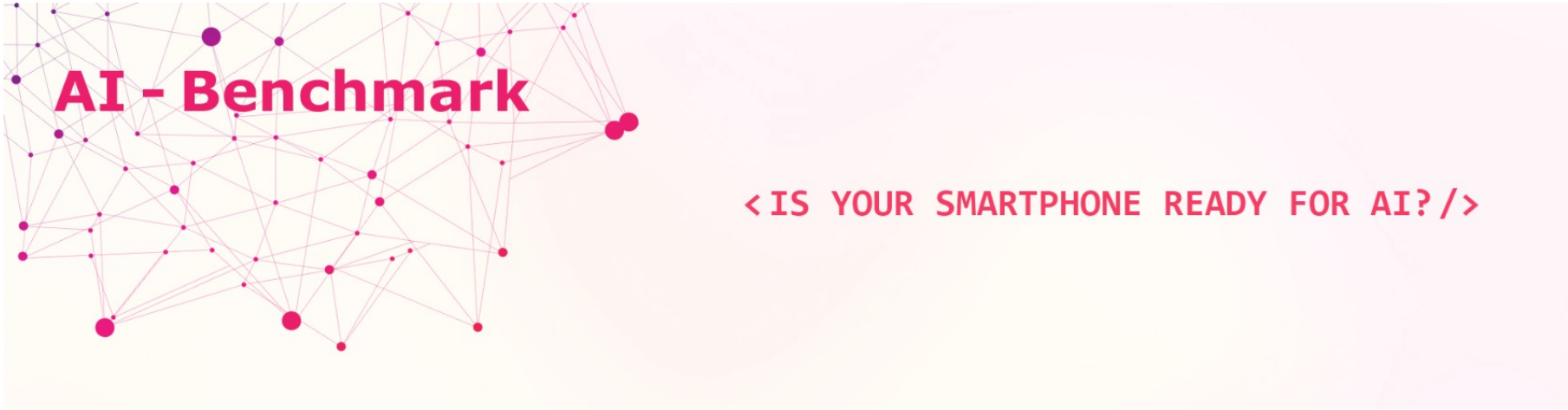



* More on https://ai-benchmark.com/

# Is your smartphone ready for AI?

## tasks

- The AI Benchmark is an **Android** application designed to check the performance and the memory limitations associated with running **AI and deep learning algorithms** on **mobile platforms**.
- It consists of several **computer vision tasks** performed by neural networks that are running directly on Android devices.




* More on https://ai-benchmark.com/

# Is your smartphone ready for AI?

## Phones – Mobile SoCs - IoT

## Performance Ranking

| Model | SoC | RAM | Year | Android | Updated | Lib | CPU-Q Score | CPU-F Score | INT8 NNAPI 1.1 | INT8 NNAPI 1.3 | INT8 Accuracy | FP16 NNAPI 1.1 | FP16 NNAPI 1.3 | FP16 Accuracy | INT8 Parallel | FP16 Parallel | INT8 NLP | FP16 NLP | INT8 Memory | FP16 Memory | AI Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Asus ROG Phone 7 | Snapdragon 8 Gen 2 | 16GB | 2023 | 13 | 7.23 | qhqh | 18.5 | 20.2 | 191 | 656 | 90.5 | 92.2 | 172 | 91.8 | 85.8 | 46.6 | 53.6 | 71 | 2800 | 2100 | 2116 |
| Asus Zenfone 10 | Snapdragon 8 Gen 2 | 8GB | 2023 | 13 | 7.23 | qhqh | 17.4 | 18.4 | 182 | 654 | 90.5 | 89.7 | 174 | 91.8 | 81.4 | 45.9 | 54.8 | 71.7 | 2900 | 2000 | 2093 |
| Xiaomi Redmi K60 Pro | Snapdragon 8 Gen 2 | 12GB | 2023 | 13 | 7.23 | qhqh | 18.8 | 20.6 | 186 | 658 | 90.5 | 89.1 | 172 | 91.8 | 85.6 | 47.7 | 54.1 | 70.7 | 2200 | 2100 | 2076 |
| Xiaomi 13 Ultra | Snapdragon 8 Gen 2 | 16GB | 2023 | 13 | 7.23 | qhqh | 17.8 | 19.3 | 186 | 656 | 90.5 | 89.6 | 173 | 91.8 | 85.6 | 45.4 | 53.3 | 70.5 | 2200 | 2100 | 2068 |
| Samsung Galaxy S23+ | Snapdragon 8 Gen 2 | 8GB | 2023 | 13 | 7.23 | qhqh | 18 | 19.4 | 184 | 652 | 90.5 | 90 | 172 | 91.8 | 83.5 | 46.4 | 54.3 | 72 | 2200 | 2100 | 2060 |
| Motorola Edge 40 Pro | Snapdragon 8 Gen 2 | 12GB | 2023 | 13 | 7.23 | qhqh | 15.8 | 17.5 | 182 | 644 | 90.5 | 88.1 | 171 | 91.8 | 80.2 | 43.3 | 53.1 | 69.9 | 2800 | 2100 | 2056 |
| Samsung Galaxy S23 Ultra | Snapdragon 8 Gen 2 | 12GB | 2023 | 13 | 7.23 | qhqh | 17.4 | 18.6 | 184 | 649 | 90.5 | 91.3 | 172 | 91.8 | 81.3 | 46.4 | 54.1 | 71.8 | 2200 | 2100 | 2053 |
| Xiaomi 13 Pro | Snapdragon 8 Gen 2 | 12GB | 2022 | 13 | 7.23 | qhqh | 17.8 | 18.9 | 184 | 651 | 90.5 | 88.6 | 171 | 91.8 | 84.6 | 45.3 | 53.5 | 69.9 | 2200 | 2100 | 2052 |
| Samsung Galaxy S23 | Snapdragon 8 Gen 2 | 8GB | 2023 | 13 | 7.23 | qhqh | 17.4 | 18.8 | 183 | 649 | 90.5 | 89.3 | 172 | 91.8 | 83.6 | 44.8 | 54.4 | 71.9 | 2200 | 2100 | 2050 |
| vivo iQOO 11 Pro | Snapdragon 8 Gen 2 | 12GB | 2022 | 13 | 7.23 | qhqh | 14.5 | 15.4 | 178 | 642 | 90.5 | 87.8 | 169 | 91.8 | 80.5 | 44.8 | 53.9 | 71 | 2800 | 2100 | 2045 |
| vivo X90 Pro+ | Snapdragon 8 Gen 2 | 12GB | 2022 | 13 | 7.23 | qhqh | 14.3 | 16 | 181 | 642 | 90.5 | 90.1 | 171 | 91.8 | 84.3 | 44.1 | 54.2 | 71 | 2200 | 2100 | 2027 |
| Xiaomi 13 | Snapdragon 8 Gen 2 | 12GB | 2022 | 13 | 7.23 | qhqh | 16.4 | 17.5 | 182 | 648 | 90.5 | 87.5 | 170 | 91.8 | 80.6 | 42.7 | 53.4 | 69.5 | 2200 | 2100 | 2026 |
| Sharp Aquos R8 Pro | Snapdragon 8 Gen 2 | 12GB | 2023 | 13 | 7.23 | qhqh | 15.2 | 16.9 | 181 | 645 | 90.5 | 87.8 | 170 | 91.8 | 72.7 | 46.6 | 53.1 | 69.4 | 2200 | 2100 | 2010 |
| ZTE nubia Red Magic 8 Pro | Snapdragon 8 Gen 2 | 12GB | 2022 | 13 | 7.23 | qhqh | 13.1 | 15.4 | 182 | 637 | 90.5 | 88.8 | 169 | 91.8 | 81.2 | 44.8 | 53.7 | 70.1 | 2200 | 2100 | 2008 |

* More on https://ai-benchmark.com/

# Is your smartphone ready for AI?

## Phones – Mobile SoCs - IoT

## Performance Ranking **DESKTOP**

| Model | TF Version | Cores | Frequency, GHz | Acceleration | Platform | RAM, GB | Year | Inference Score | Training Score | AI-Score |
|---|---|---|---|---|---|---|---|---|---|---|
| Tesla V100 SXM2 32Gb | 2.1.0 | 5120 (CUDA) | 1.29 / 1.53 | CUDA 10.1 | Debian 10 | 32 | 2018 | 17761 | 18030 | 35791 |
| Tesla V100 SXM2 16Gb | 2.1.0 | 5120 (CUDA) | 1.31 / 1.53 | CUDA 10.1 | Red Hat 7.5 | 16 | 2017 | 17251 | 17836 | 35086 |
| Tesla V100 PCIE 32Gb | 2.1.0 | 5120 (CUDA) | 1.23 / 1.38 | CUDA 10.1 | Debian 10 | 32 | 2018 | 16530 | 17865 | 34394 |
| Tesla V100 PCIE 16Gb | 2.1.0 | 5120 (CUDA) | 1.25 / 1.38 | CUDA 10.1 | Red Hat 7.5 | 16 | 2017 | 16511 | 17837 | 34347 |
| NVIDIA Quadro GV100 | 1.14.0 | 5120 (CUDA) | 1.13 / 1.63 | CUDA 10 | Debian 10 | 32 | 2018 | 16748 | 17132 | 33880 |
| NVIDIA TITAN V | 2.1.0 | 5120 (CUDA) | 1.20 / 1.46 | CUDA 10.1 | Ubuntu 18.04 | 12 | 2017 | 16192 | 17215 | 33406 |
| NVIDIA TITAN RTX | 2.1.0 | 4608 (CUDA) | 1.35 / 1.77 | CUDA 10.1 | Ubuntu 18.04 | 24 | 2018 | 16084 | 17255 | 33339 |
| GeForce RTX 2080 Ti | 2.1.0 | 4352 (CUDA) | 1.35 / 1.55 | CUDA 10 | Debian 10 | 11 | 2018 | 16042 | 16828 | 32870 |
| NVIDIA Quadro RTX 8000 | 2.1.0 | 4608 (CUDA) | 1.40 / 1.77 | CUDA 10.1 | Debian 10 | 48 | 2018 | 13014 | 14637 | 27651 |
| NVIDIA Quadro GP100 | 2.0.0 | 3584 (CUDA) | 1.30 / 1.44 | CUDA 10 | Red Hat 7.4 | 16 | 2016 | 12264 | 13436 | 25700 |
| NVIDIA TITAN Xp | 2.1.0 | 3840 (CUDA) | 1.41 / 1.58 | CUDA 10.2 | Debian 10 | 12 | 2017 | 11948 | 12922 | 24870 |
| GeForce GTX 1080 Ti | 2.1.0 | 3584 (CUDA) | 1.58 / 1.60 | CUDA 10.2 | Debian 10 | 11 | 2017 | 11914 | 12473 | 24386 |

* More on https://ai-benchmark.com/

# CV in Mobile Phones

## tasks

- **Image Recognition**, which refers to the ability to find class of objects within the images.
- **Content-Based Image Retrieval** refers to the ability to retrieve images that are similar to the query image
- **Face detection** answers the question “Is this a face?” It is a technique that identifies or locates human faces in digital images.
- **Face recognition**, answers the question “Whose face is this?” It is the task of identifying an already detected object as a known or unknown face.

# CV in Mobile Phones – Camera roll

tasks

Mobile photos are managed thanks to a photo management and editing application


Amazon

Apple

Google

Main CV Features: Semantic Classes, Object Saliency, Person Identification

# CV in Mobile Phones – Camera roll

tasks

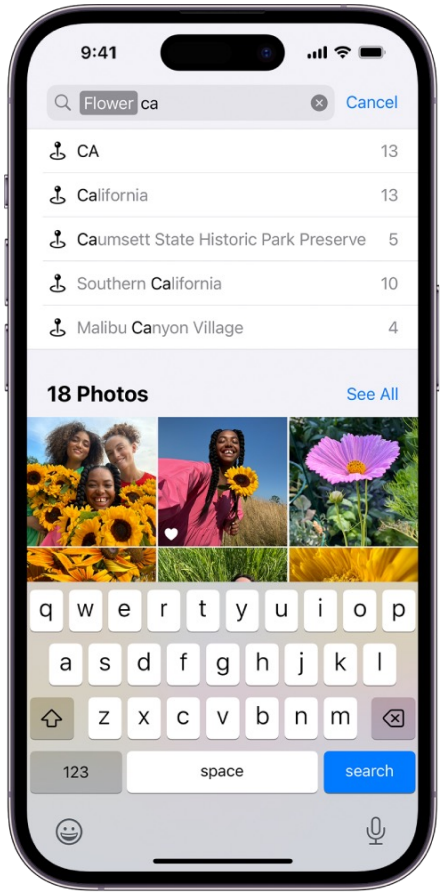
How does it work?

# CV in Mobile Phones – Camera roll

tasks

## Search for Photos on Apple camera roll:

- Date (month or year)
- Place (city or state)
- Business names (museums, for example)
- Category (beach or sunset, for example)
- Events (sports games or concerts, for example)
- A person identified in your People album
- Text (an email address or phone number, for example)
- Caption
- The person who added the photo to the library




* See **additional materials** on https://machinelearning.apple.com/research/recognizing-people-photos

# CV in Mobile Phones – Camera roll

## EXIF format

EXIF stands for "Exchangeable Image File Format", the definition first given by Japan Camera Industry Association (JCIA) in 1985.

The standard is managed by Japan Electronics and Information Technology Industries Association (JEITA) as of today. EXIF is a standard for the specifications of image and sound formats mainly used by digital cameras and scanners

| Tag Name | Content |
|---|---|
| | ---- **EXIF** ---- |
| Make | Canon |
| Model | Canon EOS 350D DIGITAL |
| Orientation | Horizontal (normal) |
| XResolution | 72 |
| YResolution | 72 |
| ResolutionUnit | inches |
| ModifyDate | 2011:06:27 12:26:57 |
| YCbCrPositioning | Co-sited |
| ExposureTime | 1/400 |
| FNumber | 14.0 |
| ExposureProgram | Program AE |
| ISO | 400 |
| ExifVersion | 0221 |
| DateTimeOriginal | 2011:06:27 12:26:57 |
| CreateDate | 2011:06:27 12:26:57 |
| ComponentsConfiguration | Y, Cb, Cr, - |
| ShutterSpeedValue | 1/400 |
| ApertureValue | 14.0 |
| ExposureCompensation | 0 |
| MeteringMode | Multi-segment |
| Flash | Off, Did not fire |
| FocalLength | 55.0 mm |
| ColorSpace | sRGB |
| ExifImageWidth | 2496 |
| ExifImageHeight | 1664 |
| FocalPlaneResolutionUnit | inches |
| CustomRendered | Normal |
| ExposureMode | Auto |
| WhiteBalance | Auto |
| SceneCaptureType | Standard |
| Compression | JPEG (old-style) |
| XResolution | 72 |
| YResolution | 72 |
| ResolutionUnit | inches |

* See **additional materials** on https://docs.fileformat.com/image/exif/

# CV in Mobile Phones – Camera roll

## EXIF format

EXIF standard includes the **tagging** and **metadata** information with the image file.

Metadata can contain information like: *camera model*, *shutter speed*, *Date* and *time*, *aperture*, *manufacturer*, *exposure time*, X *resolution*, Y *resolution* etc.

| Tag Name | Content |
|---|---|
| | ---- **EXIF** ---- |
| Make | Canon |
| Model | Canon EOS 350D DIGITAL |
| Orientation | Horizontal (normal) |
| XResolution | 72 |
| YResolution | 72 |
| ResolutionUnit | inches |
| ModifyDate | 2011:06:27 12:26:57 |
| YCbCrPositioning | Co-sited |
| ExposureTime | 1/400 |
| FNumber | 14.0 |
| ExposureProgram | Program AE |
| ISO | 400 |
| ExifVersion | 0221 |
| DateTimeOriginal | 2011:06:27 12:26:57 |
| CreateDate | 2011:06:27 12:26:57 |
| ComponentsConfiguration | Y, Cb, Cr, - |
| ShutterSpeedValue | 1/400 |
| ApertureValue | 14.0 |
| ExposureCompensation | 0 |
| MeteringMode | Multi-segment |
| Flash | Off, Did not fire |
| FocalLength | 55.0 mm |
| ColorSpace | sRGB |
| ExifImageWidth | 2496 |
| ExifImageHeight | 1664 |
| FocalPlaneResolutionUnit | inches |
| CustomRendered | Normal |
| ExposureMode | Auto |
| WhiteBalance | Auto |
| SceneCaptureType | Standard |
| Compression | JPEG (old-style) |
| XResolution | 72 |
| YResolution | 72 |
| ResolutionUnit | inches |

* See **additional materials** on https://docs.fileformat.com/image/exif/

# CV in Mobile Phones – Camera roll

tasks

## Apple photos People Recognition (on device)

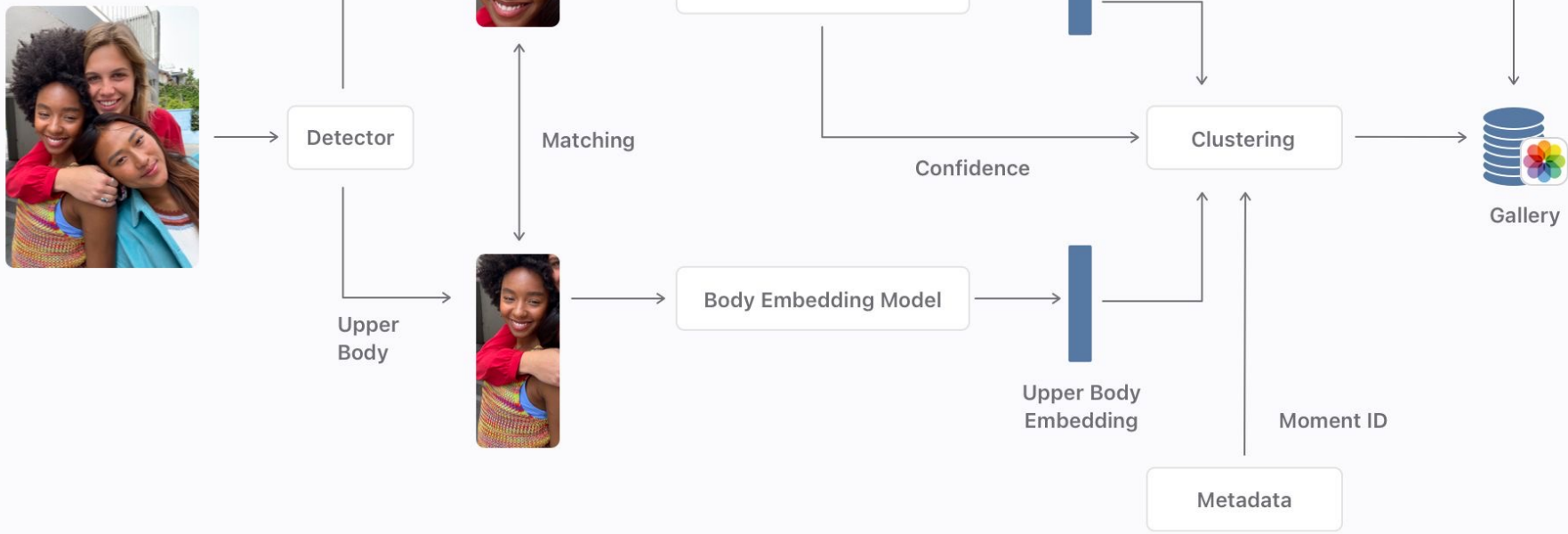

A B C D

* See **additional materials** on https://machinelearning.apple.com/research/recognizing-people-photos

# CV in Mobile Phones – Camera roll

tasks

## Apple photos People Recognition (on device)



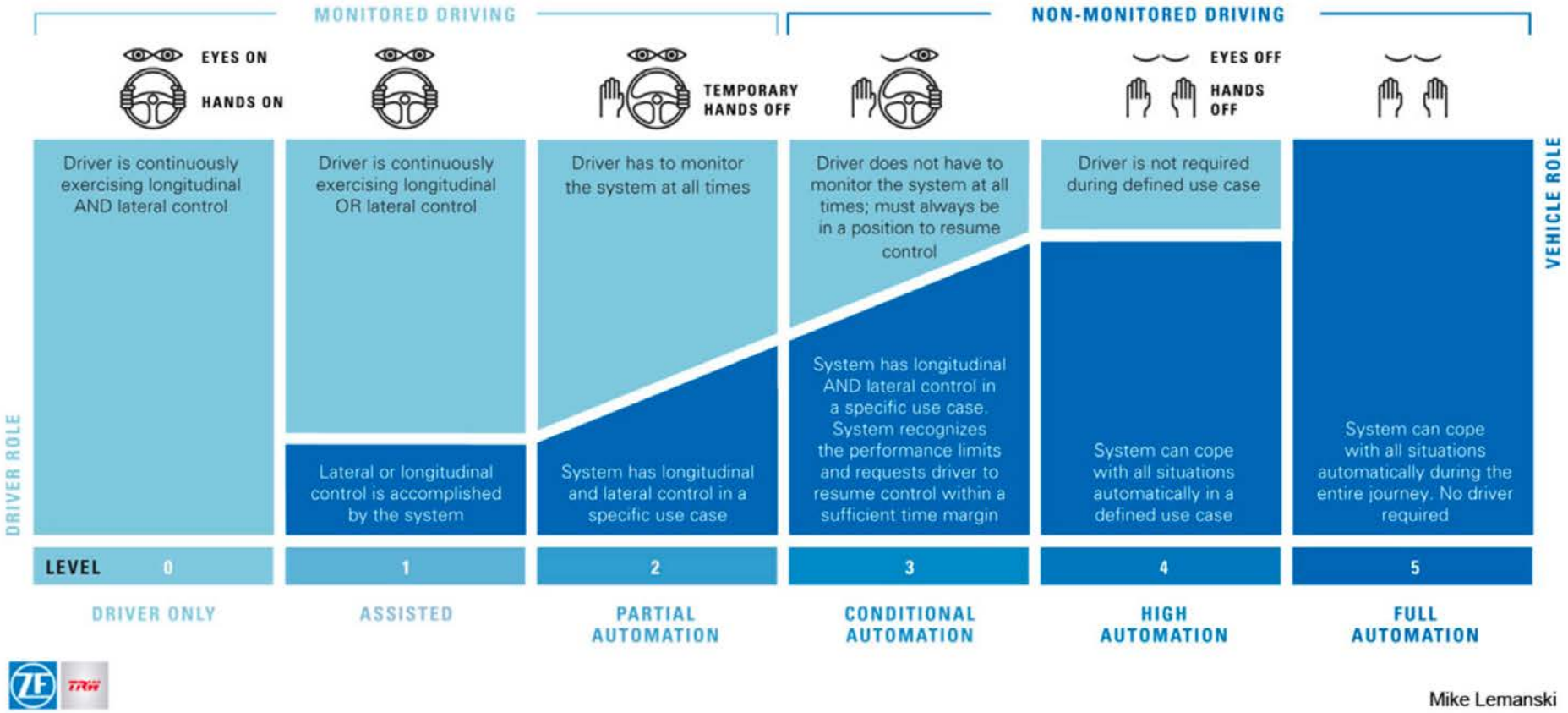

* See **additional materials** on https://machinelearning.apple.com/research/recognizing-people-photos

# CV in Autonomous Vechicle

# CV in Autonomous Vehicle

tasks

## **SAE** level of Autonomy

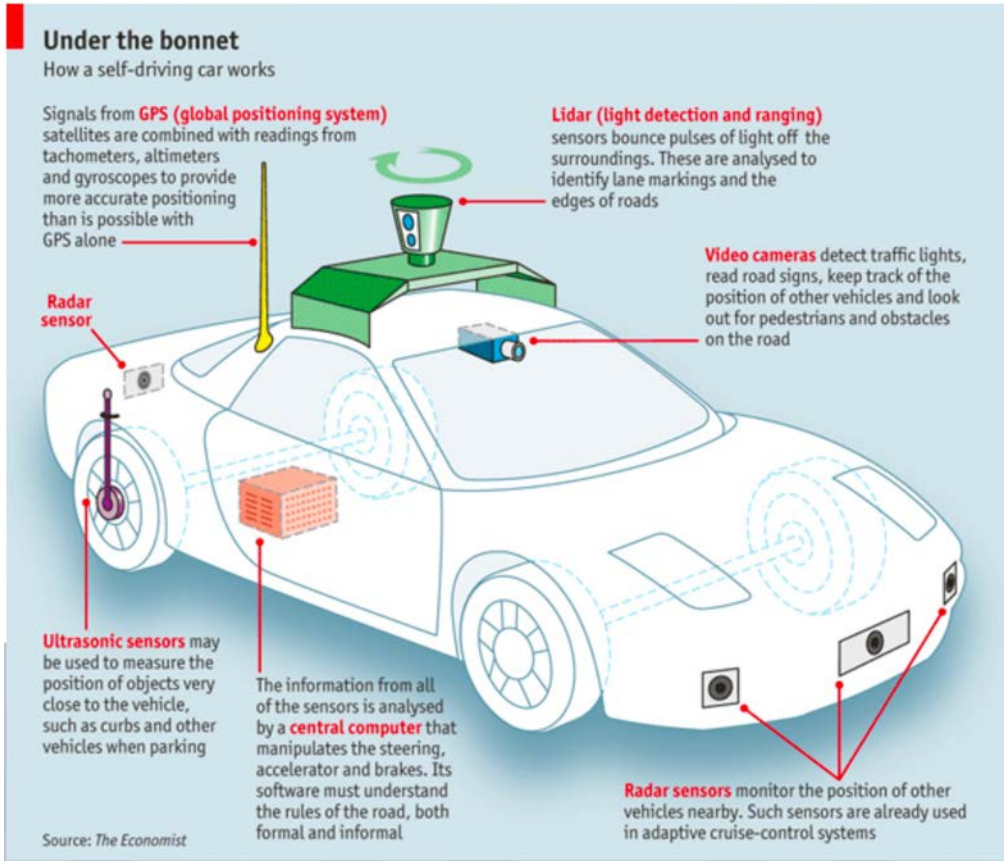



SAE - Society of Automotive Engineers

# Basic Physical Ecosystem of an AV

tasks

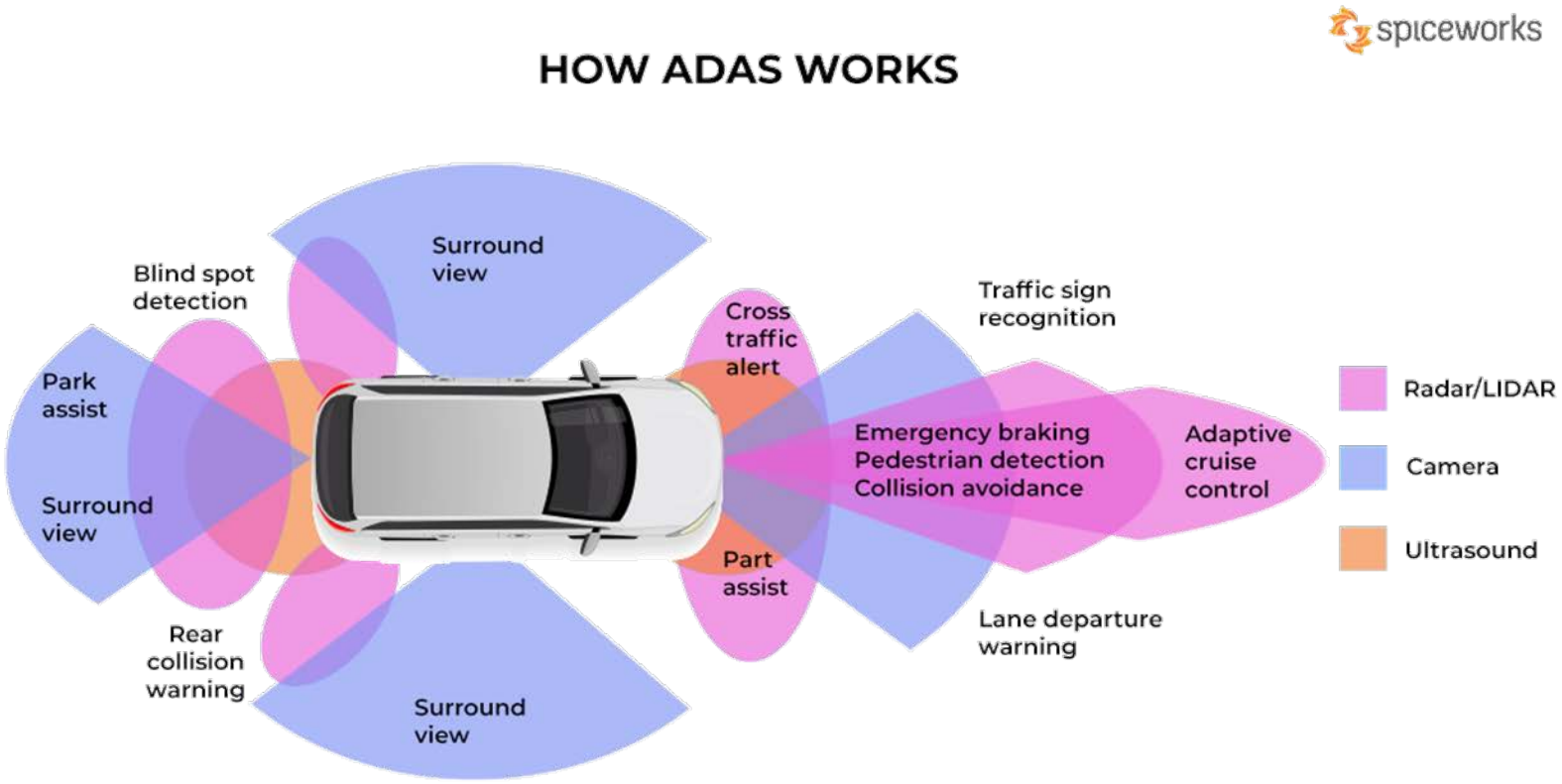
- Global Positioning System (GPS)
- Light Detection and Ranging (LIDAR)
- Cameras (Video)
- Ultrasonic Sensors
- Central Computer
- Radar Sensors
- Dedicated Short-Range Communications-Based Receiver (not pictured)



# Role of the Cameras in AV

## tasks

Cameras provide real-time obstacle detection to facilitate lane departure and track roadway information (like road signs).



# Role of the Cameras in AV

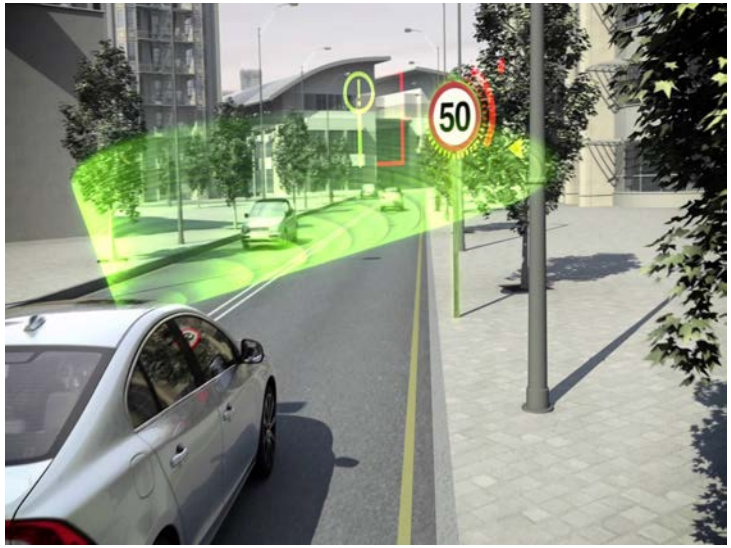
## tasks

Driver Assistance: Enhances vehicle systems for safety and improved driving when the driver is in control. Technology includes blind-spot detection, pedestrian detection, lane-departure warnings, intelligent braking, traffic-sign recognition, automatic braking, and adaptive cruise control.

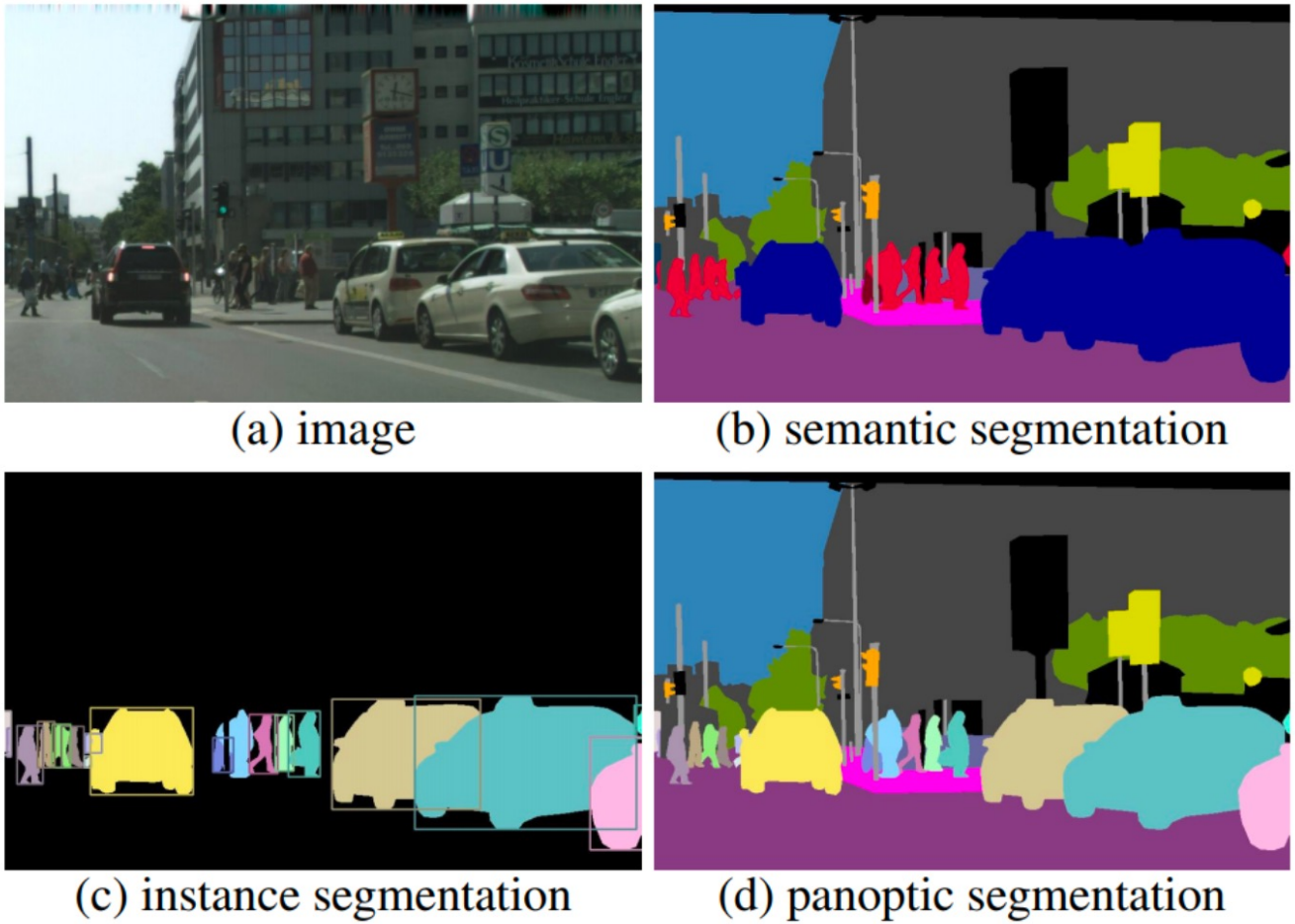
# Role of the Cameras in AV

## tasks

Driver Assistance: Enhances vehicle systems for safety and improved driving when the driver is in control. Technology includes blind-spot detection, pedestrian detection, lane-departure warnings, intelligent braking, traffic-sign recognition, automatic braking, and adaptive cruise control.



## Segmentation



(a) image (b) semantic segmentation

(c) instance segmentation (d) panoptic segmentation

# CV in Intelligent Security Cameras

# Security Camera

tasks

Surveillance is the **monitoring** of **behavior**, many activities, or information for the purpose of information gathering, influencing, managing or directing. This can include observation from a distance by means of **electronic equipment**, such as closed-circuit television adaptive cruise control.

# Security Camera

## tasks

Security Cameras are used **indoor** and **outdoor** to monitoring, control the presence of verified people or object, unwanted people, anomalous events, etc.

### Analytics

Motion Detection

No Motion Detection

Intrusion Detection

Face Detection


Parking Management

Missing Object Detection

License Plate Recognition

Trip Wire


Loitering

Crowd Management

### Notifications

Calling from Mobile App


Email with Snapshot

SMS

Video Pop-up

Audio Alarm & Buzzer

# Face Detection

# Face detection

## Challenges

Face detection  is an essential **early step** for tasks such as **face recognition**, **facial attribute classification**, **face editing**, and **face tracking**, and its performance has a direct impact on the effectiveness of those tasks.

# Face detection

## Challenges

- **Face detection** in the wild remains an **open challenge**: variations in poses, facial expressions, scale, illumination, image distortion, face occlusion, etc.

- Different from generic object detection, face detection features smaller variations in the aspect ratio, but much **larger variations in scale** (from several pixels to thousand pixels)


* See **additional materials** on Minaee, S., Luo, P., Lin, Z., & Bowyer, K. (2021). Going deeper into face detection: A survey. *arXiv preprint arXiv:2103.14983*.

# Face detection

## Challenges

- Early face detection methods were made of **hand-crafted features** extracted from the image (or from sliding windows on the image) and a classifier (or ensemble of classifiers) to detect likely face regions.

- Two classical works for face detection are the Haar Cascades classifier (also known as **Viola-Jones**) and the Histogram of Oriented Gradients (HOG) followed by SVM.


* See **additional materials** on Paul Viola and Michael J. Jones. Robust real-time face detection. International Journal of Computer Vision, 57(2):137–154, 2004. [227].

# Face detection

## Train dataset

### Wider

- 32k images
- 494k faces

### Celeba

- 200k images, 10k persons
- Landmarks, 40 binary attributes


* See **additional materials** on Paul Viola and Michael J. Jones. Robust real-time face detection. International Journal of Computer Vision, 57(2):137–154, 2004. [227].

# Face detection

## Train dataset

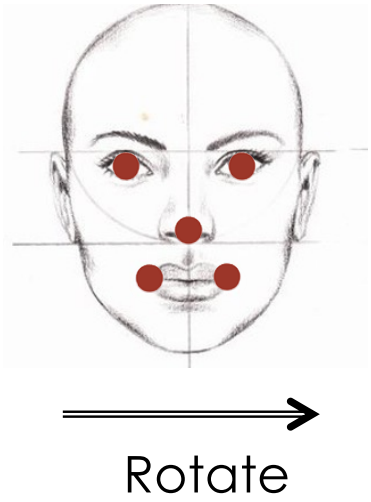
**Face Detection Data Set and Benchmark (FDDB)**

- 2845 images
- 5171 faces

# Face detection – face alignment

## Auxiliary task

- **Face** alignment: rotation
- **Goal:** make it easier for Face Recognition




# Fooling Face-Detection

## CV Dazzle

...

See **additional materials** on https://adam.harvey.studio/cvdazzle/

NEW CV Dazzle Looks N°6, N°7. Developed for Designs for a Different Future 2020 to break convolutional neural network face recognition. © Adam Harvey 2020.

# Methods to detect faces

## Categories of methods

- **Knowledge-based methods:** Encode human knowledge of what constitutes a typical face (usually the relationships between facial features)
- **Feature invariant approaches:** Aim to find structural features of a face that exist even when the pose, viewpoint, or lighting conditions vary
- **Template matching methods:** Several standard patterns stored to describe the face as a whole or the facial features separately
- **Appearance-based methods:** The models (or templates) are learned from a set of training images which capture the representative variability of facial appearance

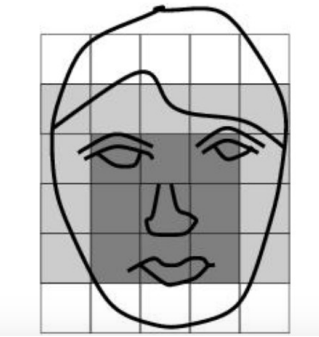
Parts of these slides are taken from Prof. M. Pelillo.

# Knowledge-based Methods

Top-down approach: Represent a face using a set of **human-coded rules**.

**Example:**

- The center part of face has uniform intensity values
- The difference between the average intensity values of the center part and the upper part is significant
- A face often appears with two eyes that are symmetric to each other, a nose and a mouth
- Use these rules to guide the search process

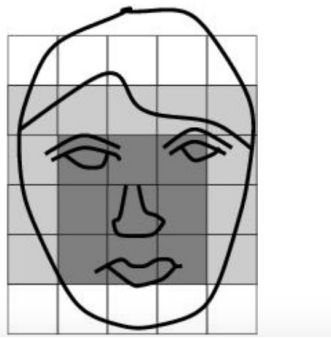


Parts of these slides are taken from Prof. M. Pelillo.

# Knowledge-based Method

[Yang and Huang 94]

- Multi-resolution focus-of-attention approach
- **Level 1** (lowest resolution): apply the rule "the center part of the face has 4 cells with a basically uniform intensity" to search for candidates
- **Level 2**: local histogram equalization followed by edge detection
- **Level 3**: search for eye and mouth features for validation



Parts of these slides are taken from Prof. M. Pelillo.

# Knowledge-based Method

[Kotropoulos & Pitas 94]

Horizontal/vertical projection to search for candidates

$$HI(x) = \sum_{y=1}^{n} I(x, y) \qquad VI(y) = \sum_{x=1}^{m} I(x, y)$$

Search eyebrow/eyes, nostril/nose for validation

Difficult to detect multiple people or in complex background

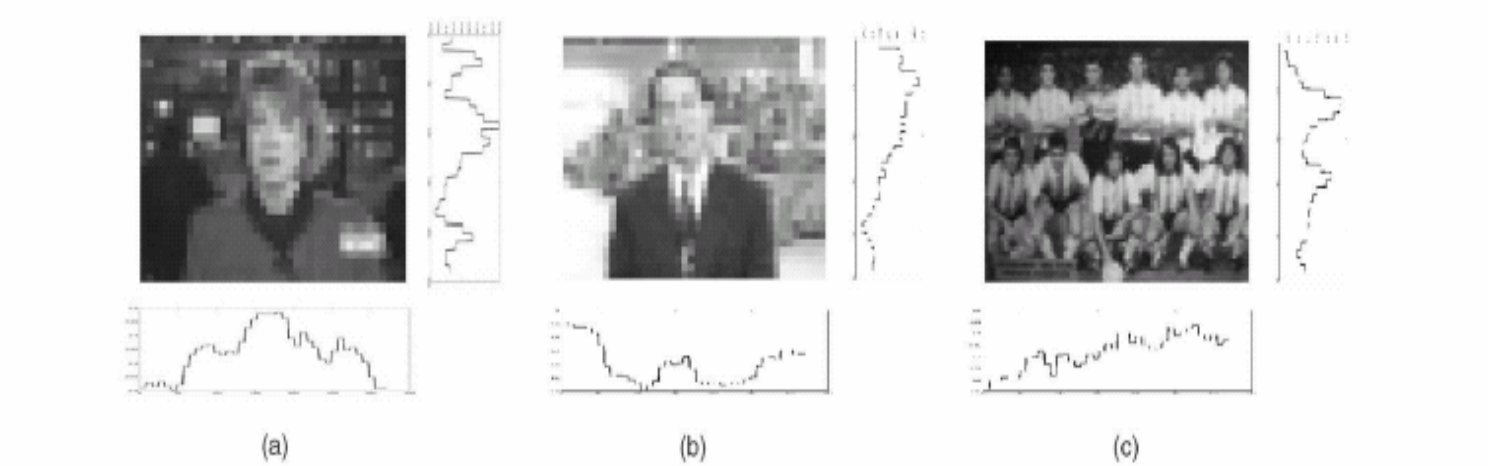



Fig. 3. (a) and (b) n = 8. (c) n = 4. Horizontal and vertical profiles. It is feasible to detect a single face by searching for the peaks in horizontal and vertical profiles. However, the same method has difficulty detecting faces in complex backgrounds or multiple faces as shown in (b) and (c).

Parts of these slides are taken from Prof. M. Pelillo.

# Knowledge-based Method

**Pros:**

- Easy to come up with simple rules to describe the features of a face and their relationships
- Based on the coded rules, facial features in an input image are extracted first, and face candidates are identified
- Work well for face localization in uncluttered background

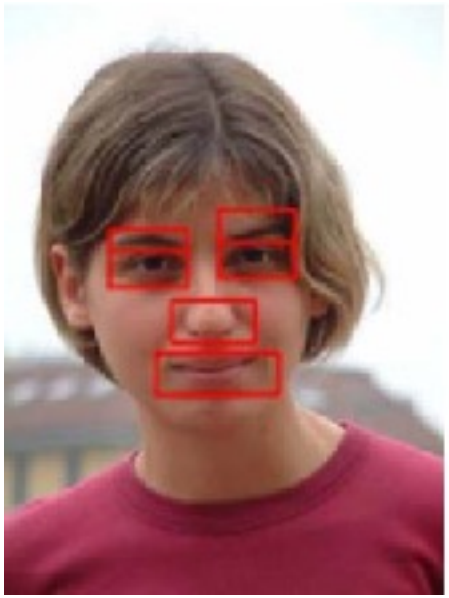
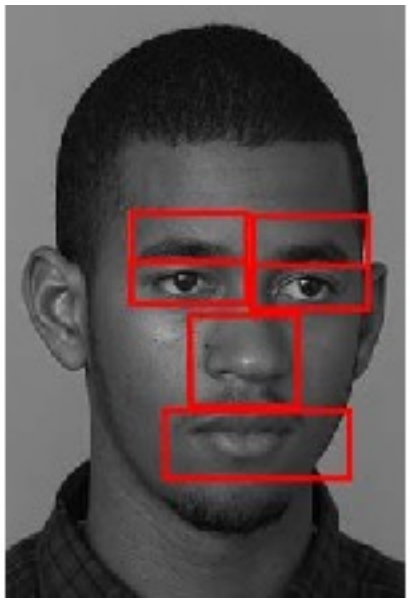
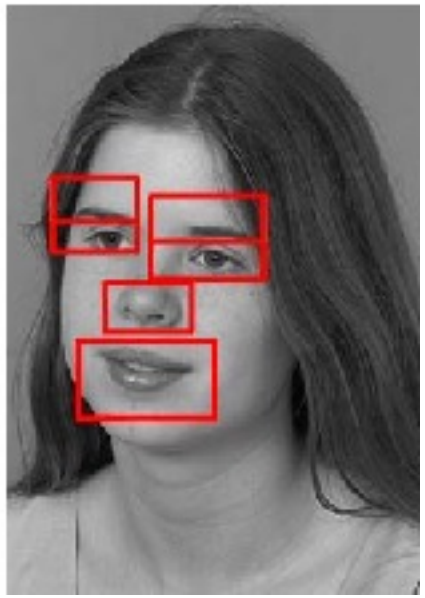
**Cons:**

- Difficult to translate human knowledge into rules precisely: detailed rules fail to detect faces and general rules may find many false positives
- Difficult to extend this approach to detect faces in different poses: implausible to enumerate all the possible cases

Parts of these slides are taken from Prof. M. Pelillo.

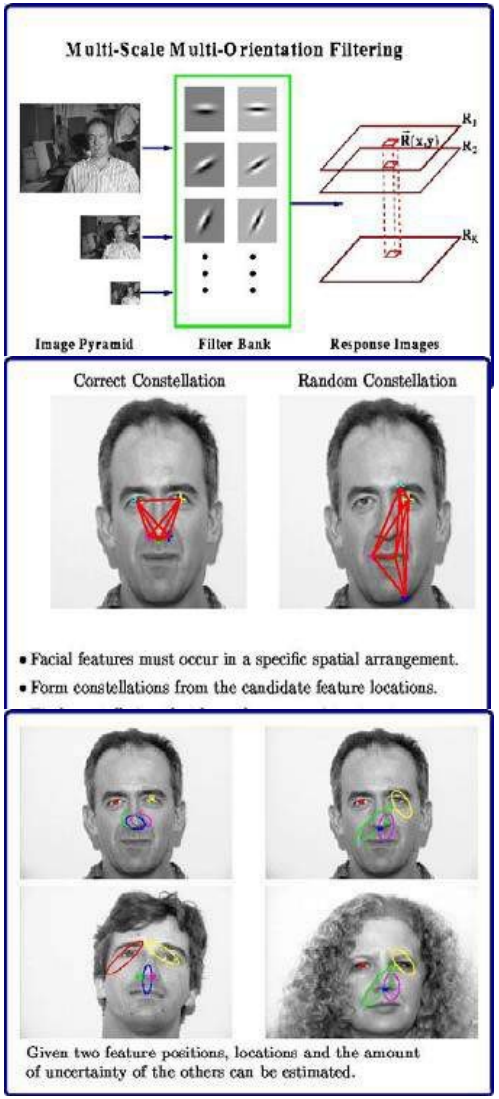
# Feature-based Method

- Bottom-up approach: Detect facial features (eyes, nose, mouth, etc) first
- Facial features: edge, intensity, shape, texture, color, etc
- Aim to detect invariant features
- Group features into candidates and verify them

# Random Graph Matching [Leung et al. 95]

- Formulate as a problem to find the correct geometric arrangement of facial features
- Facial features are defined by the average responses of multi-scale filters
- Graph matching among the candidates to locate faces



# Feature-based Method

**Pros:**

- Features are invariant to pose and orientation change

**Cons:**

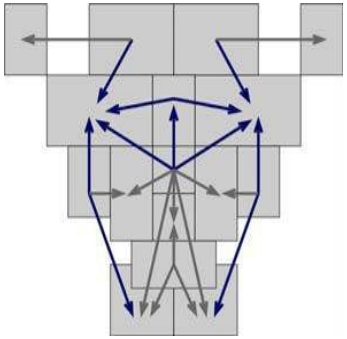
- Difficult to locate facial features due to several corruption (illumination, noise, occlusion)
- Difficult to detect features in complex background

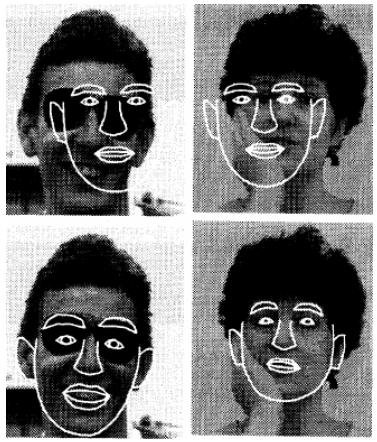
# Template-Matching methods

## Face templates

- Store a template
  - Predefined: based on edges or regions
  - Deformable: based on facial contours (e.g., Snakes)
- Templates are hand-coded (not learned)
- Use correlation to locate faces

**Face templates**





# Template-based methods

## Pros

- Simple

## Cons

- Templates needs to be initialized near the face images
- Difficult to enumerate templates for different poses (similar to knowledge-based methods)

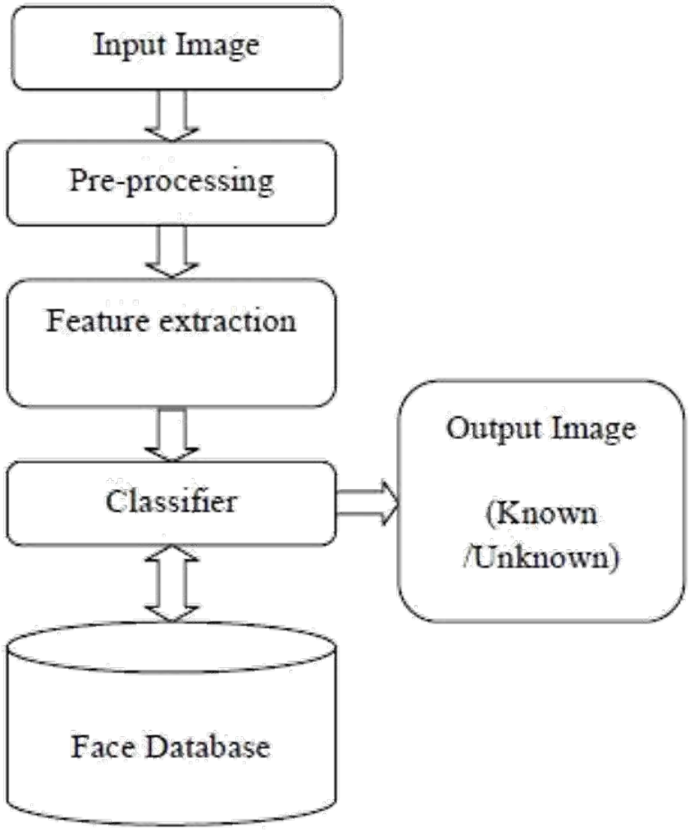
# Appearance-based methods

## Feature extraction + classification

### General idea

- Collect a **large set** of (resized) **face** and **non-face images** and train a classifier to discriminate them.

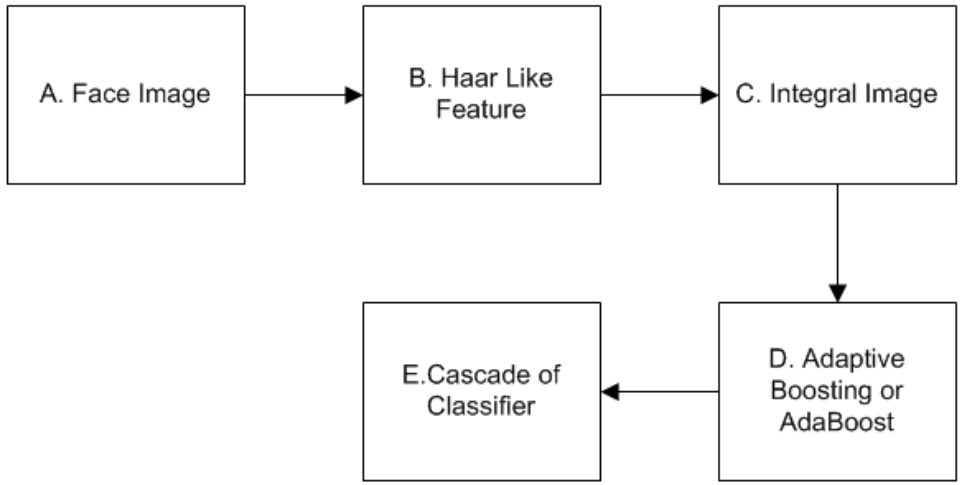
- Given a **test image**, detect faces by applying the **classifier** at each **position** and **scale** of the image.



# The Viola-Jones Face Detector

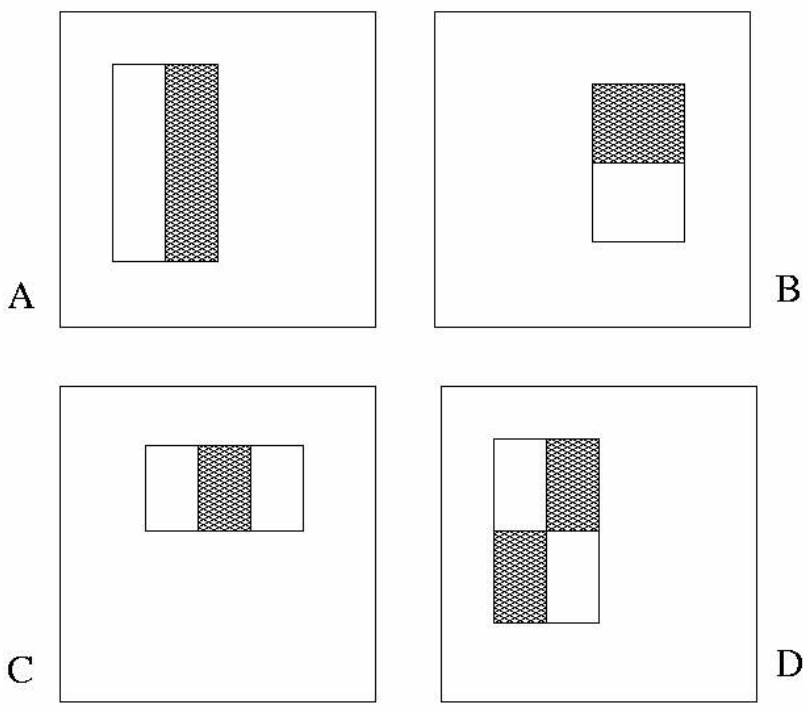
- A seminal approach to **real-time object detection**
- Training is slow, but detection is very fast

- Key ideas
  - **Integral images** for fast feature evaluation
  - **Boosting** for feature selection
  - **Attentional cascade** for fast rejection of non-face windows



# Rectangular Image Features

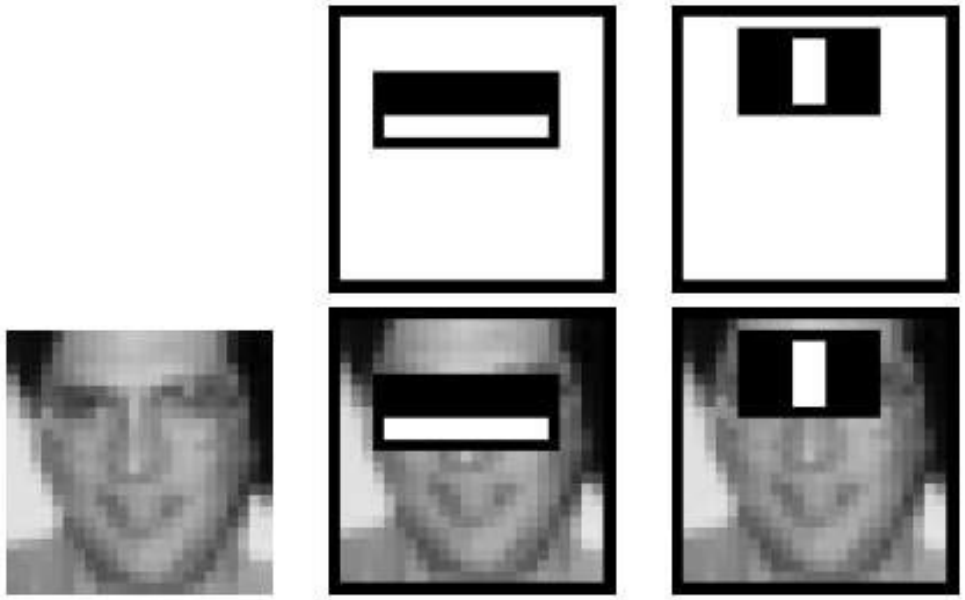
## Robust features




*Value = ∑ (pixels in white area) – ∑ (pixels in black area)*

# Rectangular Image Features

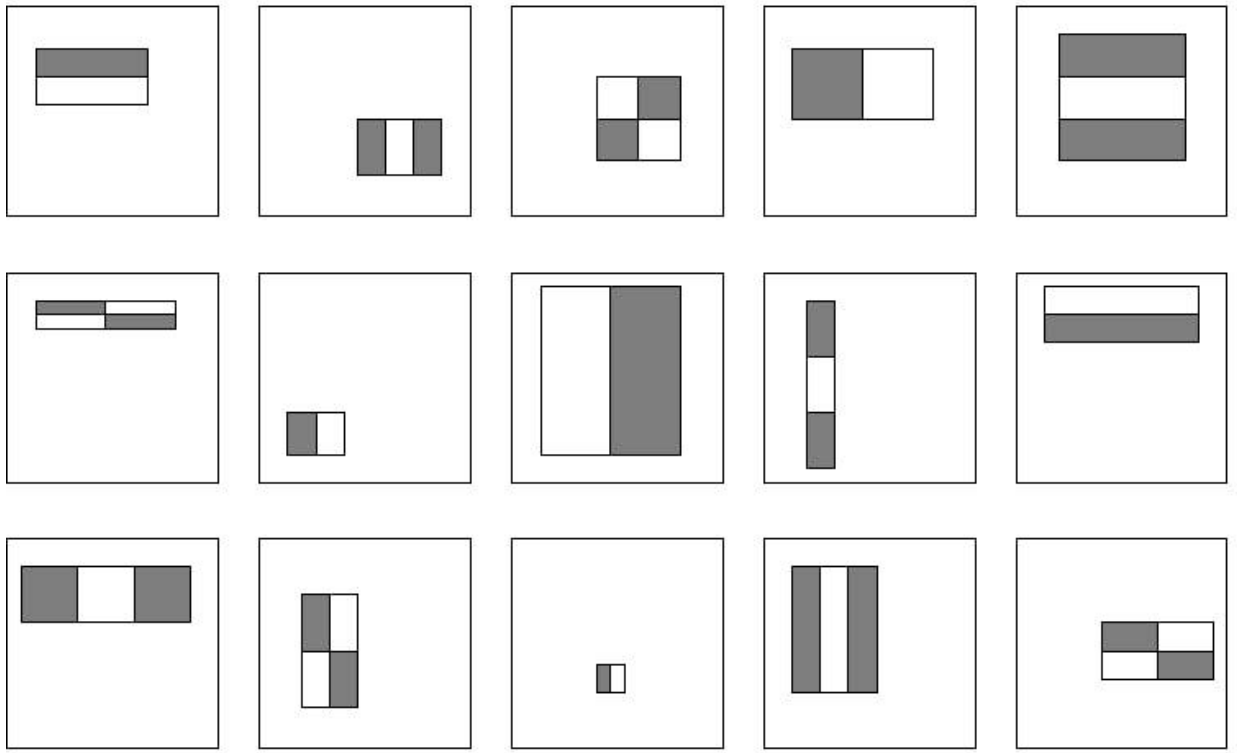
## Haar-like features



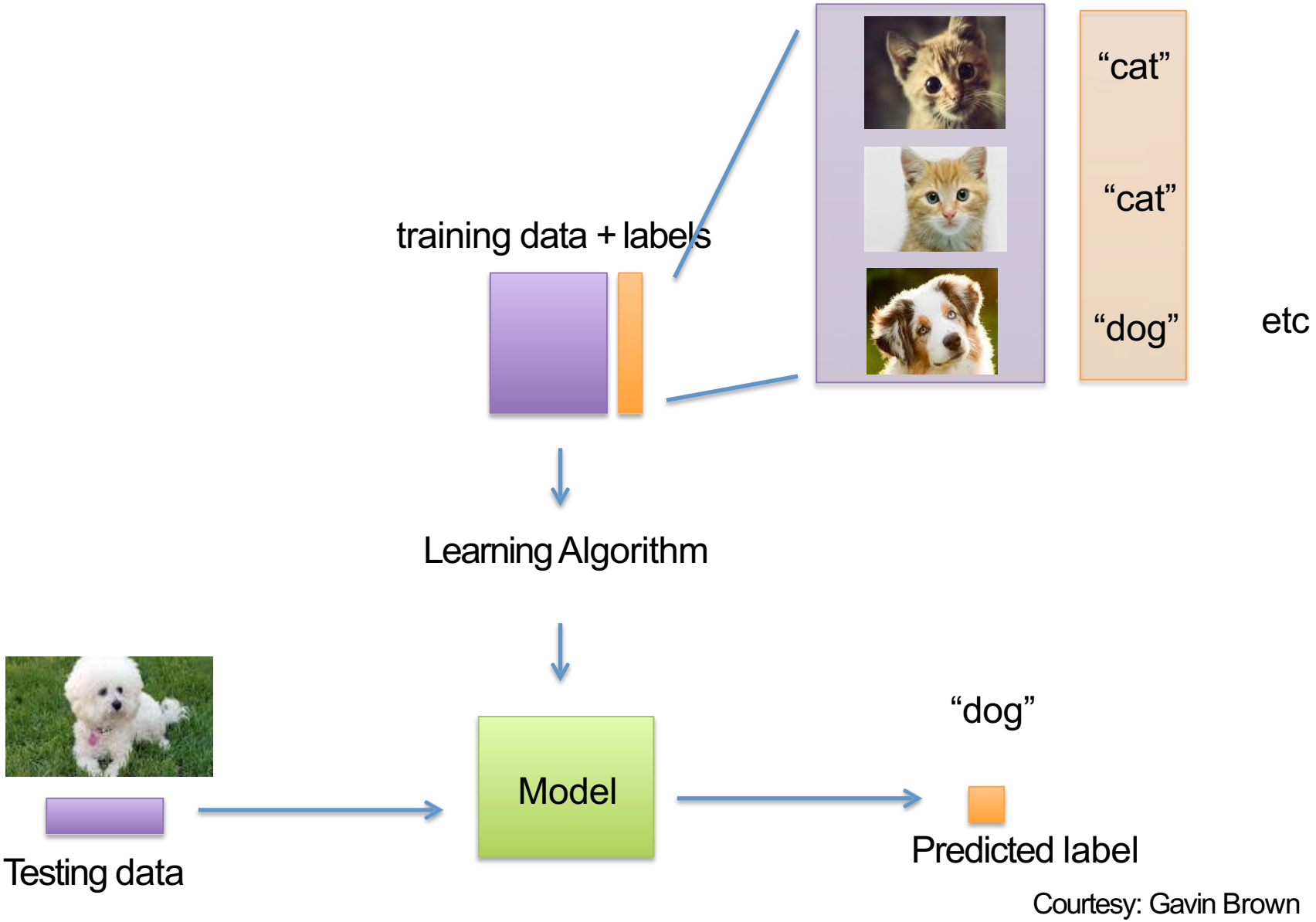
Forehead, eye features can be captured

# Feature selection

- For a 24x24 detection region, the number of possible rectangle features is ~160,000!

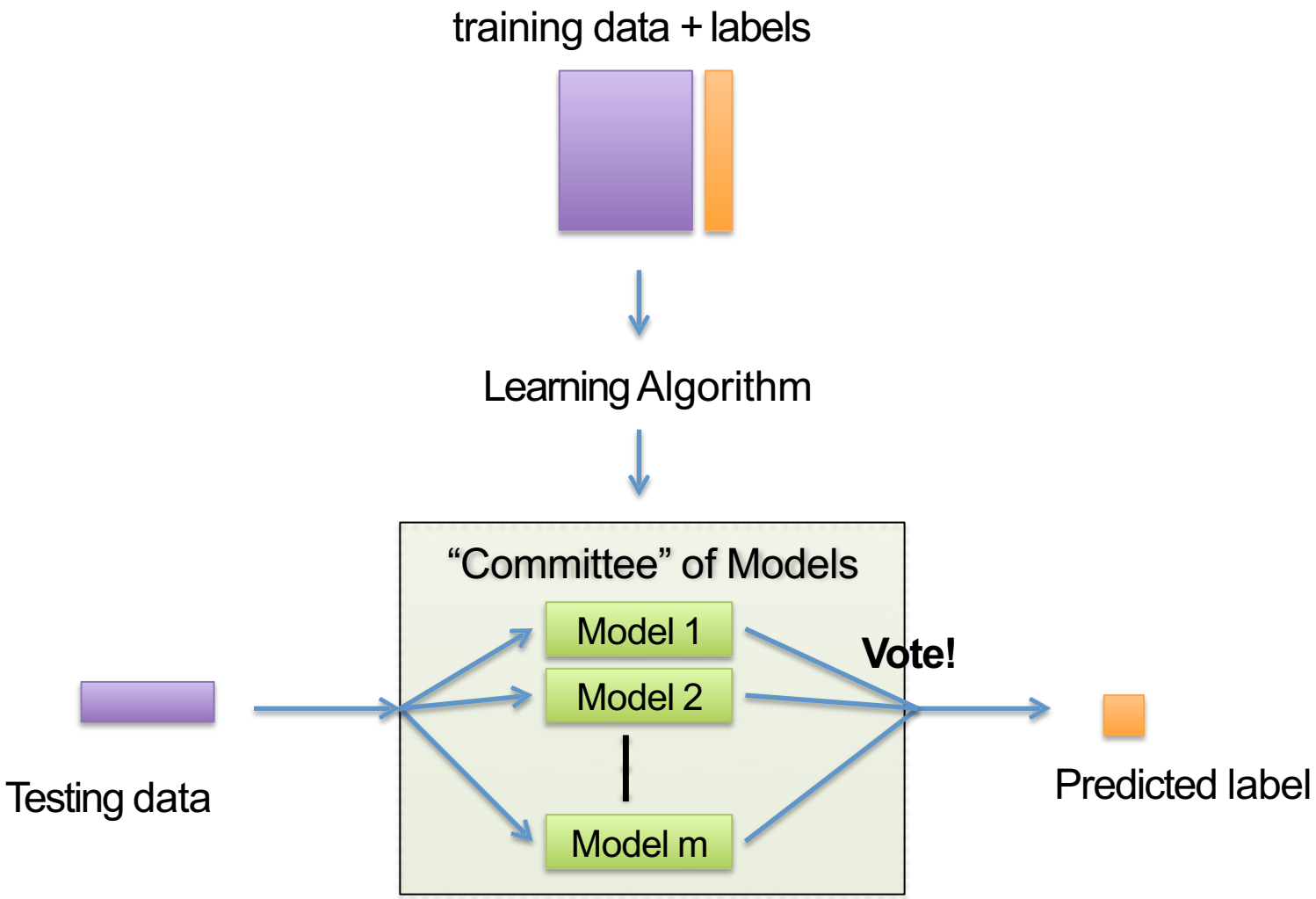


# The Supervised Learning Pipeline (the two-class case)



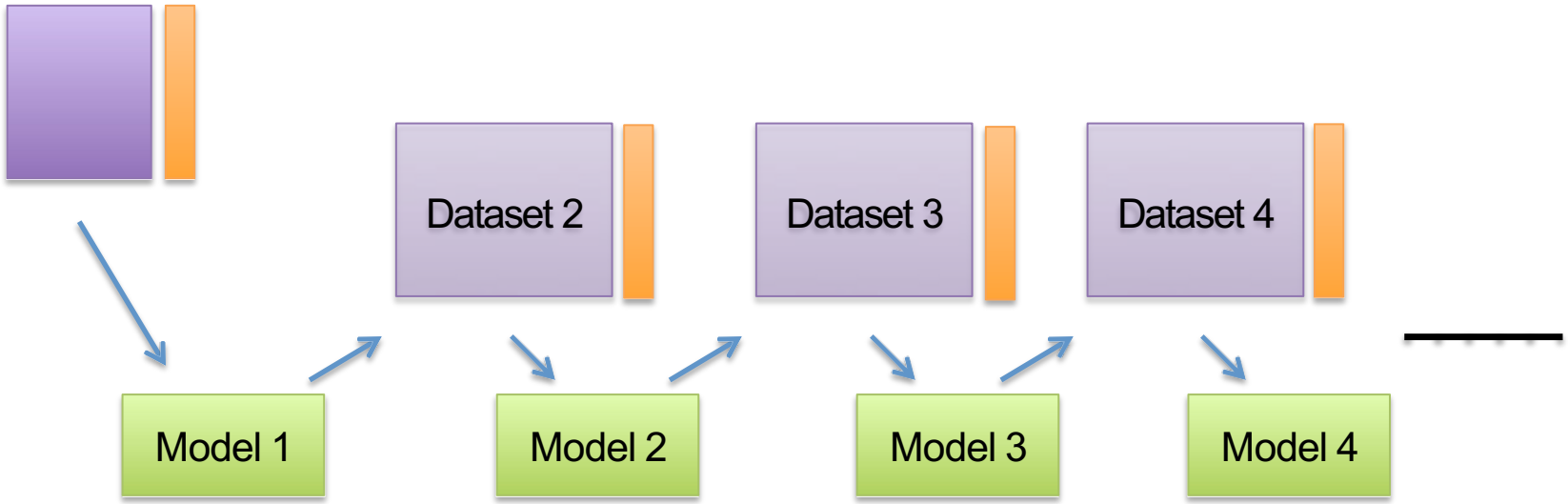

Courtesy: Gavin Brown

# The Ensemble approach




Courtesy: Gavin Brown

# Boosting




"Boosting" algorithms build an ensemble, sequentially.

Each model corrects the mistakes of its predecessors.

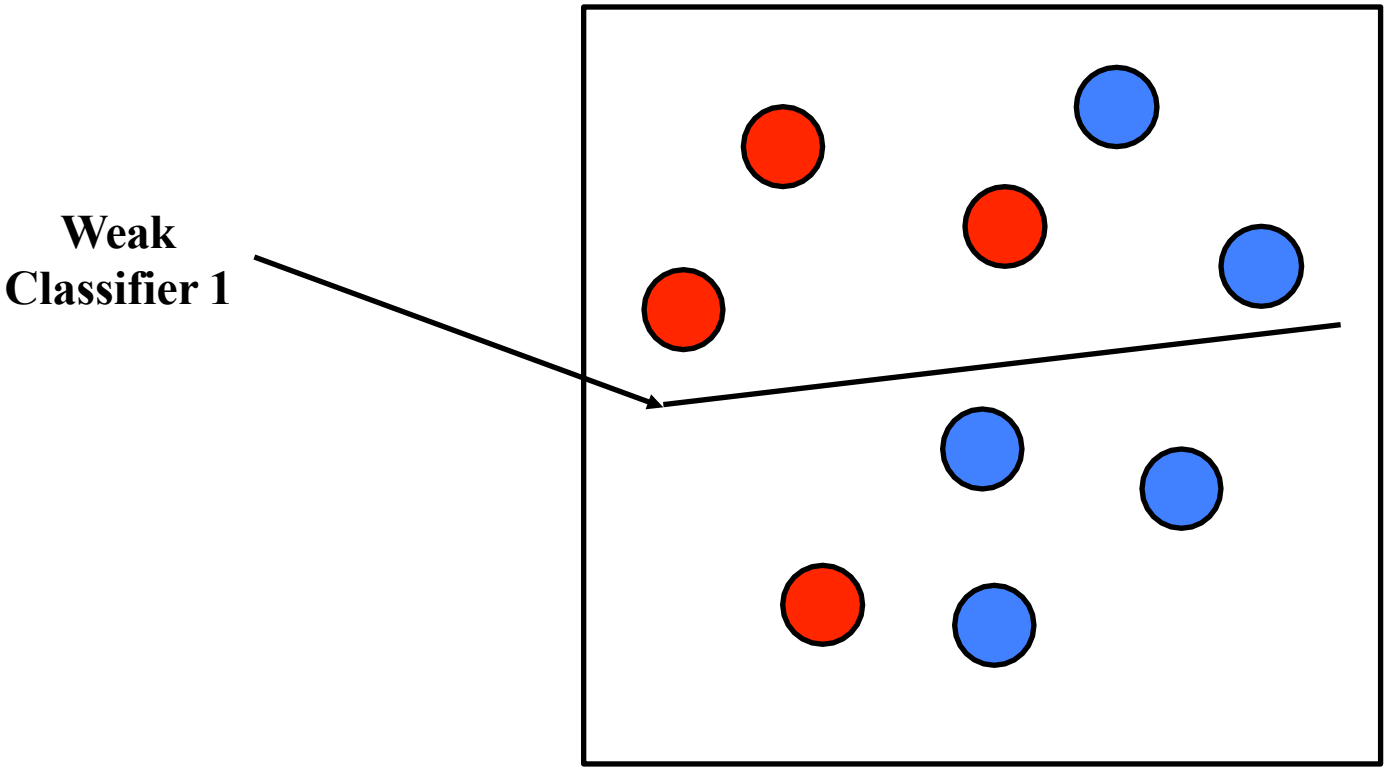
Courtesy: Gavin Brown

# Boosting

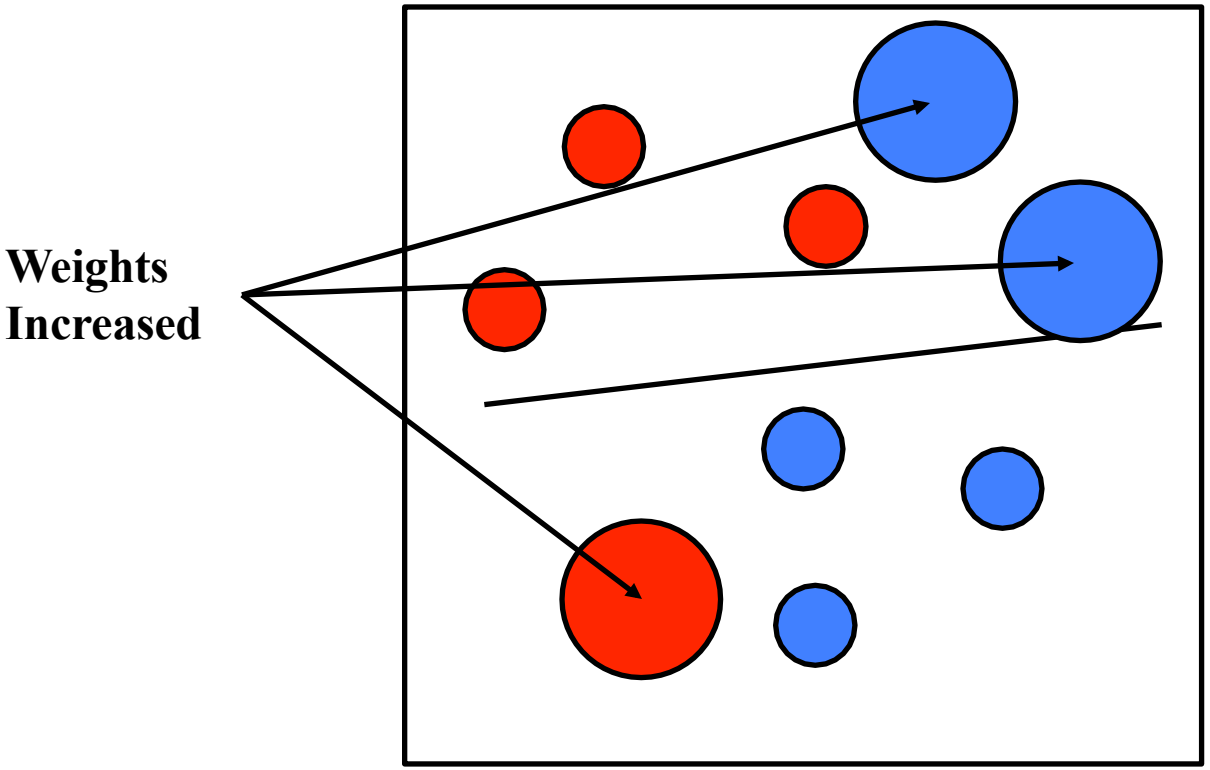
- Boosting is a classification scheme that works by combining weak learners into a more accurate ensemble classifier
  - A weak learner need only do better than chance

- Training consists of multiple boosting rounds
  - During each boosting round, we select a weak learner that does well on examples that were hard for the previous weak learners
  - “Hardness” is captured by weights attached to training examples

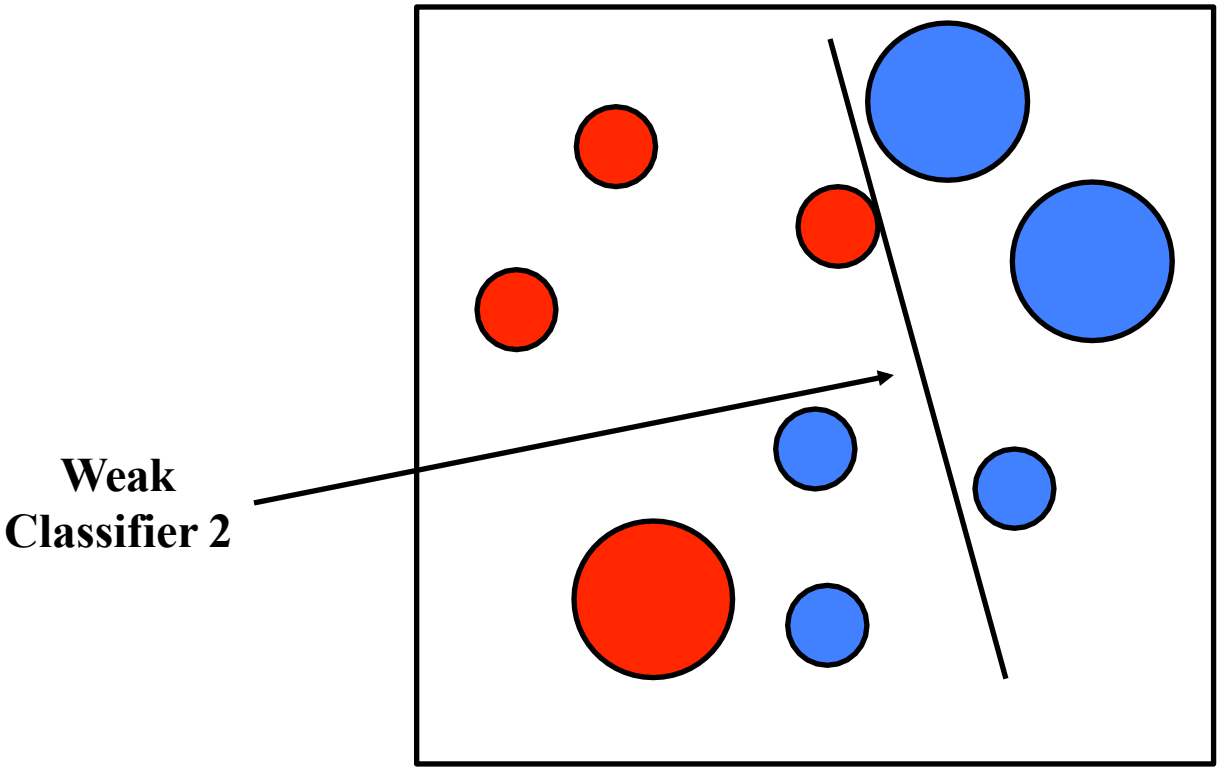
# Boosting at work



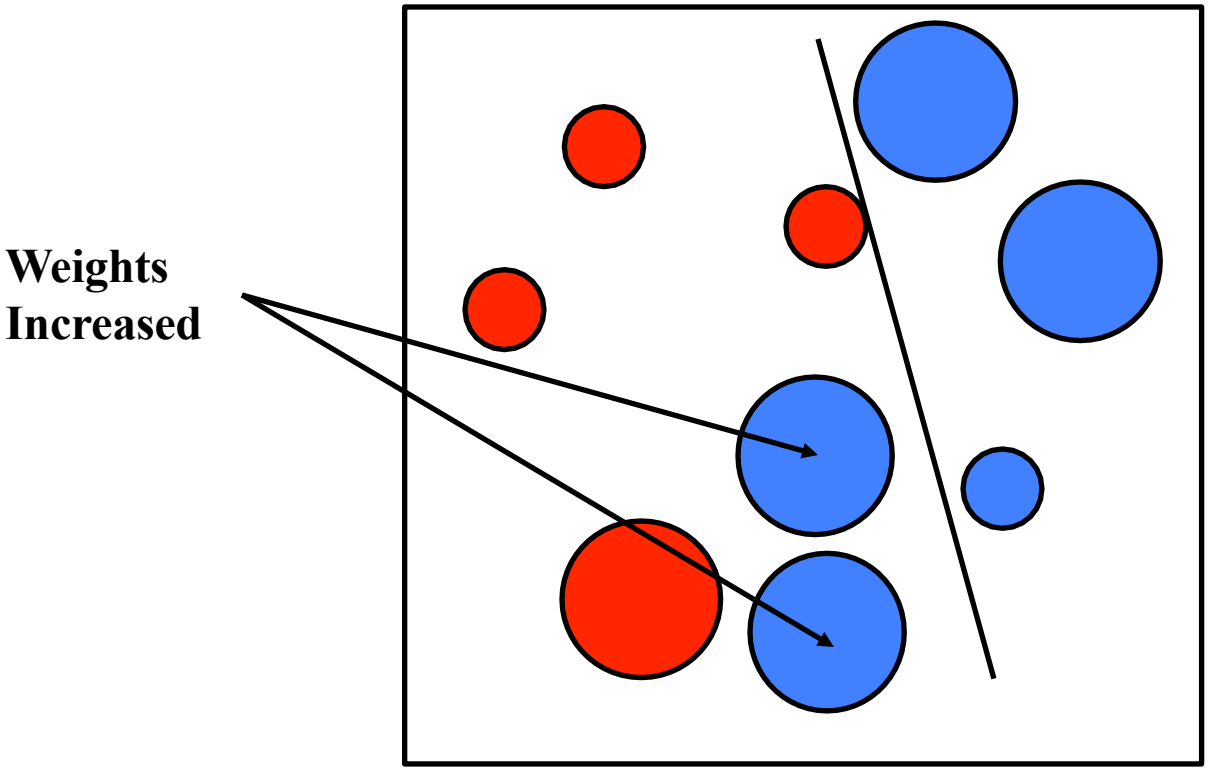
# Boosting at work



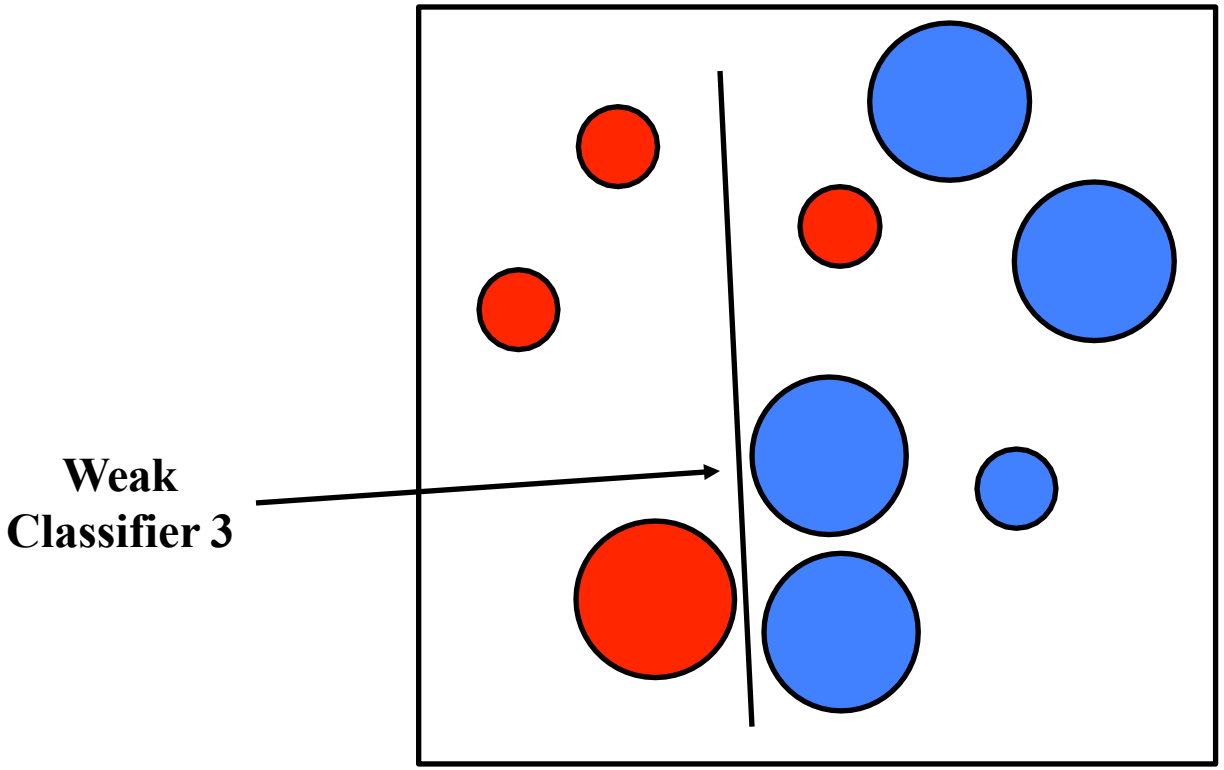
# Boosting at work



# Boosting at work

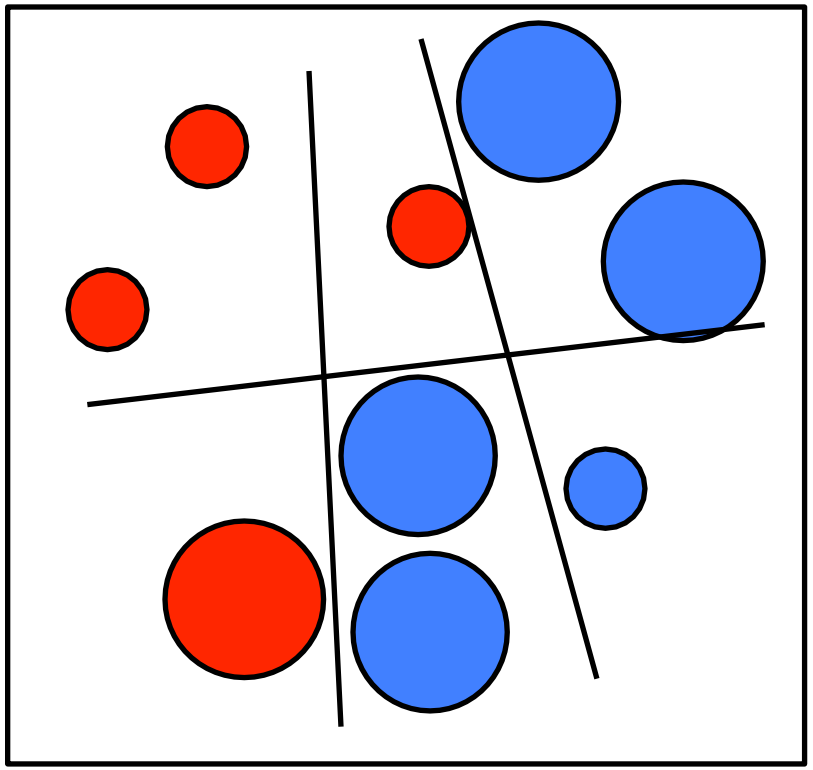


# Boosting at work



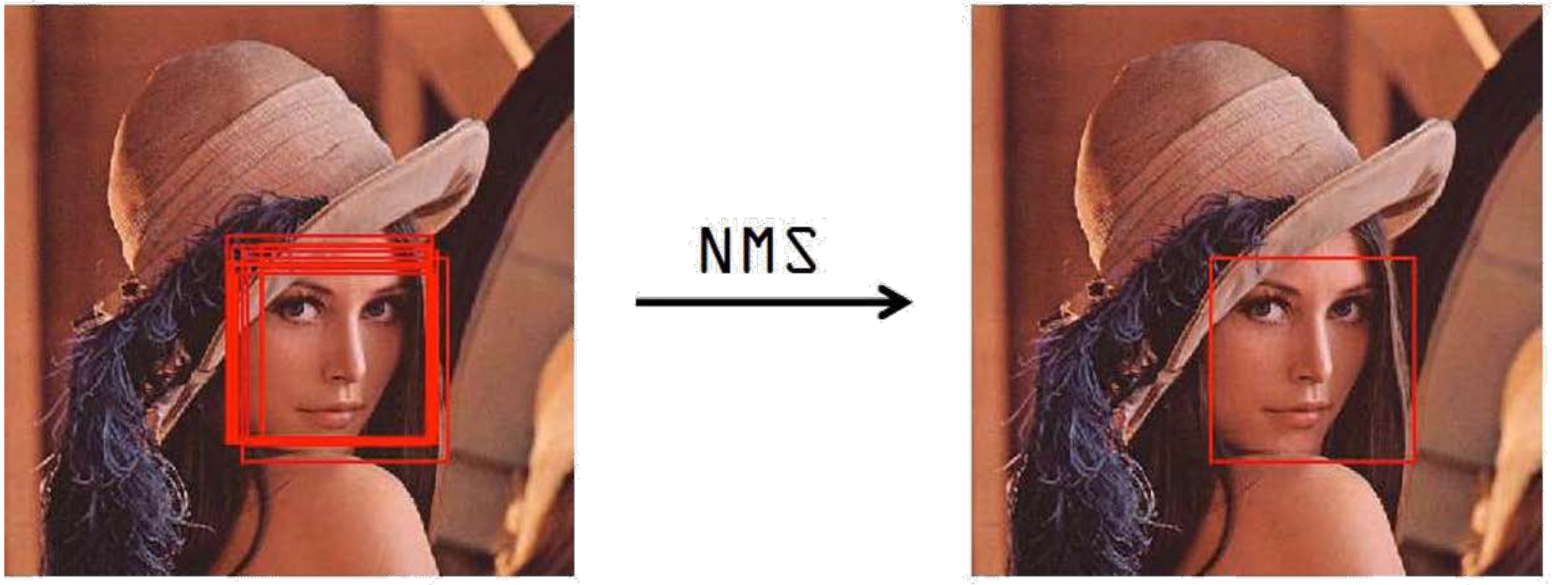
# Boosting at work

**Final classifier is a combination of weak classifiers**



# Non-maximum Suppression

- The set of detections are first partitioned into disjoint subsets
- Two detections are in the same subset if their regions overlap
- Each partition yields a single final detection
- The corners of the final bounding region are the average of the corners of all detections in the set



# The implemented system

## Training Data

- 4916 hand labeled faces
- 10000 non faces
- Faces are normalized
  - Scale, translation

## Many variations

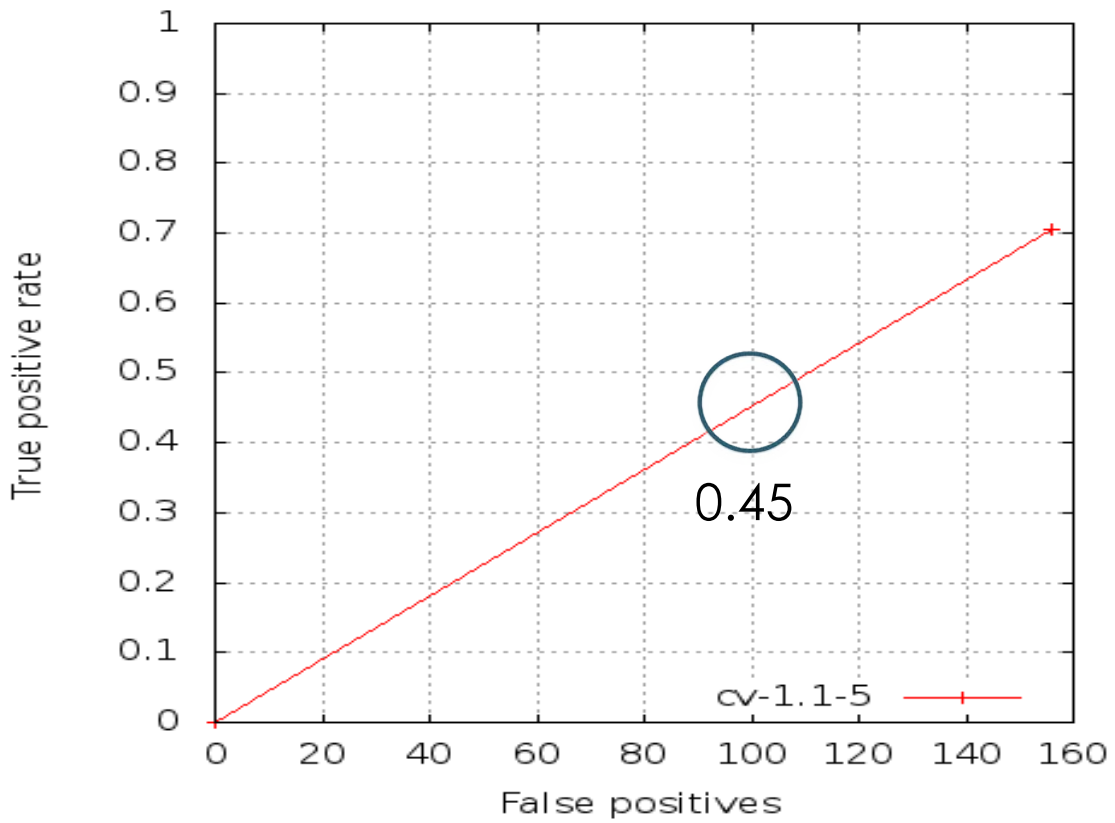
- Across individuals
- Illumination
- Pose (rotation both in plane and out)

# The implemented system

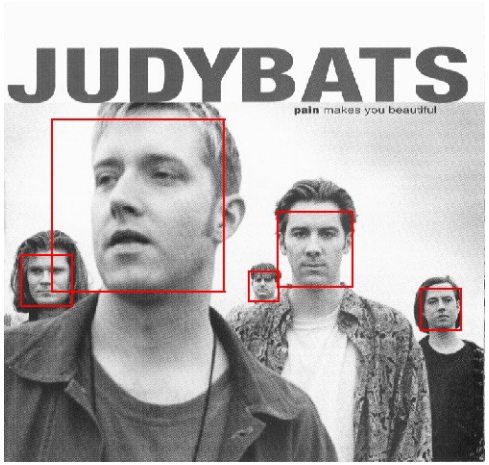
OpenCV implementation

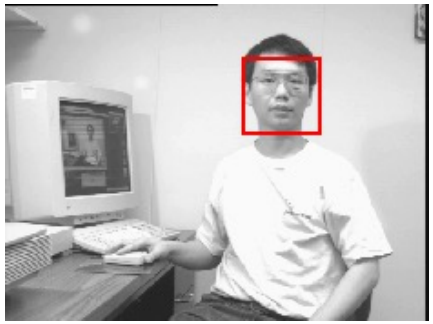
- Fast: ~100ms  on CPU
- Not accurate

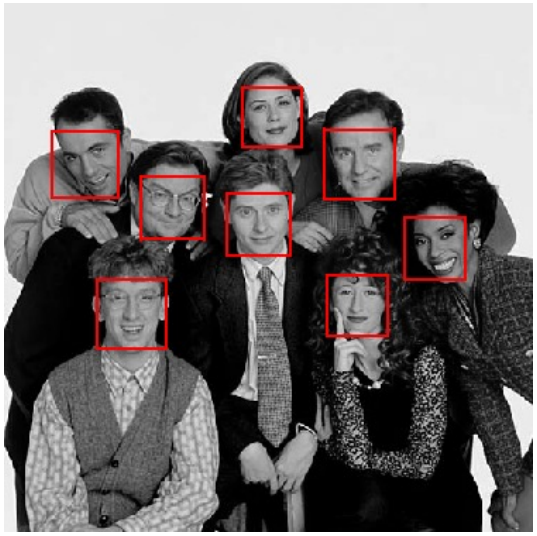
FDDB results

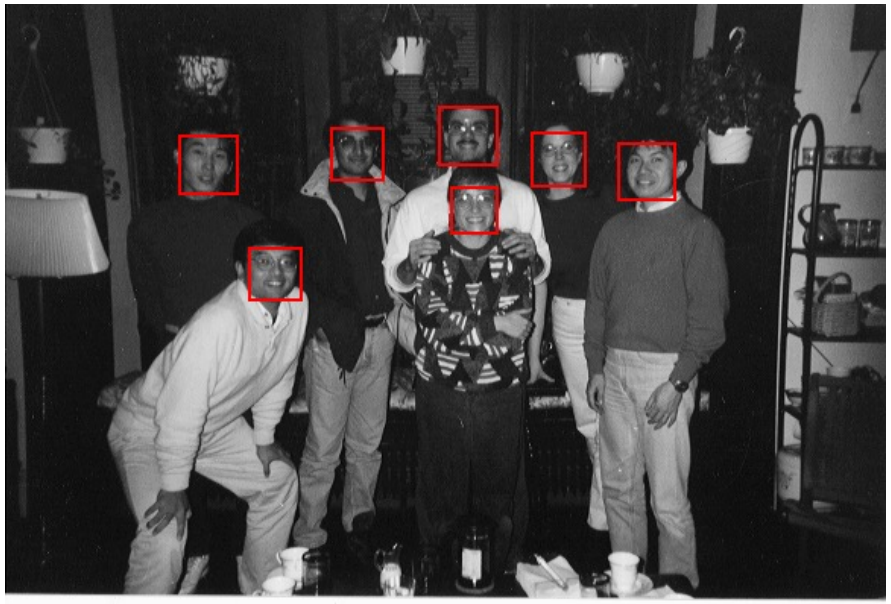


# Viola-Jones at work



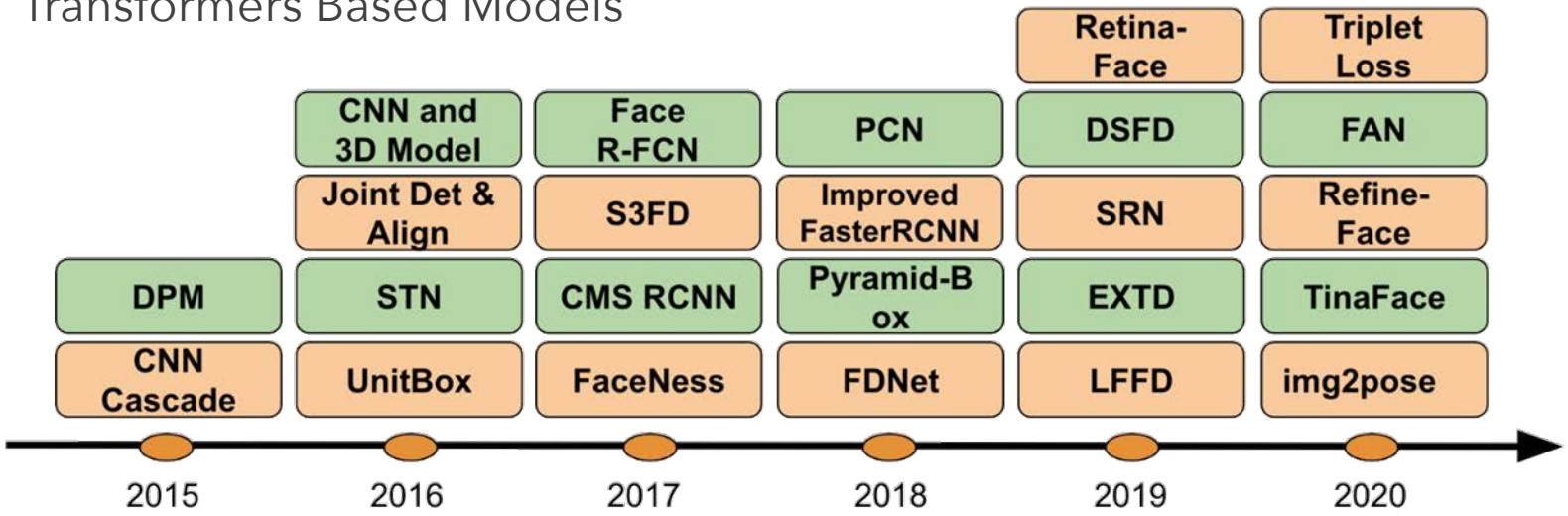




# Neural Face detection

## Deep Learning methods

- There are many approaches proposed for face detection using different deep learning architectures:
  - Cascade-CNN Based Models
  - R-CNN Based Models
  - Single Shot Detector Models
  - Feature Pyramid Network Based Models
  - Transformers Based Models

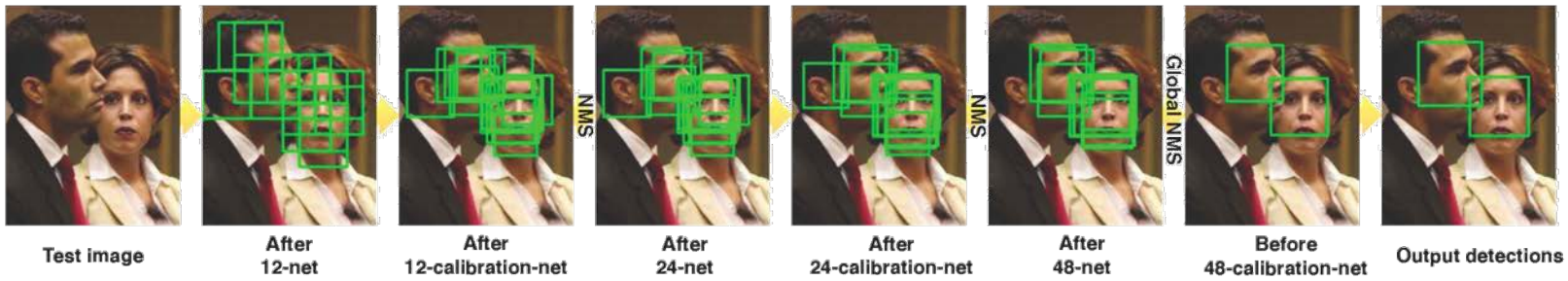


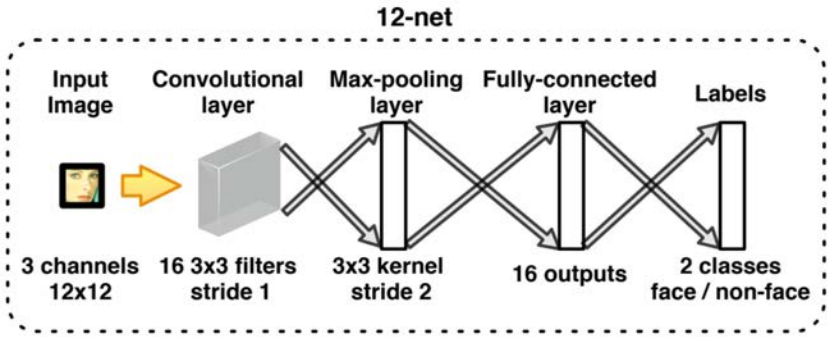

* See **additional materials** on Minaee, S., Luo, P., Lin, Z., & Bowyer, K. (2021). Going deeper into face detection: A survey. *arXiv preprint arXiv:2103.14983*.

# Cascade-CNN Based Models

## Deep Learning methods

- It is one of the early deep models for face detection, based on a convolutional neural network cascade.
- The proposed method runs at 14 FPS on a single CPU core for VGA-resolution images and 100 FPS using a GPU.



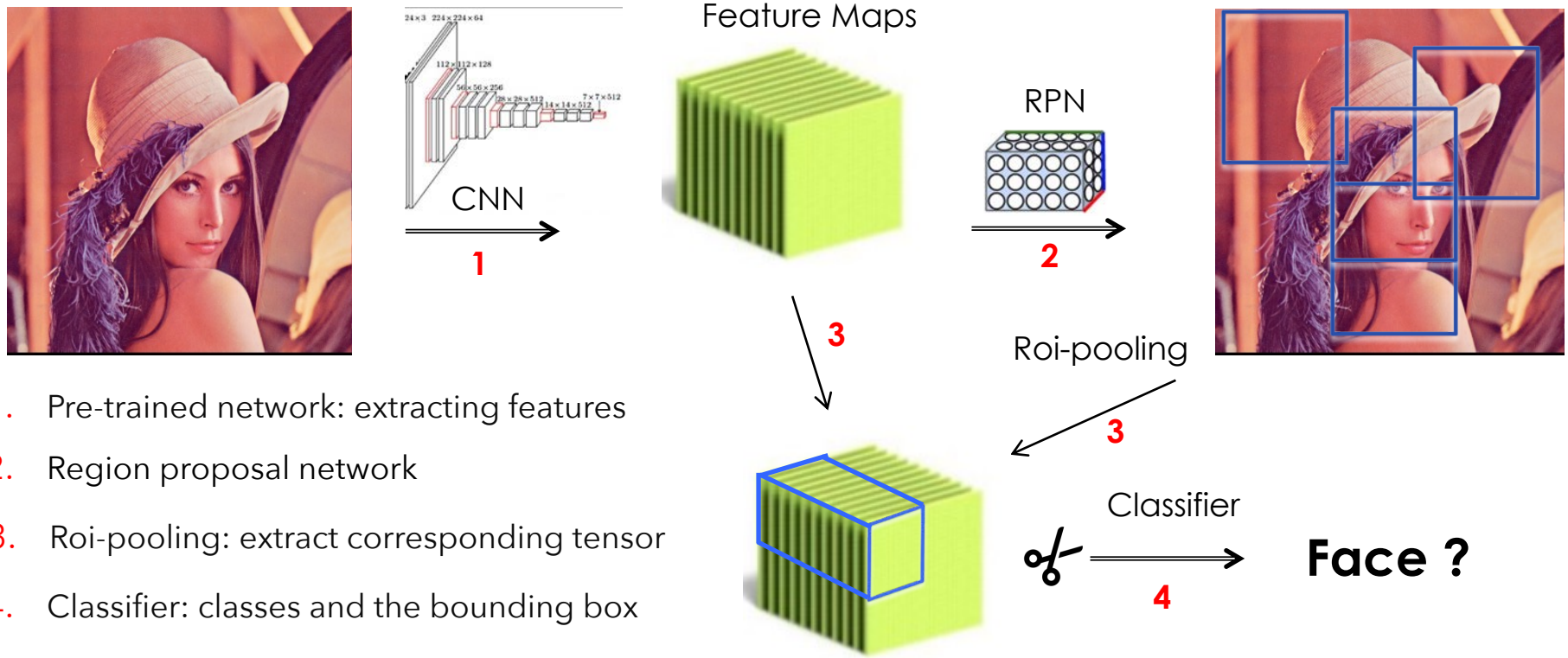




* See **additional materials** on Minaee, S., Luo, P., Lin, Z., & Bowyer, K. (2021). Going deeper into face detection: A survey. *arXiv preprint arXiv:2103.14983*.

# R-CNN and Faster-RCNN Based Models

## Deep Learning methods

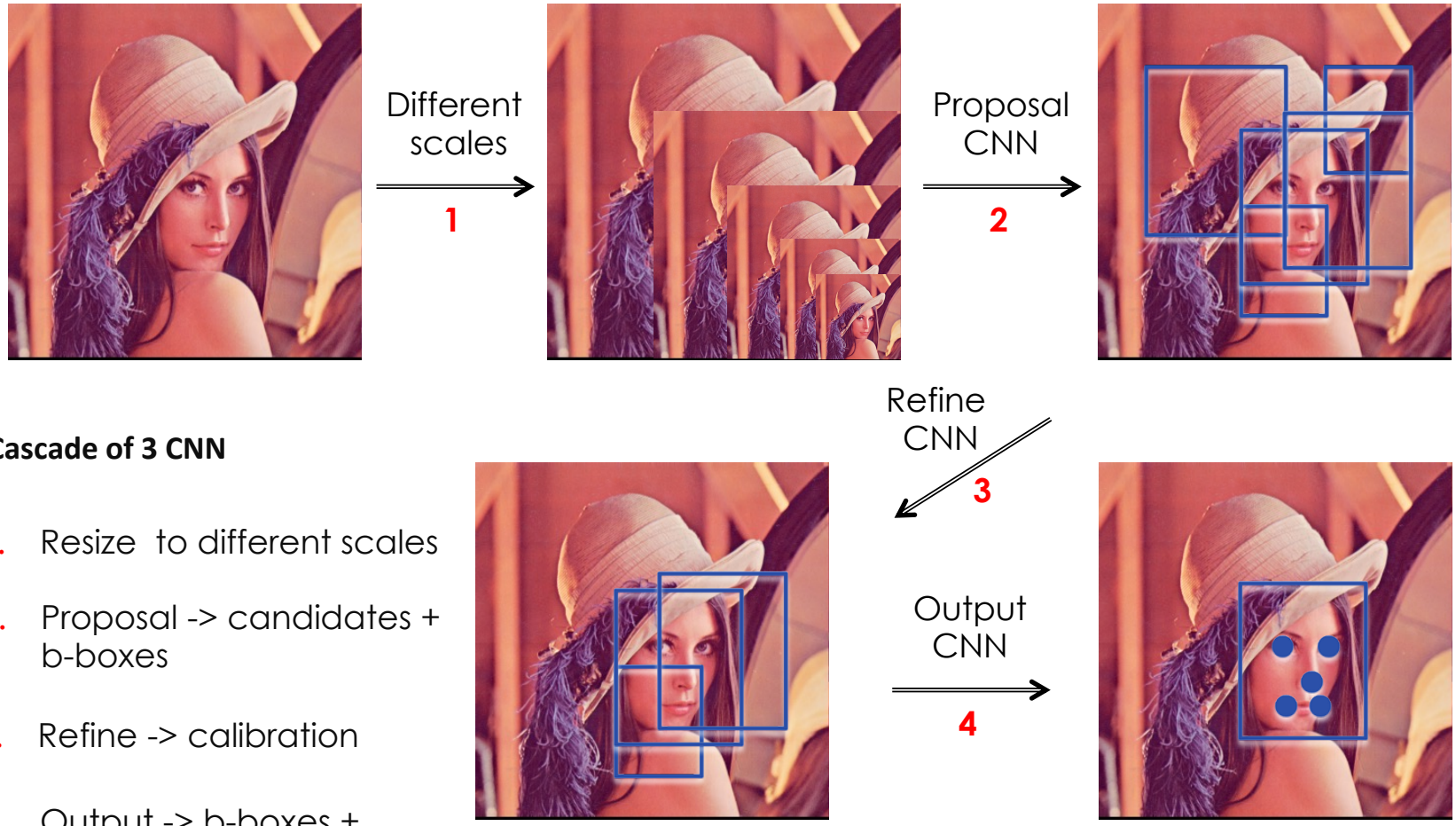
- Region proposal-based CNN models have been very successful for object detection, and have also been applied to face detection
- Classifier: classes and the bounding box




1. Pre-trained network: extracting features
2. Region proposal network
3. Roi-pooling: extract corresponding tensor
4. Classifier: classes and the bounding box

* See **additional materials** on Minaee, S., Luo, P., Lin, Z., & Bowyer, K. (2021). Going deeper into face detection: A survey. *arXiv preprint arXiv:2103.14983*.

# MTCNN - Multi-Task Cascaded Convolutional Neural Networks




**Cascade of 3 CNN**

1. Resize to different scales
2. Proposal -> candidates + b-boxes
3. Refine -> calibration
4. Output -> b-boxes + landmarks

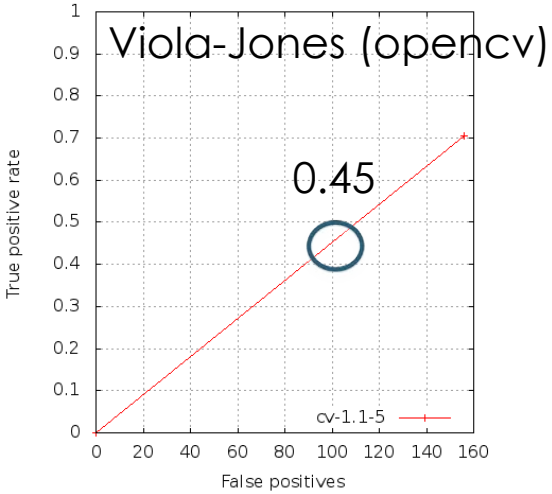
* See **additional materials** on Minaee, S., Luo, P., Lin, Z., & Bowyer, K. (2021). Going deeper into face detection: A survey. *arXiv preprint arXiv:2103.14983*.

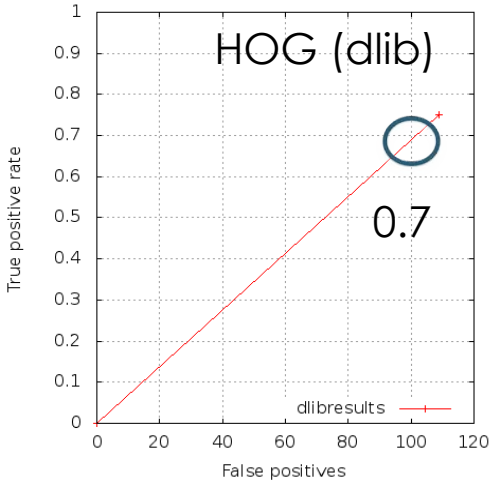
# Comparison: Viola-Jones vs R-FCN

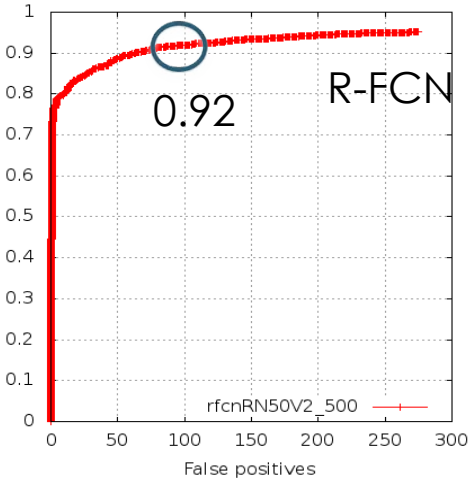
## Deep Learning methods

- Region 92% accuracy (R-FCN)
- 40ms on GPU (slow)

FDDB results



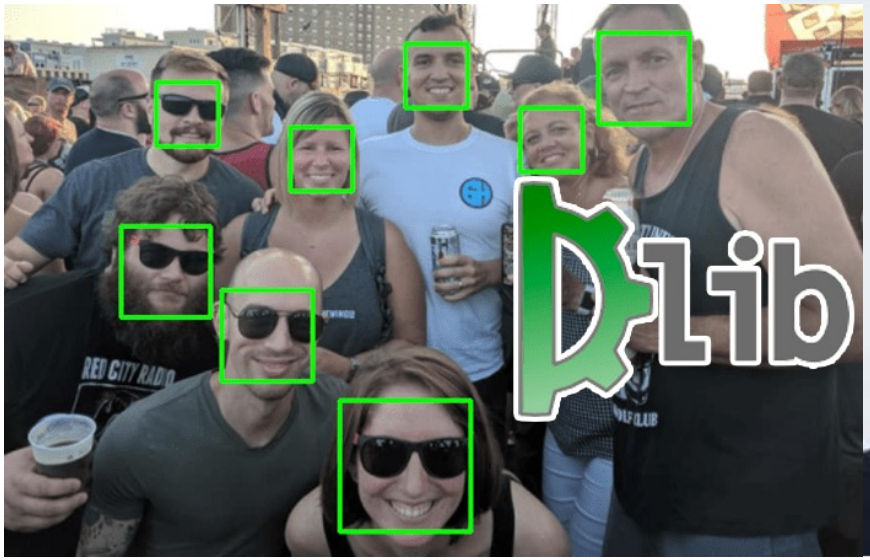



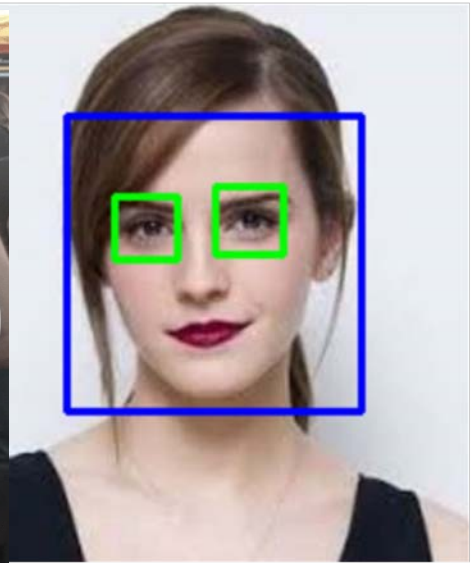



# Face detection libraries

## Methods

http://dlib.net/
https://opencv.org/



OpenCV

# Face Recognition

# Face Recognition

## Definition

**Face recognition** is the task of identifying a person from the analysis of the face. Face recognition is classified into two parts:

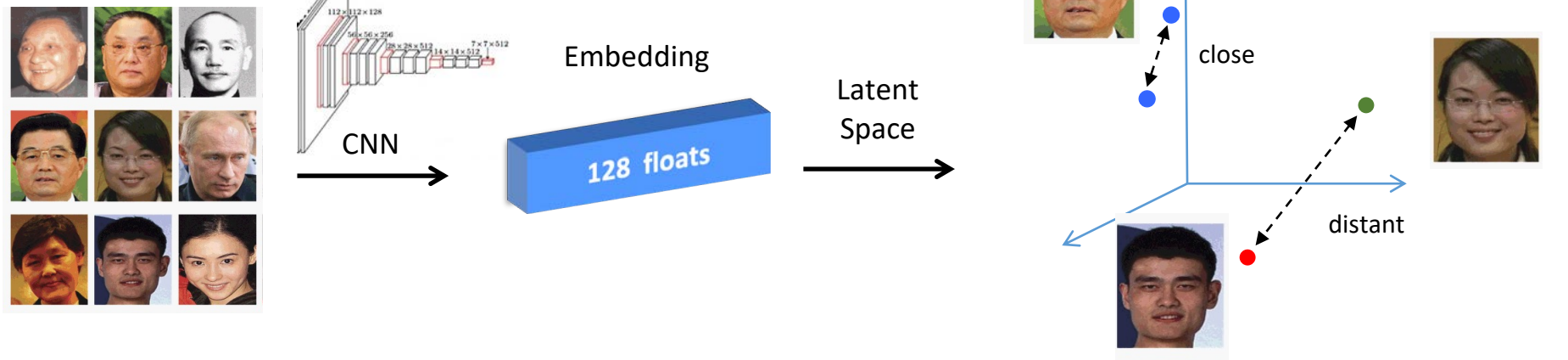
- **face identification** is a computer vision task that involves recognizing and distinguishing between different individuals based on their facial features.
- **face verification** is a computer vision task that involves determining whether two given facial images or representations belong to the same person.

* See **additional materials** on Wang, M., & Deng, W. (2021). Deep face recognition: A survey. *Neurocomputing*, *429*, 215-244.

# Face Recognition

## Deep Learning methods

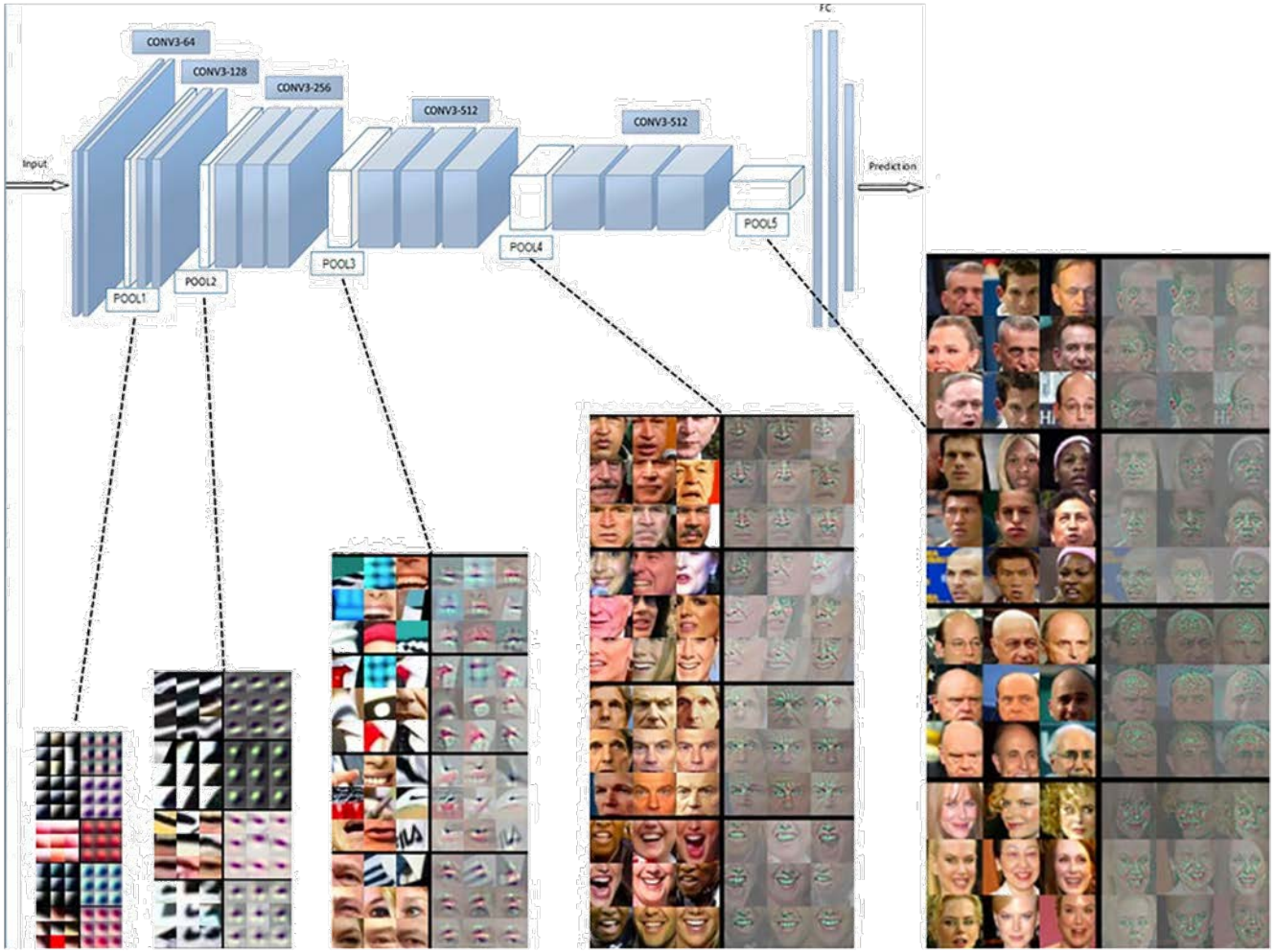
- Goal – to compare faces
- How? To learn metric
- To enable Zero-shot learning



# Face Recognition

## Deep Learning methods

The hierarchical architecture that stitches together pixels into invariant face representation. The output is a compressed feature vector that represents the face.

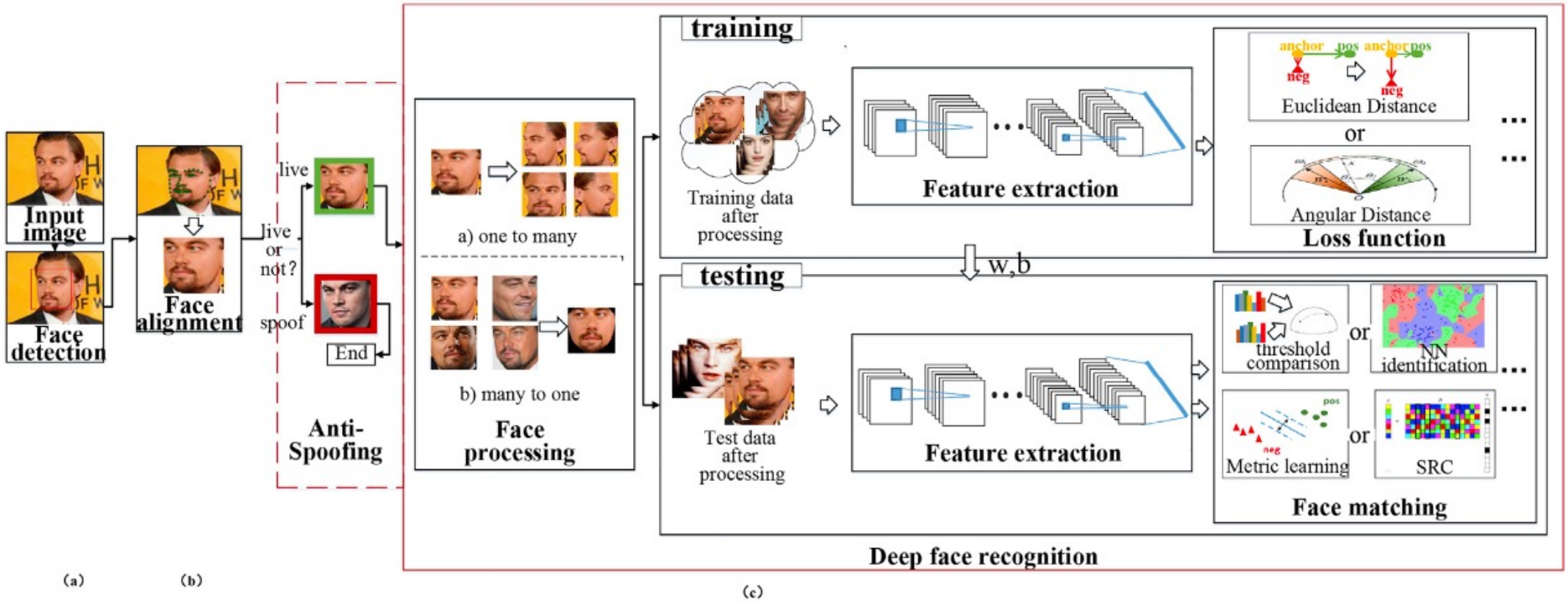


# Face Recognition (FR)

## Pipeline

### Deep FR pipeline

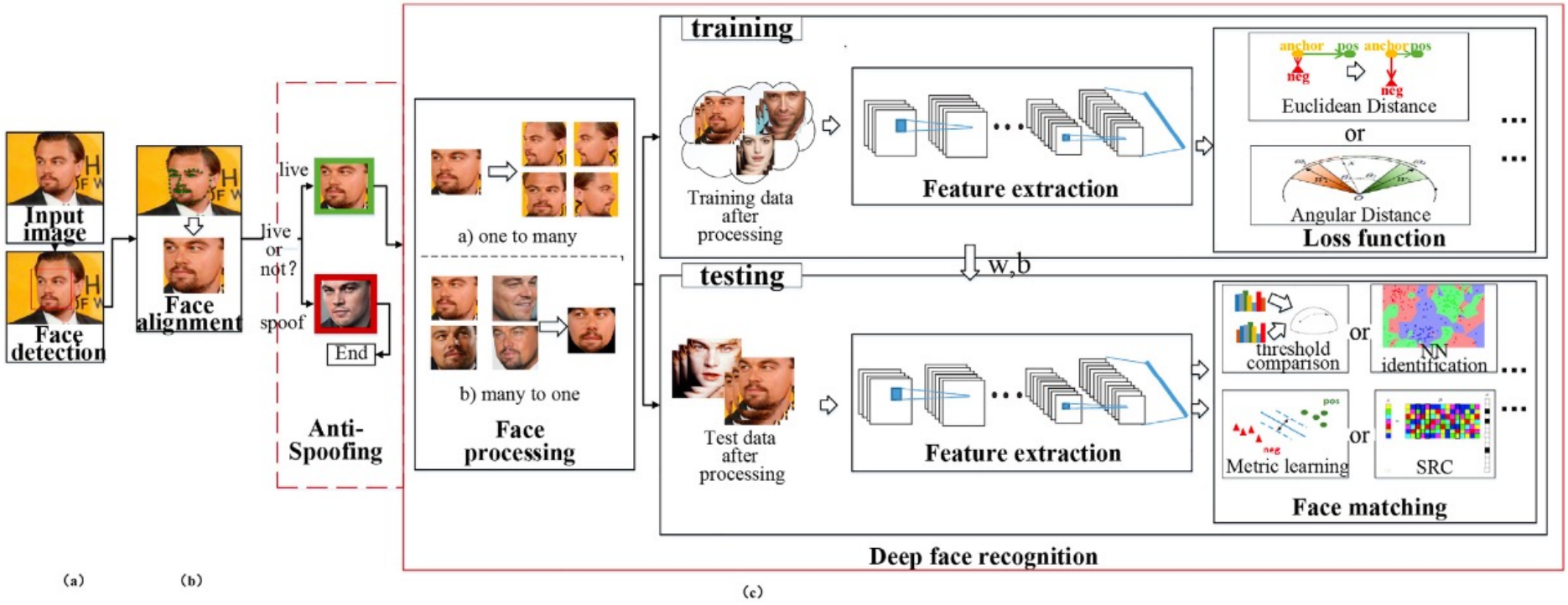
- First, a face detector is used to localize faces in images or videos.
- Second, with the facial landmark detector, the faces are aligned to normalized canonical coordinates.
- Third, the FR module is implemented with these aligned face images.




Wang, M., & Deng, W. (2021). Deep face recognition: A survey. *Neurocomputing*, *429*, 215-244.

## Deep FR pipeline

- Before a face image is fed to an FR module, face anti-spoofing, which recognizes whether the face is live or spoofed, is applied to avoid different types of attacks.




Wang, M., & Deng, W. (2021). Deep face recognition: A survey. *Neurocomputing*, *429*, 215-244.

## Deep FR pipeline

- FR module consists of **face processing**, **deep feature extraction** and **face matching**, and it can be described as follows:

$$M\left[F(P_i(I_i)), F\left(P_j\left(I_j\right)\right)\right]$$

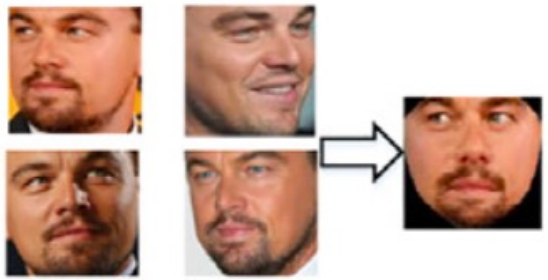
- where $I_i$ and $I_j$ are two face images, respectively. $P$ stands for **face processing** to handle intra-personal variations before training and testing, such as poses, illuminations, expressions and occlusions.
- $F$ denotes **feature extraction**, which encodes the identity information.
- $M$ means a **face matching** algorithm used to compute similarity scores of features to determine the specific identity of faces.

## Face processing

- ***One-to-many augmentation***. These methods generate many patches or images of the pose variability from a single image to enable deep networks to learn pose-invariant representations.

a) one to many

- ***Many-to-one normalization***. These methods recover the canonical view of face images from one or many images of a non-frontal view; then, FR can be performed as if it were under controlled conditions.

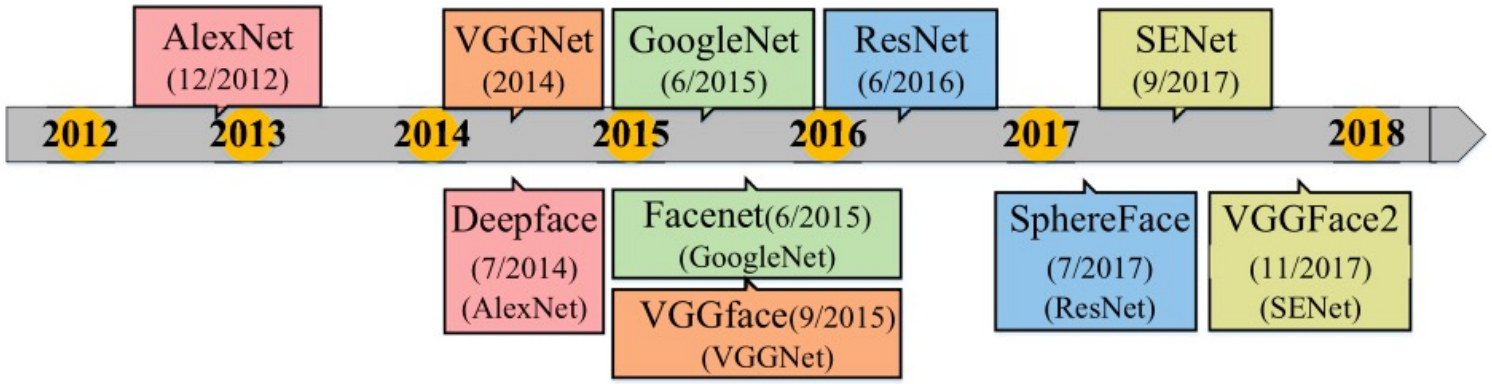


b) many to one

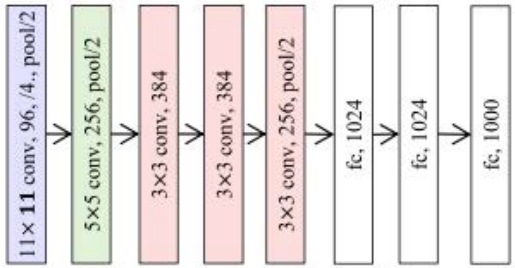
## Deep features extraction

- The architectures can be categorized as backbone and assembled networks

| Network Architectures | Subsettings |
|---|---|
| backbone network | mainstream architectures: AlexNet [80,81,38], VGGNet [37,47,82], GoogleNet [83,38], ResNet [84,82], SENet [39] |
| | light-weight architectures [85,86,61,87] |
| | adaptive architectures [88–90] |
| | joint alignment-recognition architectures [91–94] |
| assembled networks | multipose [95–98], multipatch [58–60,99,34,21,35], multitask [100] |

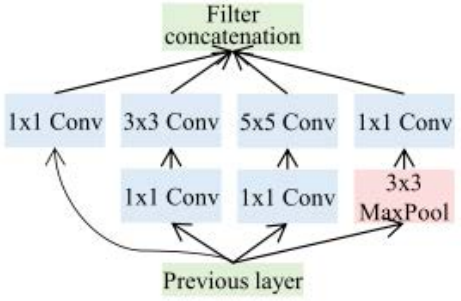


## Deep features extraction

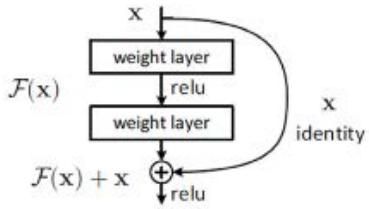



(a) Alexnet

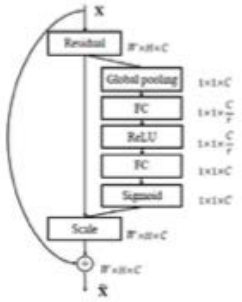

(b) VGGNet




(c) GoogleNet




(d) ResNet



(e) SENet

# Face Recognition

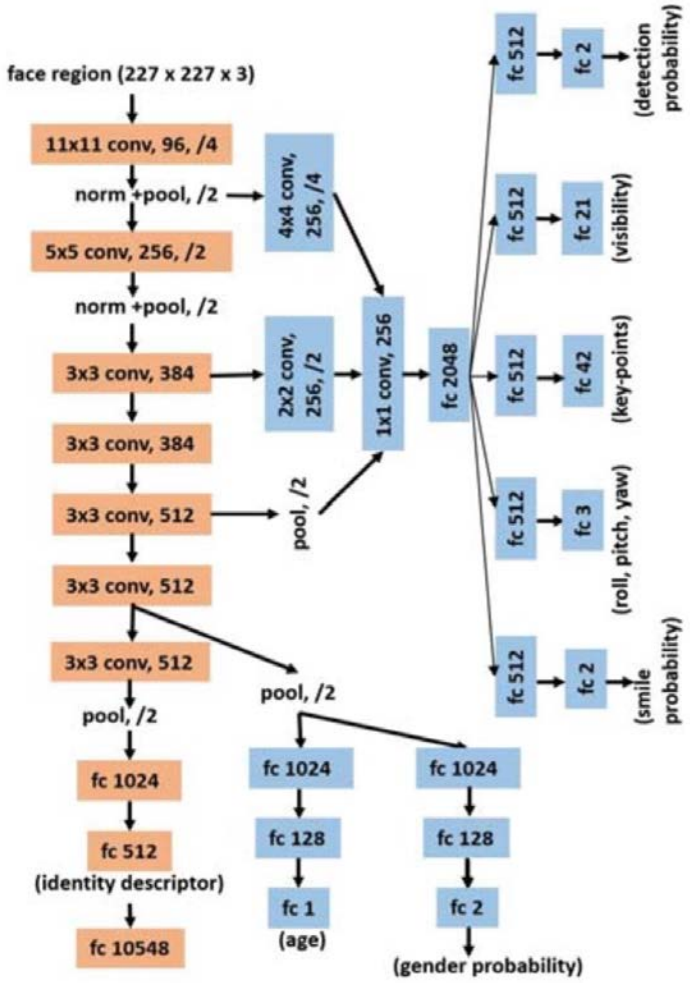
## Pipeline

## All-In-One Face

A multi-purpose CNN which can simultaneously detect faces, extract key-points and pose angles, determine smile expression and gender and estimate age from any unconstrained image of a face.

Additionally, it assigns an identity descriptor to each face which can be used for face recognition and verification.

Orange represents the pre-trained network from Sankaranarayanan et al., while blue represents added layers for MTL.

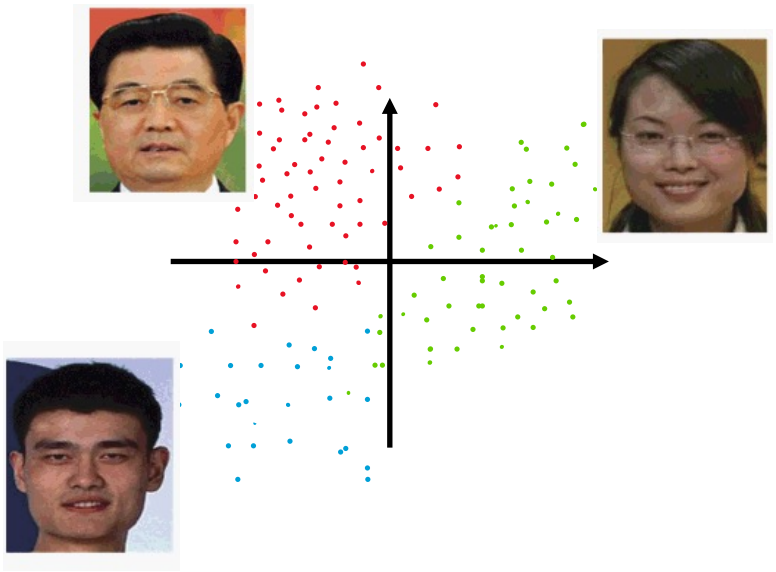



- See **additional materials** on Ranjan, R., Sankaranarayanan, S., Castillo, C. D., & Chellappa, R. (2017, May). An all-in-one convolutional neural network for face analysis. In 2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017) (pp. 17-24). IEEE.
- Sankaranarayanan, S., Alavi, A., Castillo, C. D., & Chellappa, R. (2016, September). Triplet probabilistic embedding for face verification and clustering. In 2016 IEEE 8th international conference on biometrics theory, applications and systems (BTAS) (pp. 1-8). IEEE.

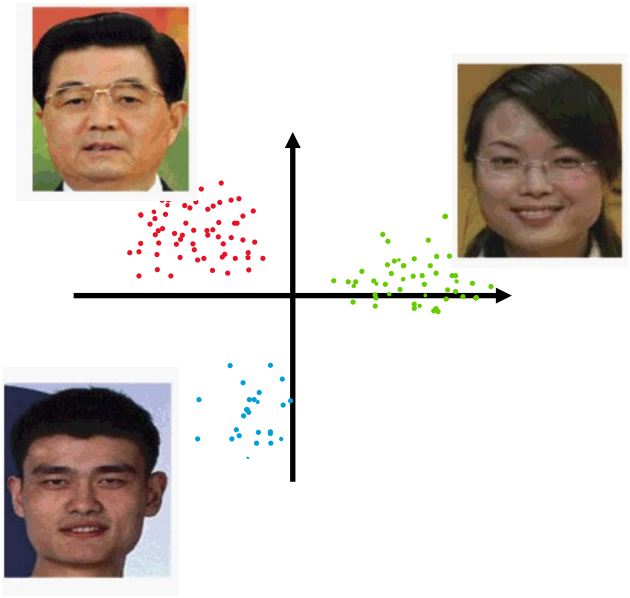
# Metric learning

## Embedding space

- The loss function for face recognition should **map samples** into a **feature space** so that samples from the **same identity** are **more compact** and samples from **different identities** are **more separated**.



Cross Entropy



Other losses

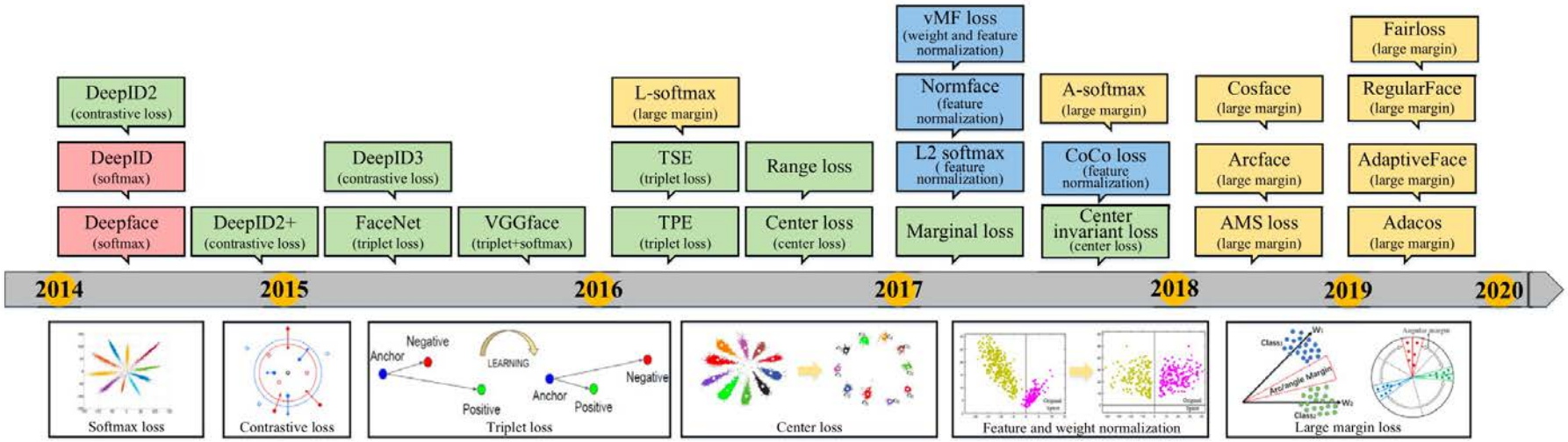
## Loss function

- The softmax loss is not sufficiently effective for FR because intra-variations could be larger than inter-differences and more discriminative features are required when recognizing different people.

| Loss Functions | Brief Description |
|---|---|
| Euclidean-distance-based loss | These methods reduce intra-variance and enlarge inter-variance based on Euclidean distance. [21,35,36,101,102,82,38,37,80,81,58,103] |
| angular/cosine-margin-based loss | These methods make learned features potentially separable with larger angular/cosine distance. [104,84,105–108] |
| softmax loss and its variations | These methods modify the softmax loss to improve performance, e.g. features or weights normalization. [109–115] |

# Face Recognition

## Losses in Face Detection




- Red, green, blue and yellow rectangles represent deep methods using softmax, Euclidean-distance-based loss, angular/cosine-margin- based loss and variations of softmax, respectively.

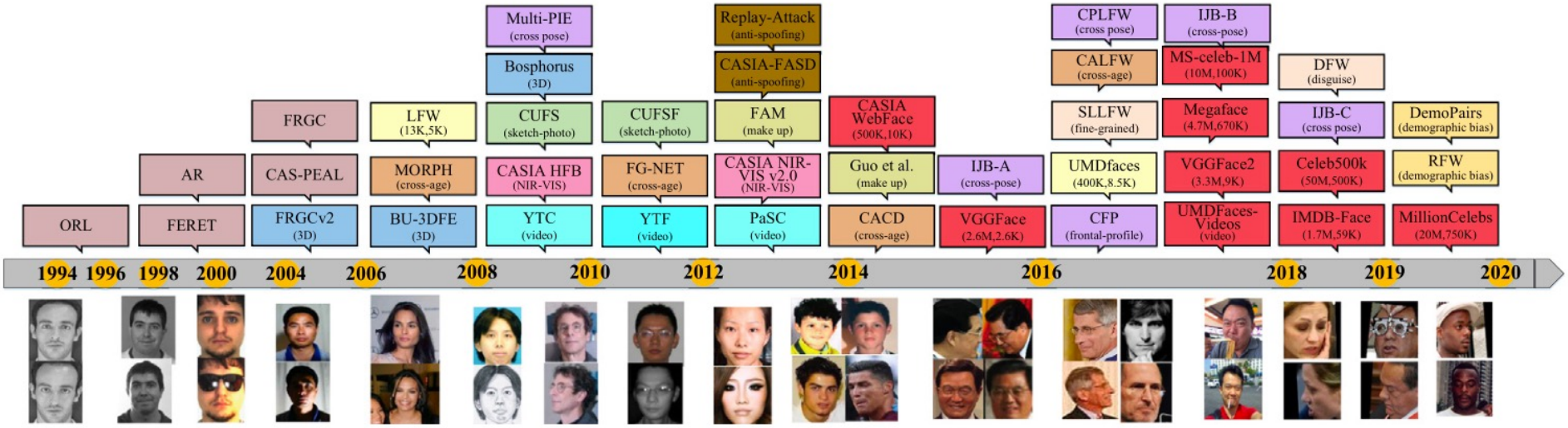
* See **additional materials** on Minaee, S., Luo, P., Lin, Z., & Bowyer, K. (2021). Going deeper into face detection: A survey. *arXiv preprint arXiv:2103.14983*.

# Face Recognition task

## Dataset

The evolution of FR datasets.

Before 2007, early works in FR focused on controlled and small-scale datasets. In 2007, LFW dataset was introduced which marks the beginning of FR under unconstrained conditions. In 2014, CASIA-Webface provided the first widely-used public training dataset.

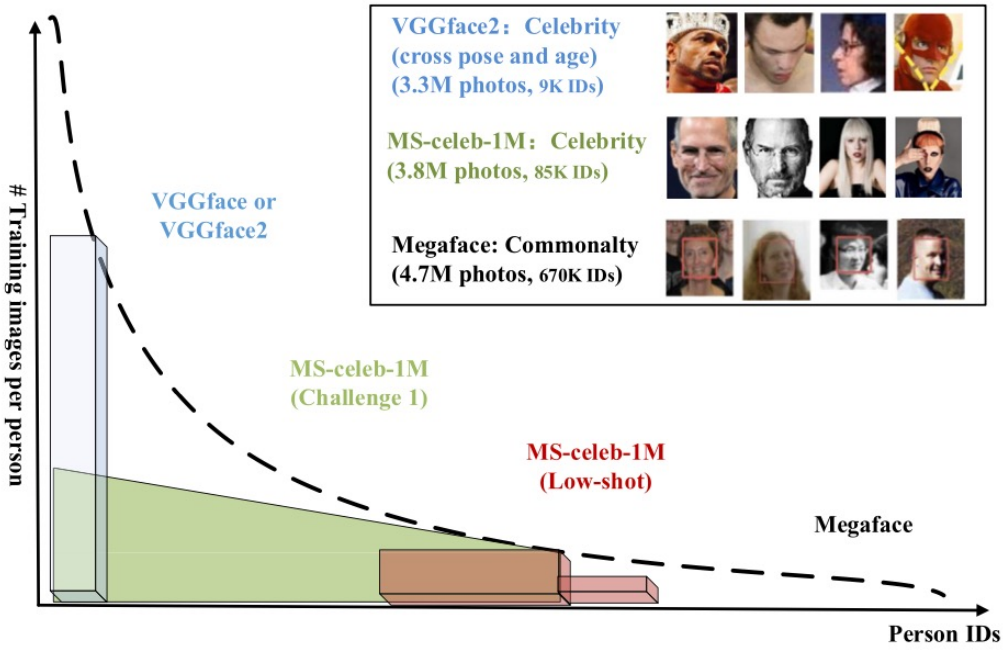



Red rectangles represent training datasets, and other color rectangles represent different testing datasets.

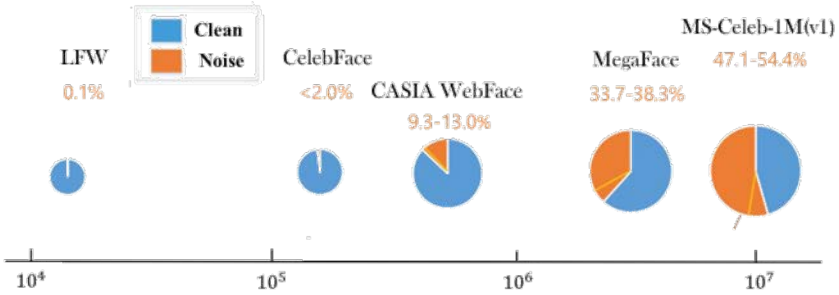
# Face Recognition task

## Dataset

### Data analysis




Estimated noise percentage of datasets.



# Face Recognition task

## LFW Database

Labeled Faces in the Wild Home

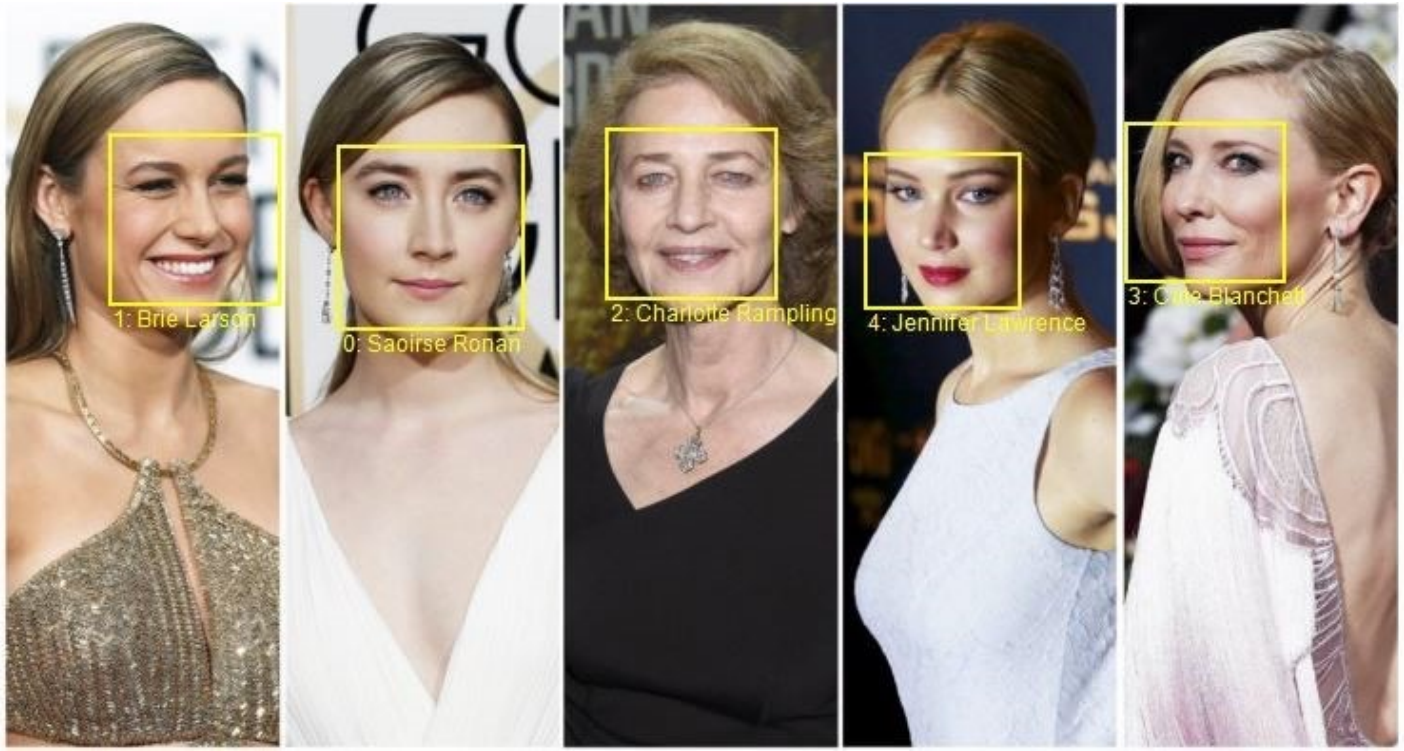
- 13k images from the web
- 1680 persons have >= 2 photos

* See **additional materials** on https://vis-www.cs.umass.edu/lfw/

# Face Recognition task

## MSCeleb Database

- Top 100k celebrities
- 10 Million images, 100 per person
- Noisy: constructed by leveraging public search engines




* See **additional materials** on https://www.microsoft.com/en-us/research/project/ms-celeb-1m-challenge-recognizing-one-million-celebrities-real-world/

# Image Recognition (attributes)

# Image Quality (IQ)

## definition

**Image quality (IQ)** is a characteristic of an image that measures the perceived image degradation due to imaging artifacts.

### IMAGE QUALITY FACTOR

Sharpness, noise, dynamic range, tone reproduction, contrast, color accuracy, distortion, vignetting, lateral chromatic aberration, lens are, color moiré, artifacts, etc.

# Image Quality (IQ)

## definition

**Image quality (IQ)** is a characteristic of an image that measures the perceived image degradation due to imaging artifacts.

### IMAGE QUALITY FACTOR

Sharpness, noise, dynamic range, tone reproduction, contrast, color accuracy, distortion, vignetting, lateral chromatic aberration, lens are, color moiré, artifacts, etc.

# Image Quality Assessment

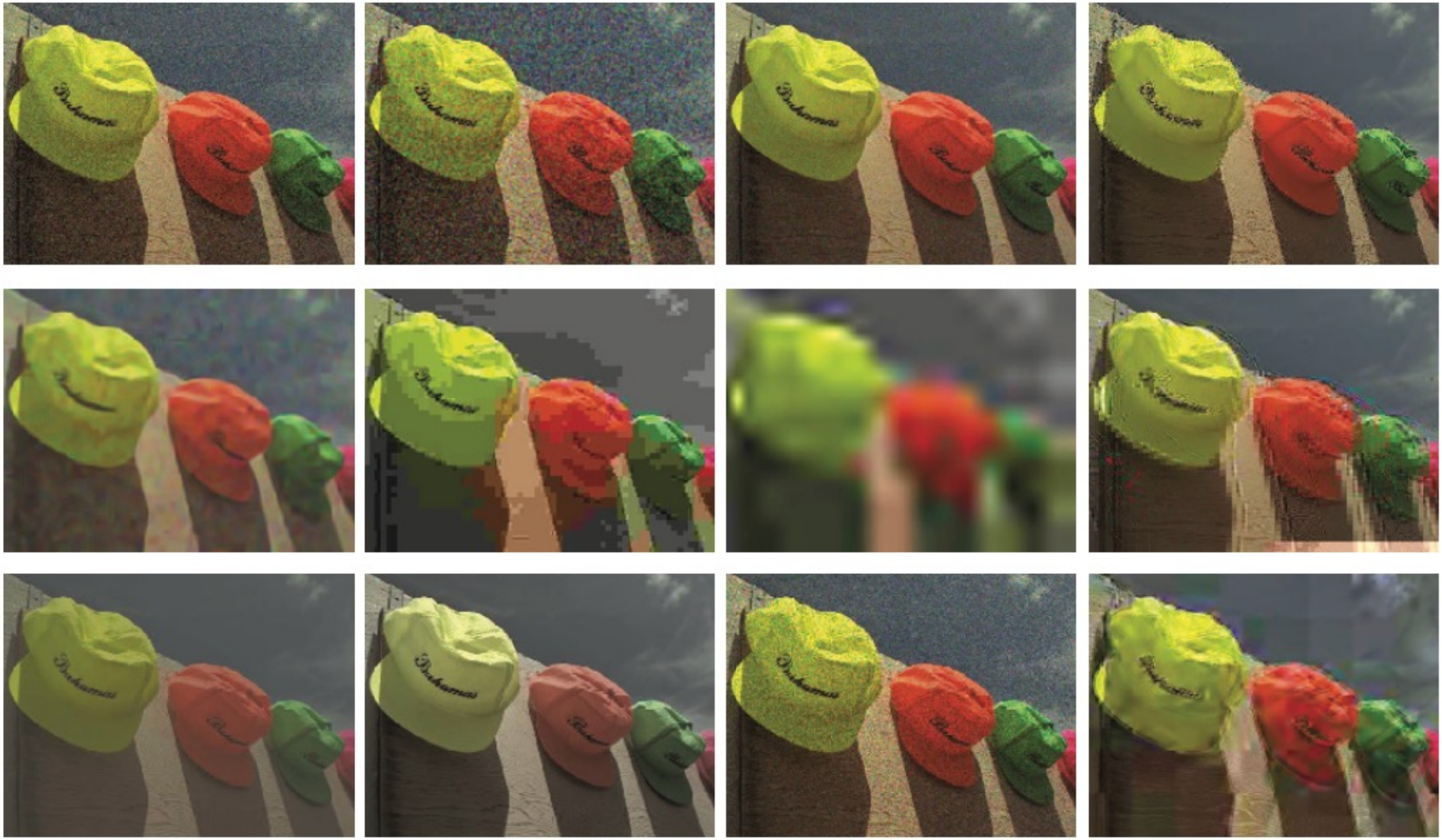
## Definition

- Image quality is a characteristic of an image that measures the **perceived image degradation**
- It plays an important role in various image processing application.
- Goal of image quality assessment is to supply quality metrics that can predict perceived image quality automatically.
- Two Types of image quality assessment
  - **Subjective quality assessmen**t
  - **Objective quality assessment**

* See **additional materials** on Zhai, G., & Min, X. (2020). Perceptual image quality assessment: a survey. Science China Information Sciences, 63, 1-52.

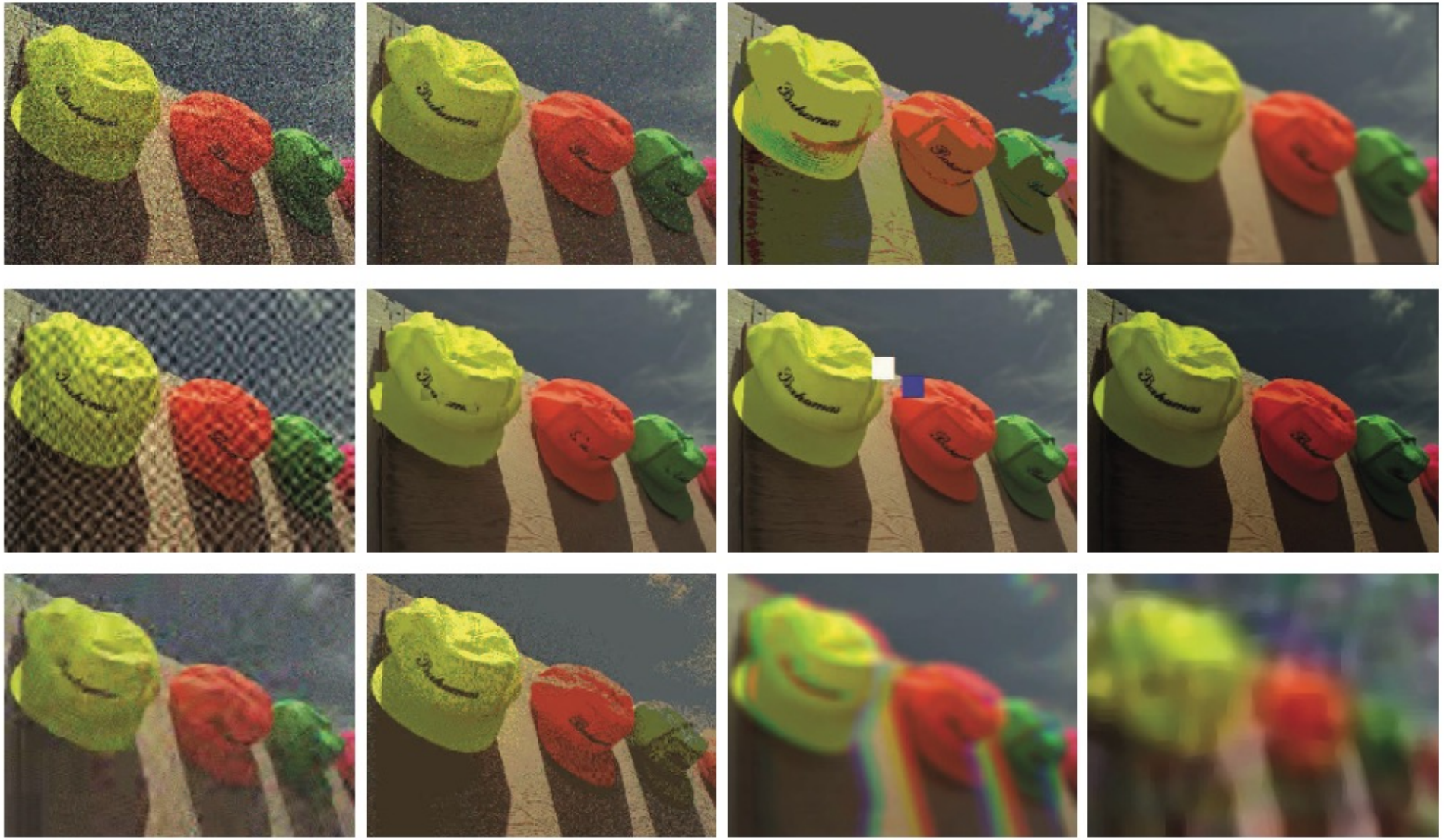
# Image Quality

## Example of distortions



# Image Quality

## Example of distortions



# Subjective Quality Measure

## Definition

- The best way to find quality of an image is to look at it because human eyes are the ultimate viewer.
- **Subjective image** quality is concerned with how image is perceived by a viewer and give his or her opinion on a particular image.
- The mean opinion score (**MOS**) has been used for subjective quality assessment from many years.
- Too **Inconvenient**, **time consuming** and **expensive**

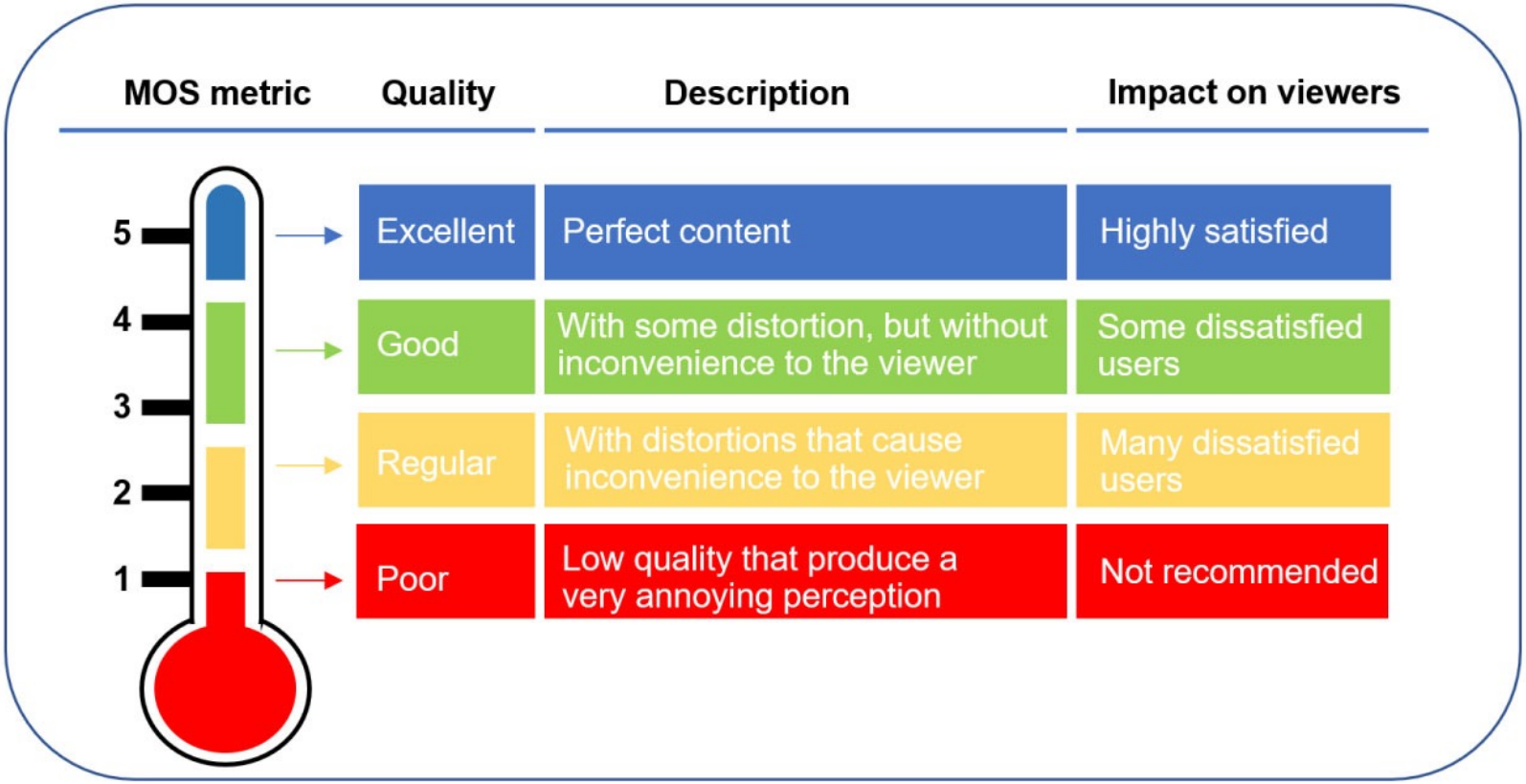
$$MOS = \frac{\sum_{n=1}^{N} R_n}{N}$$

* See **additional materials** on Zhai, G., & Min, X. (2020). Perceptual image quality assessment: a survey. Science China Information Sciences, 63, 1-52.

# MOS

## Definition

- MOS score of 1 is worst image quality and 5 is best.




* See **additional materials** on Huynh-Thu, Q.; Garcia, M. N.; Speranza, F.; Corriveau, P.; Raake, A. (2011-03-01). "Study of Rating Scales for Subjective Quality Assessment of High-Definition Video". IEEE Transactions on Broadcasting. 57 (1): 1–14

# Objective Quality Measure

## Definition

- Mathematical models that approximate results of subjective quality assessment
- The goal of **objective evaluation** is to develop a **quantitative measure** that can **predict perceived image quality**
- It plays a variety of roles
  - To monitor and **control image quality** for quality control systems
  - To **benchmark** image processing systems;
  - To **optimize** algorithms and parameters;
  - To help home users **better manage** their **digital photos** and evaluate their expertise in photographing.

* See **additional materials** on Zhai, G., & Min, X. (2020). Perceptual image quality assessment: a survey. Science China Information Sciences, 63, 1-52.

# Objective Quality Measure

## Definition

- Pearson's Linear Correlation Coefficient (**PLCC**)
- Spearman's Rank-order Correlation Coefficient (**SROCC**)
- Mean Square Error (**MSE**)
- Peak Signal-to-Noise Ratio (**PSNR**)

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Where,

r = Pearson Correlation Coefficient

$x_i$ = x variable samples

$y_i$ = y variable sample

$\bar{x}$ = mean of values in x variable

$\bar{y}$ =mean of values in y variable

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

$\rho$ = Spearman's rank correlation coefficient

$d_i$ = difference between the two ranks of each observation

$n$ = number of observations

$$MSE = \frac{1}{m\,n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2$$

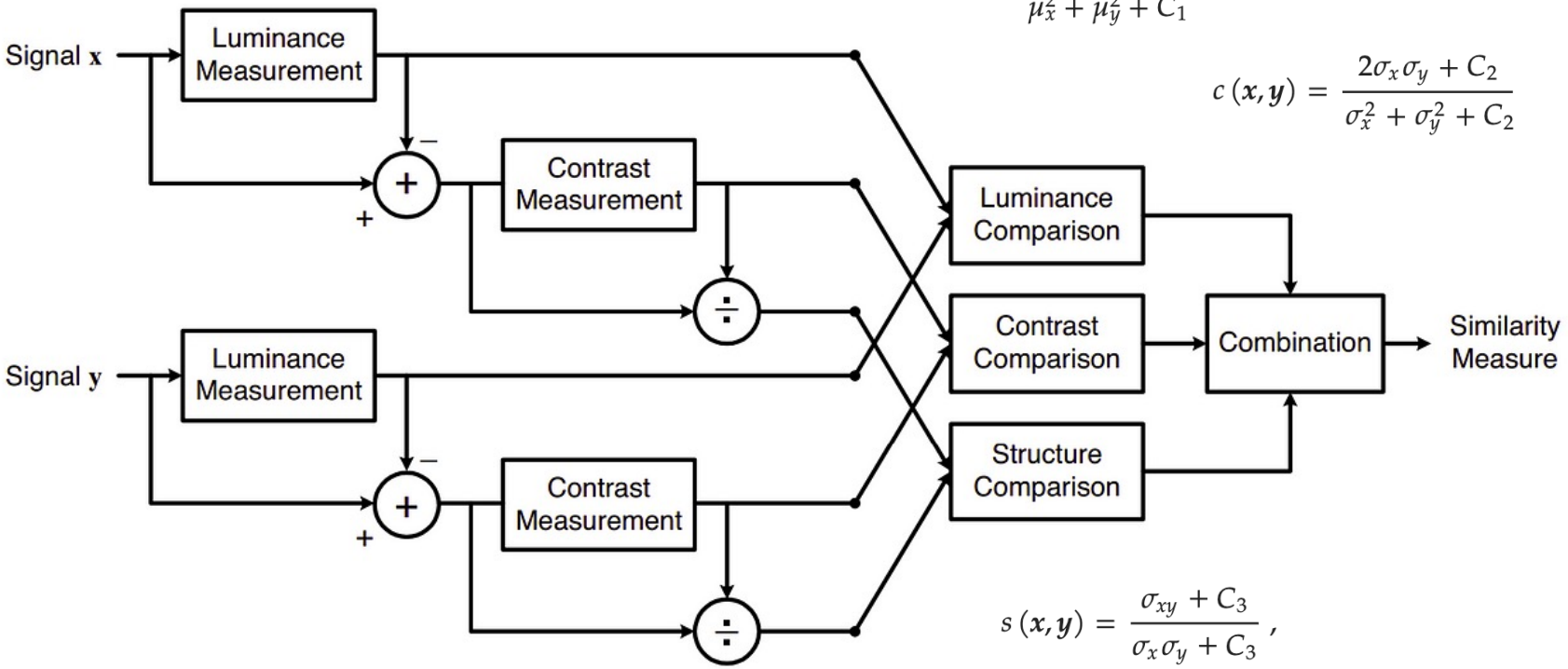
$$PSNR = 10 \cdot \log_{10} \left( \frac{MAX_I^2}{MSE} \right)$$
$$= 20 \cdot \log_{10} \left( \frac{MAX_I}{\sqrt{MSE}} \right)$$

## SSIM – Structural similarity index

The SSIM is designed to improve on traditional metrics like PSNR and MSE, which have proved to be inconsistent with human eye perception. Based on human visual system

$$l(\boldsymbol{x},\boldsymbol{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$

$$c(\boldsymbol{x},\boldsymbol{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$



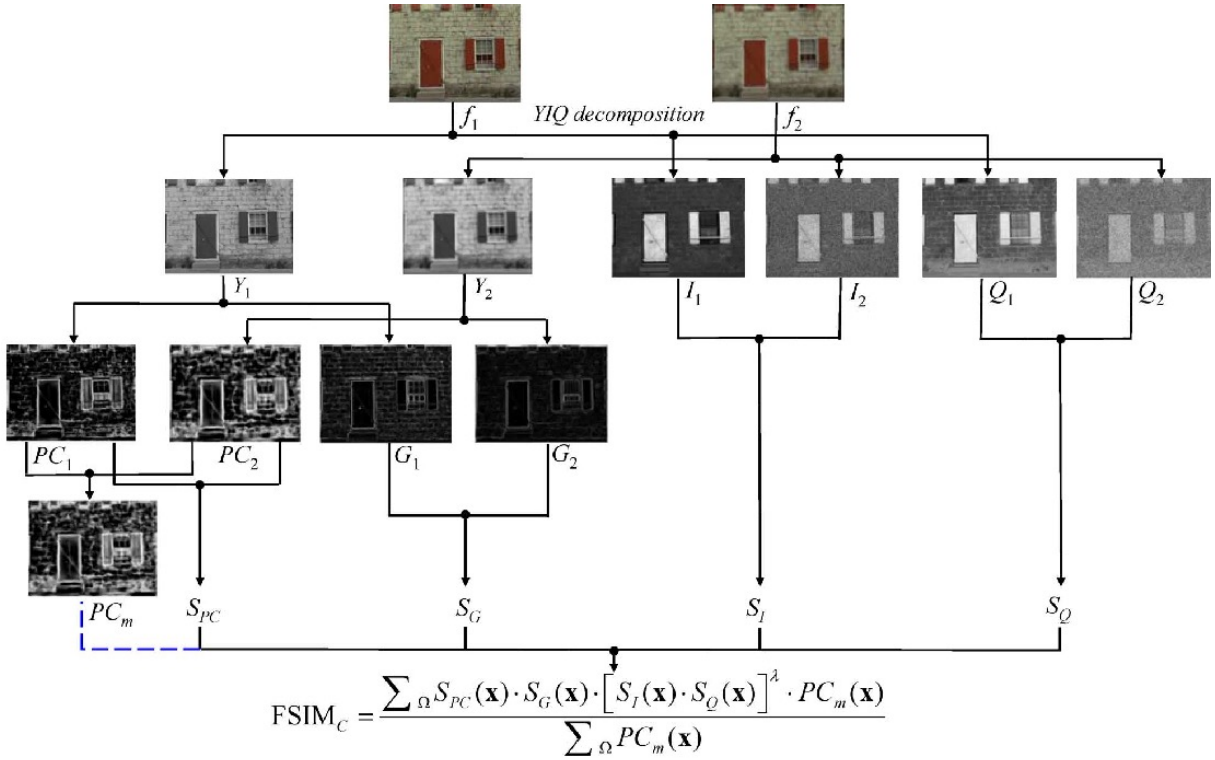

$$s(\boldsymbol{x},\boldsymbol{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3},$$

# Objective Quality Measure

## SSIM

### FSIM (Feature Similarity Indexing Method)

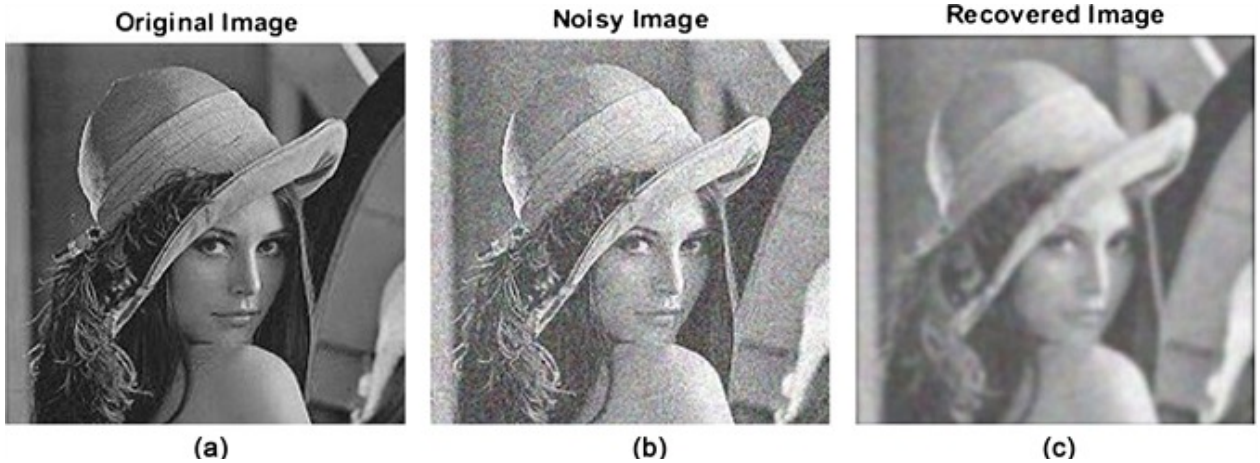
Feature Similarity Index Method maps the features and measures the similarities between two images. To describe FSIM we need to describe two criteria more clearly. They are: Phase Congruency (PC) and Gradient Magnitude (GM).



# Objective Quality Measure

## SSIM – MSE – PSNR

### Comparison

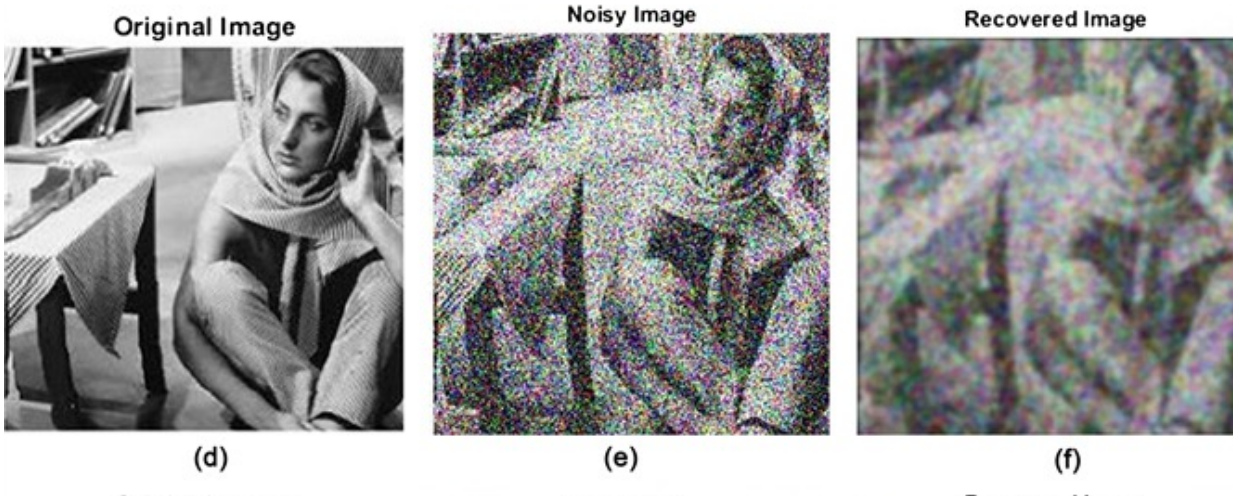



| Image | Noise Level (Gaussian-variance) | Quality Assessment Techniques | | | |
|---|---|---|---|---|---|
| | | MSE | PSNR | SSIM | FSIM |
| Lena | 0.2 | 21.56 | 21.54 | 0.78 | 0.89 |
| | 0.4 | 16.81 | 16.81 | 0.74 | 0.86 |
| | 0.6 | 14.18 | 14.18 | 0.7 | 0.8 |
| Barbara | 0.2 | 21.95 | 21.95 | 0.8 | 0.88 |
| | 0.4 | 17.8 | 17.79 | 0.75 | 0.84 |
| | 0.6 | 15.49 | 15.48 | 0.7 | 0.8 |
| Cameramen | 0.2 | 21.64 | 21.65 | 0.73 | 0.84 |
| | 0.4 | 17.29 | 17.31 | 0.76 | 0.86 |
| | 0.6 | 15.27 | 15.28 | 0.76 | 0.87 |

# Objective Quality Measure

## SSIM – MSE – PSNR

### Comparison

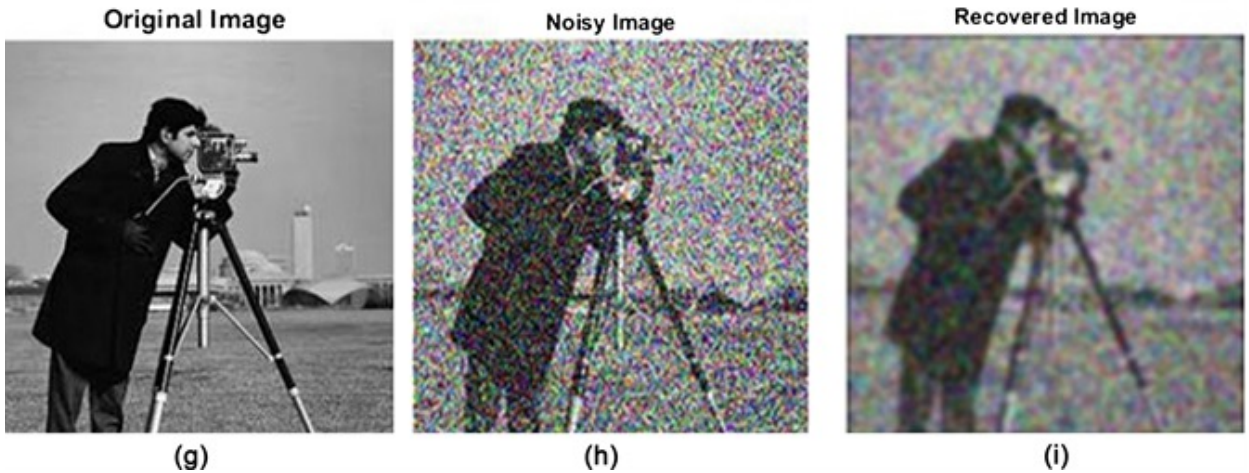



| Image | Noise Level (Gaussian-variance) | Quality Assessment Techniques | | | |
|---|---|---|---|---|---|
| | | MSE | PSNR | SSIM | FSIM |
| Lena | 0.2 | 21.56 | 21.54 | 0.78 | 0.89 |
| | 0.4 | 16.81 | 16.81 | 0.74 | 0.86 |
| | 0.6 | 14.18 | 14.18 | 0.7 | 0.8 |
| Barbara | 0.2 | 21.95 | 21.95 | 0.8 | 0.88 |
| | 0.4 | 17.8 | 17.79 | 0.75 | 0.84 |
| | 0.6 | 15.49 | 15.48 | 0.7 | 0.8 |
| Cameramen | 0.2 | 21.64 | 21.65 | 0.73 | 0.84 |
| | 0.4 | 17.29 | 17.31 | 0.76 | 0.86 |
| | 0.6 | 15.27 | 15.28 | 0.76 | 0.87 |

# Objective Quality Measure

## SSIM – MSE – PSNR

### Comparison

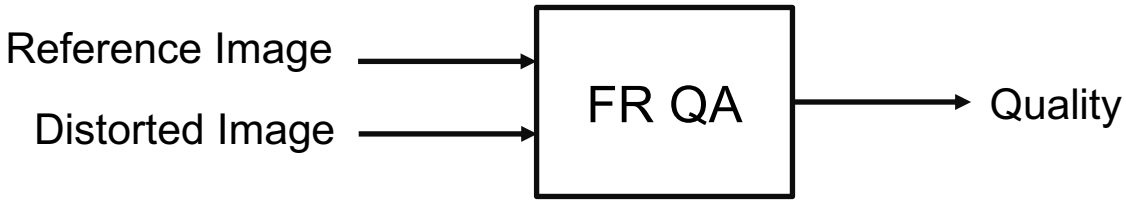



| Image | Noise Level (Gaussian-variance) | Quality Assessment Techniques | | | |
|---|---|---|---|---|---|
| | | MSE | PSNR | SSIM | FSIM |
| Lena | 0.2 | 21.56 | 21.54 | 0.78 | 0.89 |
| | 0.4 | 16.81 | 16.81 | 0.74 | 0.86 |
| | 0.6 | 14.18 | 14.18 | 0.7 | 0.8 |
| Barbara | 0.2 | 21.95 | 21.95 | 0.8 | 0.88 |
| | 0.4 | 17.8 | 17.79 | 0.75 | 0.84 |
| | 0.6 | 15.49 | 15.48 | 0.7 | 0.8 |
| Cameramen | 0.2 | 21.64 | 21.65 | 0.73 | 0.84 |
| | 0.4 | 17.29 | 17.31 | 0.76 | 0.86 |
| | 0.6 | 15.27 | 15.28 | 0.76 | 0.87 |

# Image Quality Assessment

## IQA categories

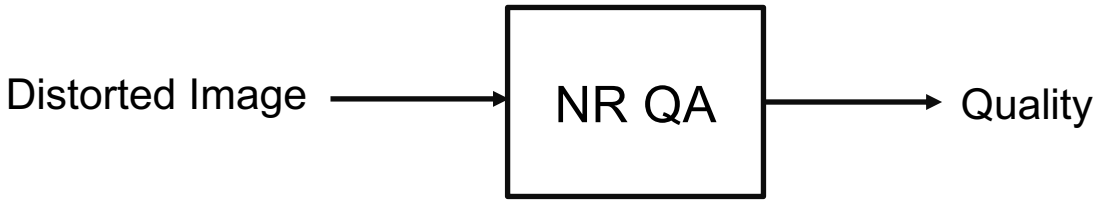
- **Full-Reference** (FR) QA
- No-Reference (NR) QA
- Reduced-Reference (RR) QA



# Image Quality Assessment

## IQA categories

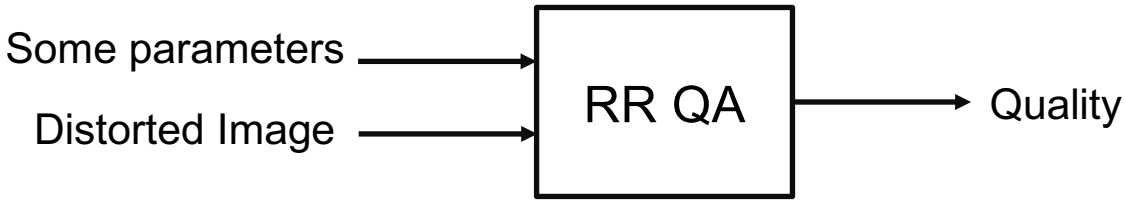
- Full-Reference (FR) QA
- **No-Reference** (NR) QA – Blind
- Reduced-Reference (RR) QA



# Image Quality Assessment

## IQA categories

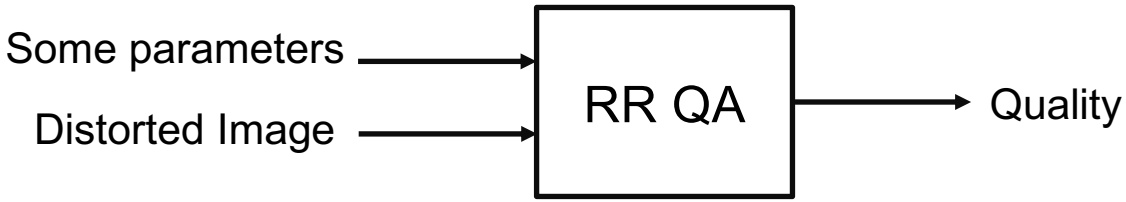
- Full-Reference (FR) QA
- No-Reference (NR) QA
- **Reduced-Reference** (RR) QA



# Image Quality Assessment

## IQA categories

- Full-Reference (FR) QA
- No-Reference (NR) QA
- **Reduced-Reference** (RR) QA



# Image Quality Assessment

## IQA categories

### LIVE In the Wild Image Quality Challenge Database*

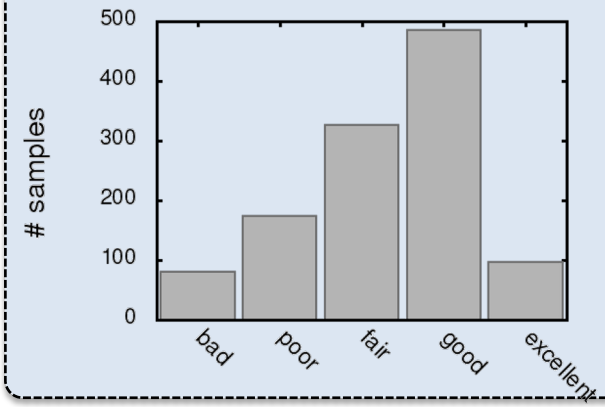
- 1,162 images affected by diverse authentic distortions and genuine artifacts;
- Over 350,000 opinion scores from over 8,100 unique human observers;
- The **mean opinion score (MOS)** of each image is the average of individual ratings (a number ranging from 0 to 100) across subjects.

**PREVIOUS DATABASES**

By adding synthetic distortions to high-quality images:
LIVE, CSIQ, TID2008 and TID2013

Ghadiyaram and Bovik:

*"images captured using typical real-world mobile camera devices are usually afflicted by complex mixtures of multiple distortions, which are not necessarily well-modeled by the synthetic distortions found in existing databases"*

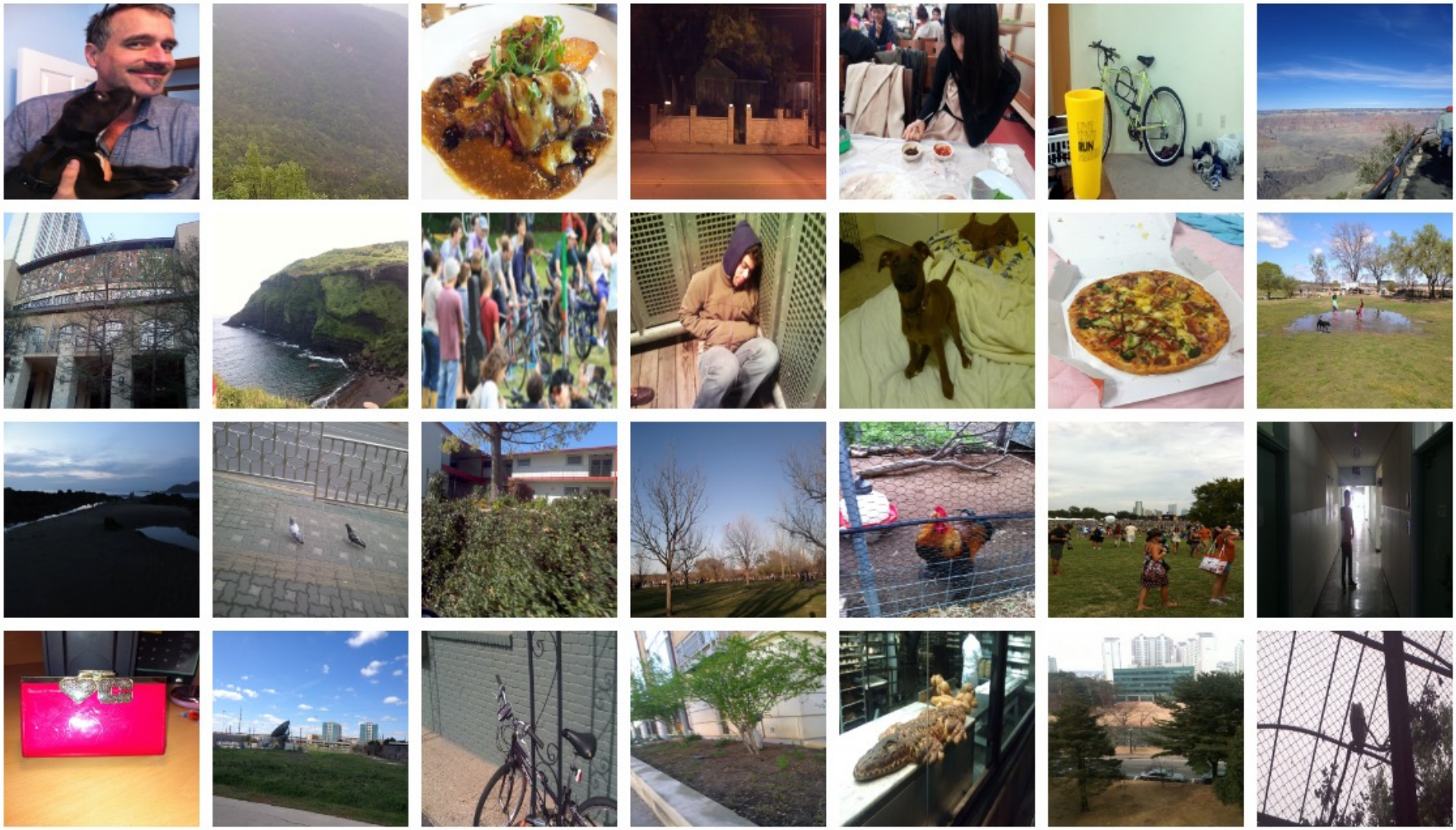



* Ghadiyaram, Deepti, and Alan C. Bovik. "***Massive online crowdsourced study of subjective and objective picture quality.***" IEEE Transactions on Image Processing 25.1 (2016): 372-387.

# Image Quality Assessment

## IQA categories

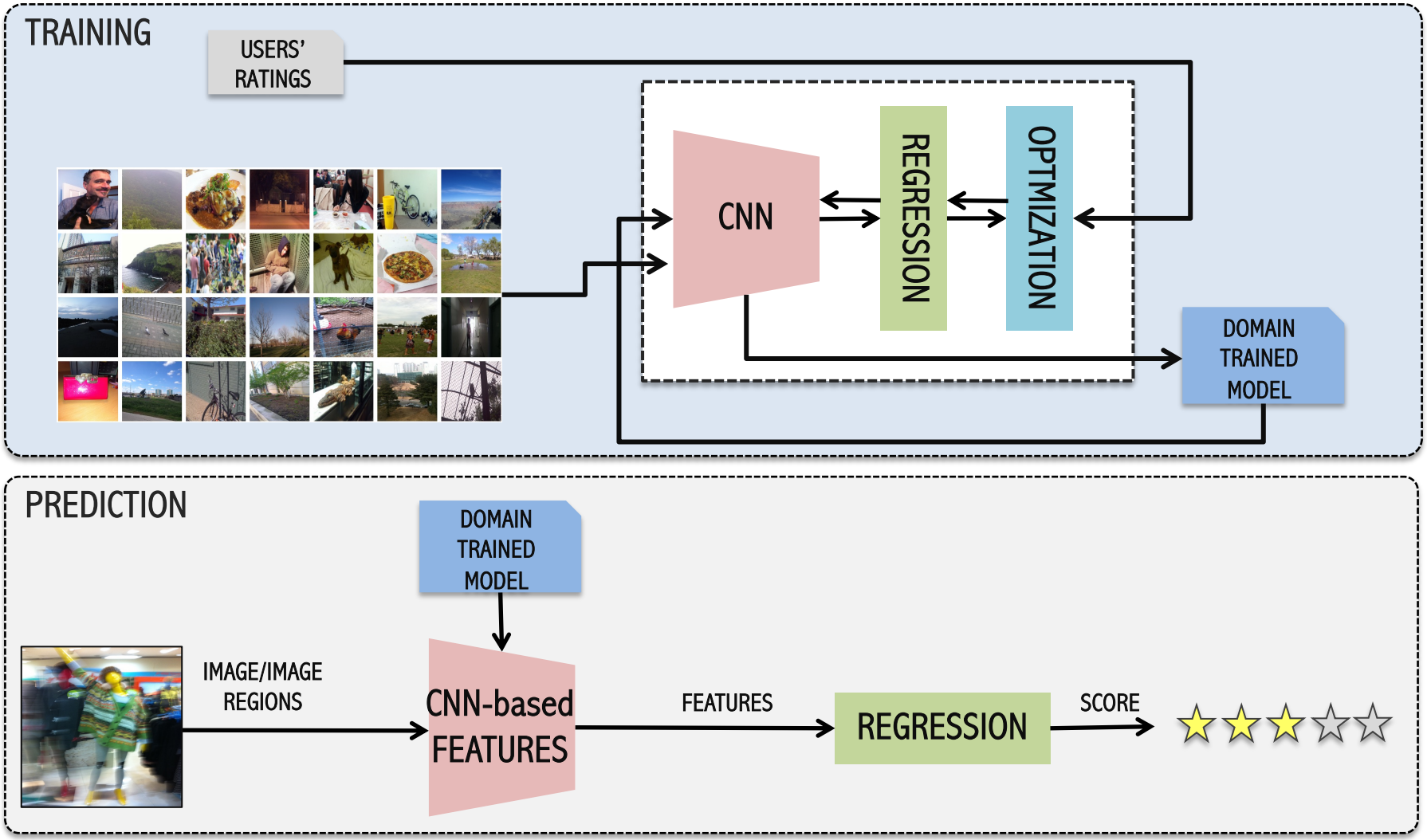
LIVE In the Wild Image Quality Challenge Database*



# No-Reference Image Quality Assessment

## Pipeline

- **Non-Reference** (NR) QA



# No-Reference Image Quality Assessment

## Pipeline

- **Non-Reference** (NR) QA

(a)
QS: 67.94 MOS: 75.69

(b)
QS: 63.91 MOS: 71.27

(c)
QS: 73.37 MOS: 82.67

(d)
QS 42.71 MOS: 16.13

(e)
QS: 44.33 MOS: 52.64

(f)
QS: 42.79 MOS: 42.79

# Image Aesthetics (IAE)

## definition

**Photography** is the art or practice of taking and processing photographs.
**Image Aesthetics** is how people usually characterize beauty in this form of art.


### IMAGE AESTHETICS FACTOR

Humans' aesthetic evaluation is subjective. Many research proposed various approaches that can be divided into two groups: aesthetic classification and aesthetic regression.

# Image Aesthetics (IAE)

## definition

**Photography** is the art or practice of taking and processing photographs.
**Image Aesthetics** is how people usually characterize beauty in this form of art.


### IMAGE AESTHETICS FACTOR

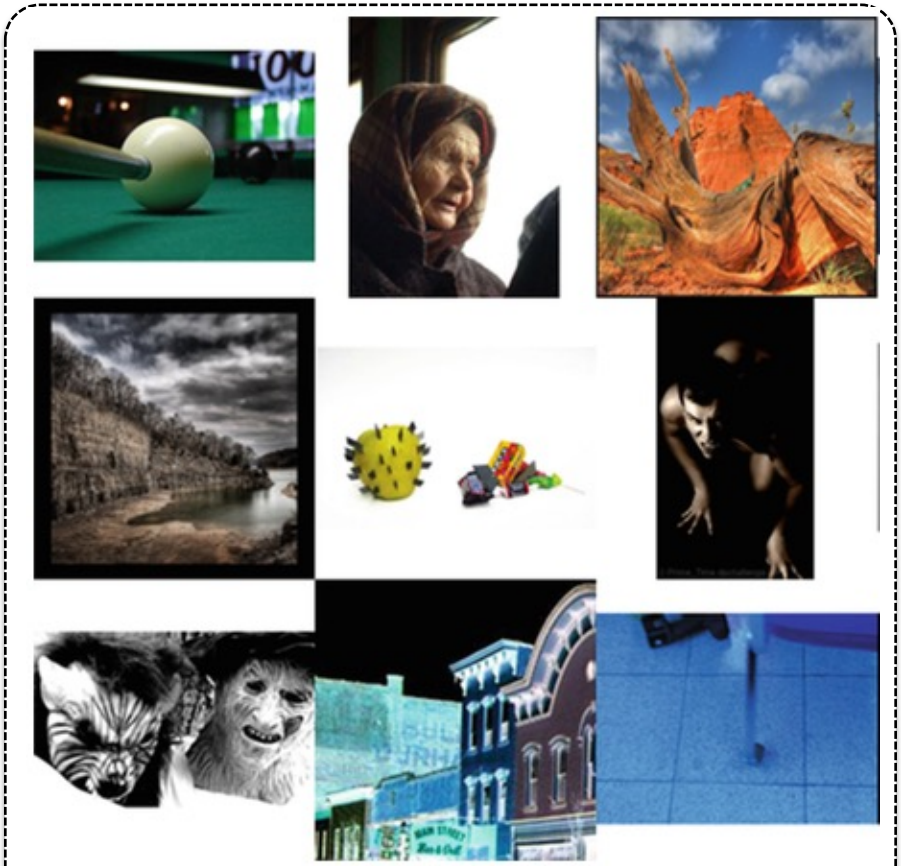
Humans' aesthetic evaluation is subjective. Many research proposed various approaches that can be divided into two groups: aesthetic classification and aesthetic regression.

# Image Aesthetics (IAE)

## Pipeline

AVA Database*:

- 255,000 images and meta-data obtained from the on-line community of photography amateurs www.dbchallenge.com;
- a **wide variety** of subjects;
- aesthetic **ratings** ranging from 1 to 10;
- semantic annotations consisting in **66 textual tags** describing the semantics of the images;
- photographic style annotations corresponding **to 14 photographic** techniques.



Sample images from the AVA database sorted by their aesthetic score (decreasing from top-left to bottom-right).

Murray, N., Marchesotti, L., Perronnin, F.: *Ava: a large-scale database for aesthetic visual analysis*. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2408–2415. IEEE (2012)

# Image Aesthetics (IAE)

## Results

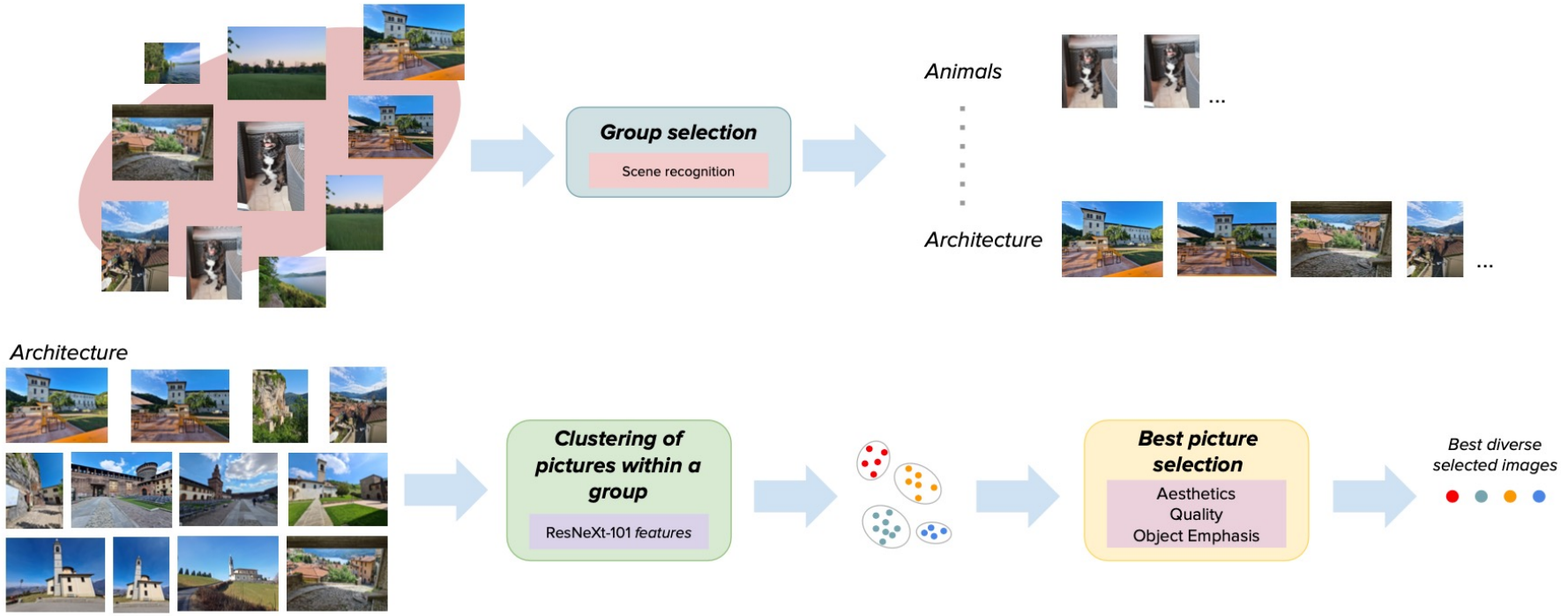
Output produced by the proposed method on sample images from the AVA dataset. For each image, the aesthetic score and the attributes predicted by the proposed method are reported (ground-truth is in brackets). “N/A” means that the dataset does not provide any style annotation for the image


Celona, L., Leonardi, M., Napoletano, P., & Rozza, A. (2022). Composition and style attributes guided image aesthetic assessment. *IEEE Transactions on Image Processing*, *31*, 5009-5024.

# Image Collection Summarization

## Pipeline

The proposed method first groups images based on their semantic content, and then selects the most diverse and aesthetically pleasing images to represent each category.




Celona, L., Leonardi, M., Napoletano, P., & Rozza, A. (2022). Composition and style attributes guided image aesthetic assessment. *IEEE Transactions on Image Processing*, *31*, 5009-5024.

# Summing up

# Summing up

## datasets

https://pollev.com/paolonapoletano587

# QUESTIONS?