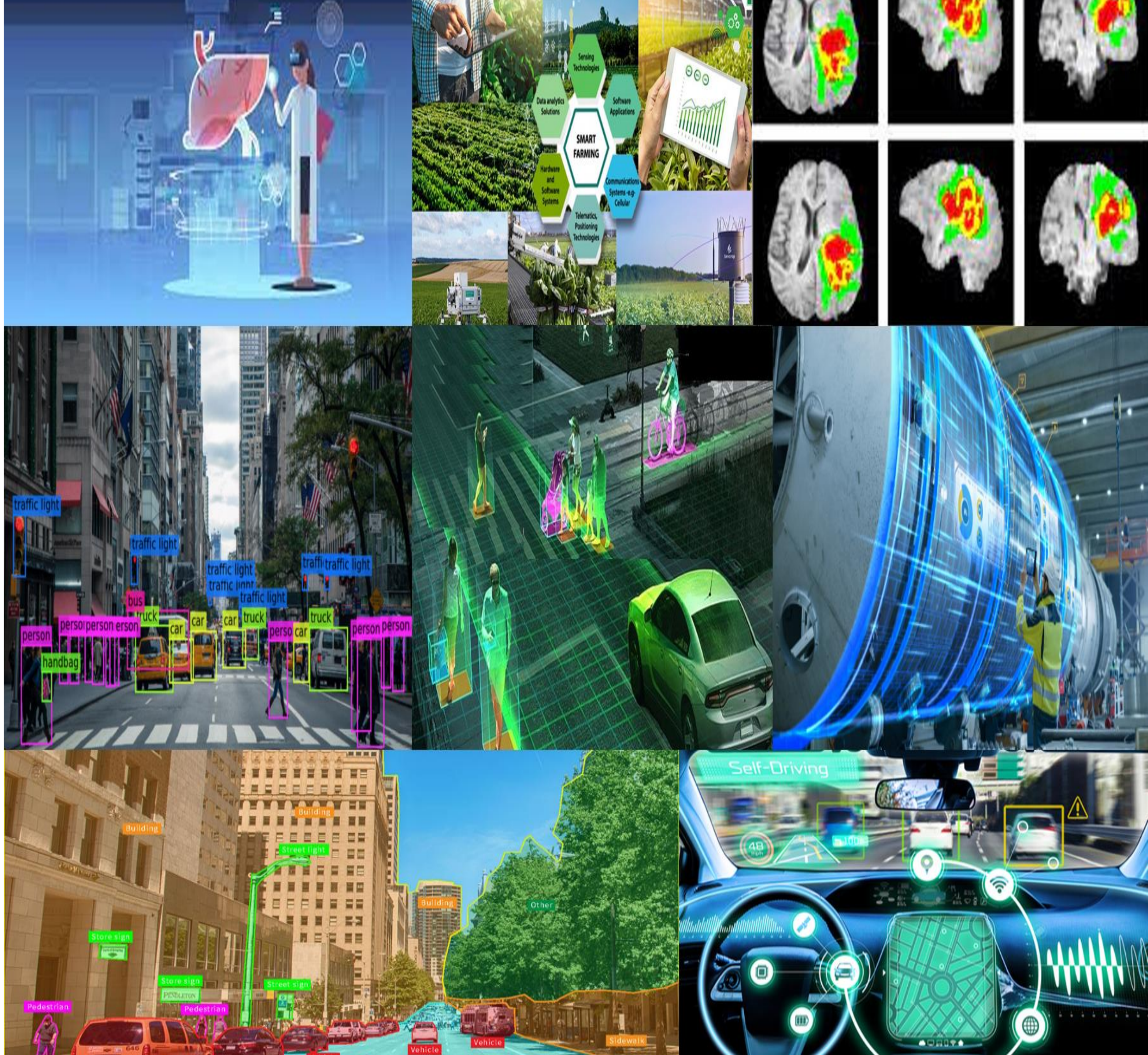
ECE 595: Image Processing and Computer Vision with Deep Learning

Dr. Jiji C. V.

Professor, Shiv Nadar University
Chennai, India

27/06/2023



Contents

- Introduction
- Image Processing- conventional solutions
- Computer Vision – conventional solutions
- Why DL for Image Processing and Computer Vision
- What all we cover in this course ?
  - CNN
  - GAN
  - Transformers
- Image Restoration and Computer Vision with Deep Learning
- Lab sessions

Image Restoration

- Degraded input image: Clean output image
- Image Enhancement and Image Restoration
- Image Enhancement : subjective
  - Intensity transformations
  - Contrast enhancement
  - Histogram processing
  - Spatial filtering

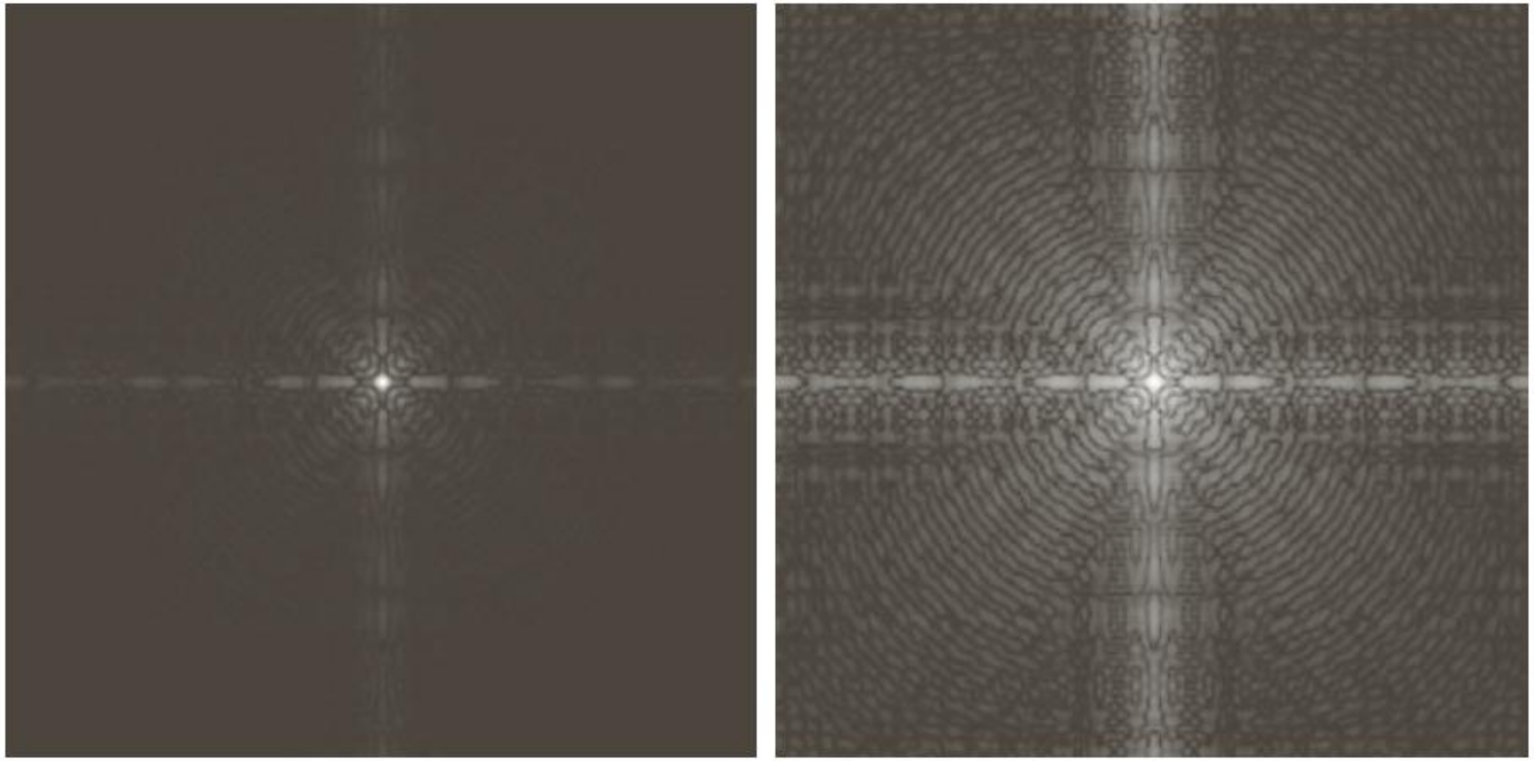
Example: Image Negatives

a b

FIGURE 3.4
(a) Original digital mammogram.
(b) Negative image obtained using the negative transformation in Eq. (3.2-1).
(Courtesy of G.E. Medical Systems.)

Small lesion

Example: Log Transformations

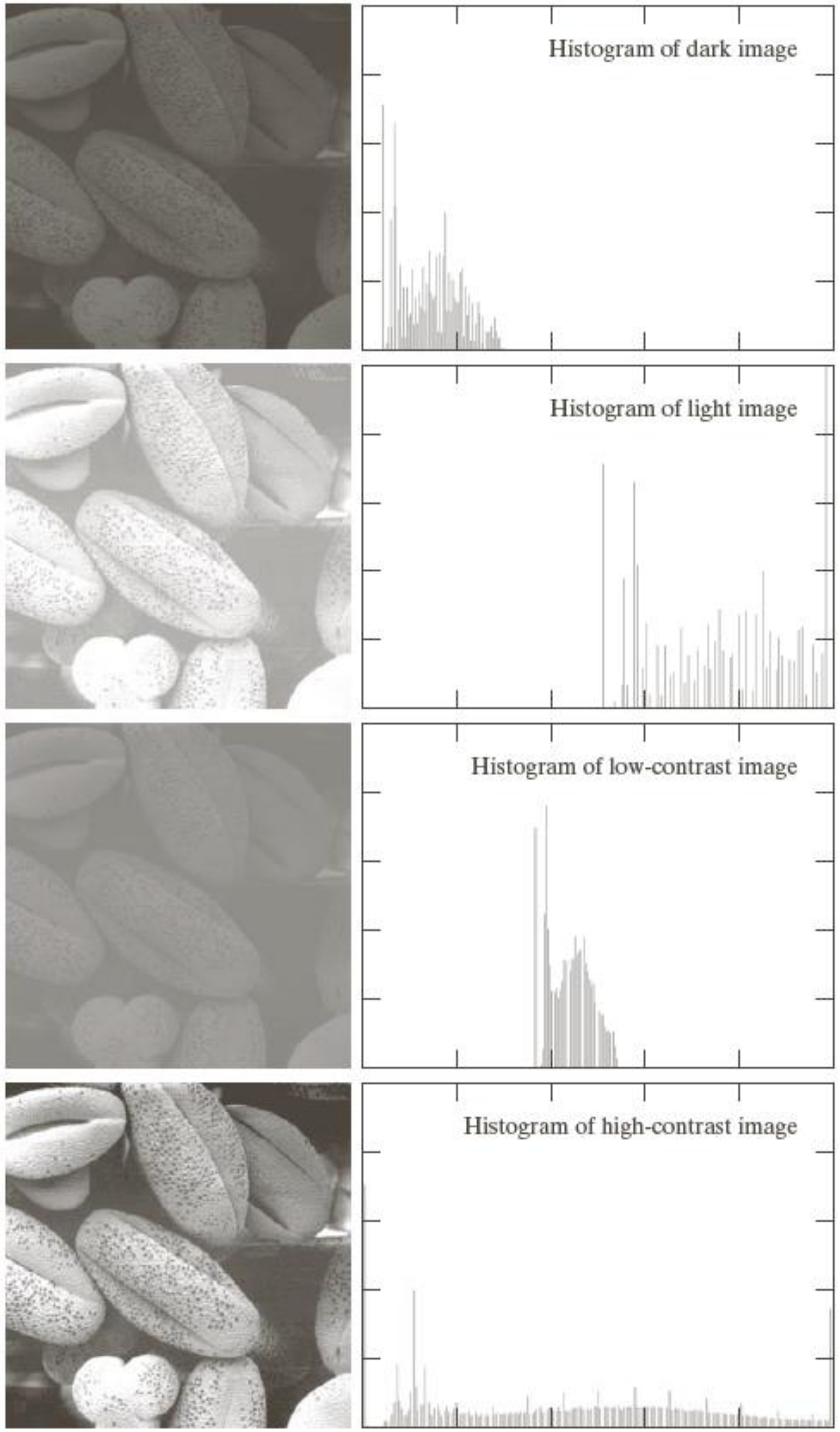


a b

FIGURE 3.5
(a) Fourier spectrum.
(b) Result of applying the log transformation in Eq. (3.2-2) with $c = 1$.

a	b
c	d

FIGURE 3.10 Contrast stretching. (a) Form of transformation function. (b) A low-contrast image. (c) Result of contrast stretching. (d) Result of thresholding. (Original image courtesy of Dr. Roger Heady, Research School of Biological Sciences, Australian National University, Canberra, Australia.)


Histogram of dark image
Histogram of light image
Histogram of low-contrast image
Histogram of high-contrast image

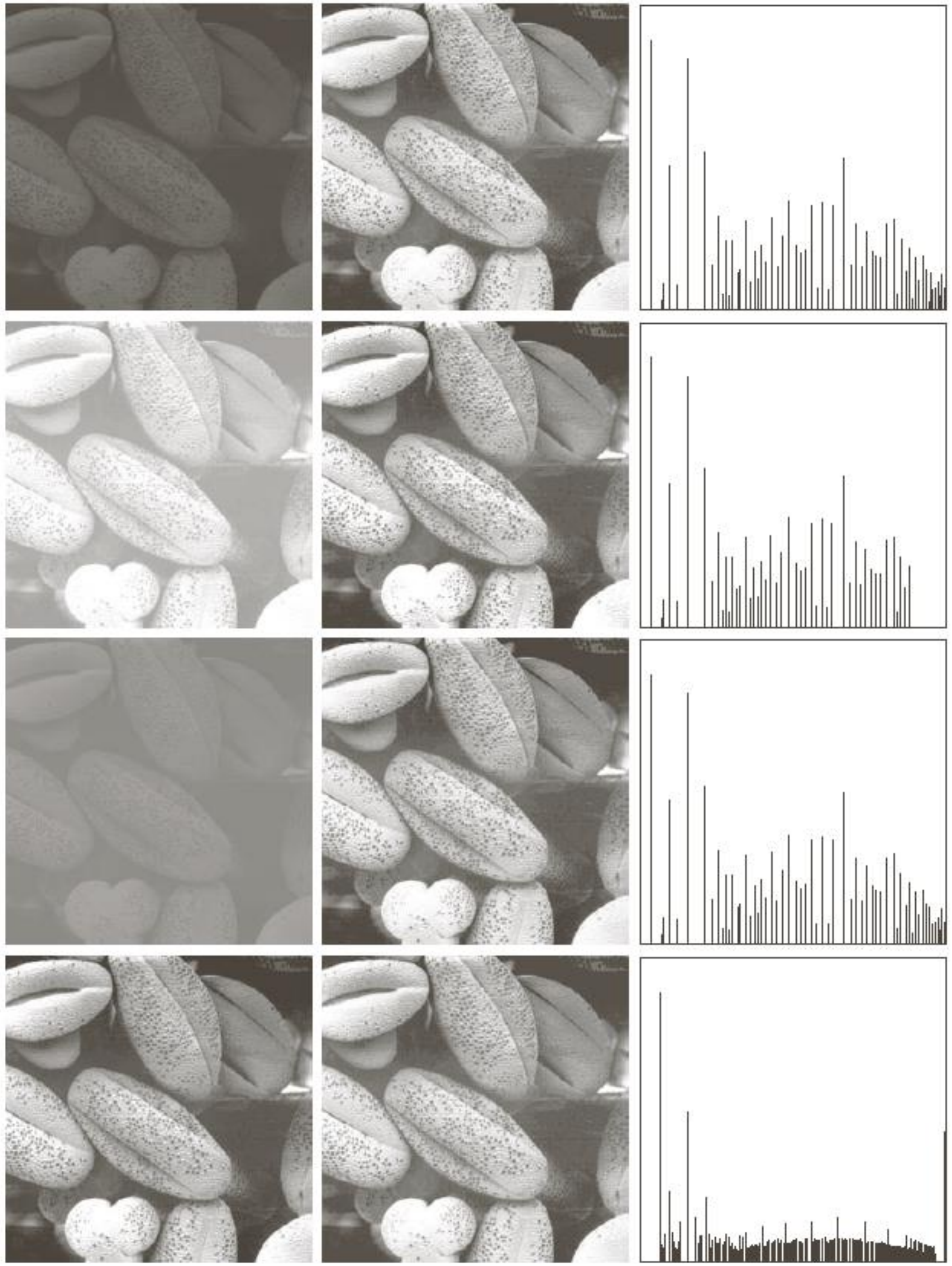


FIGURE 3.20 Left column: images from Fig. 3.16. Center column: corresponding histogram-equalized images. Right column: histograms of the images in the center column.

Example:

Combining Spatial Enhancement Methods

Goal:

Enhance the image by sharpening it and by bringing out more of the skeletal detail

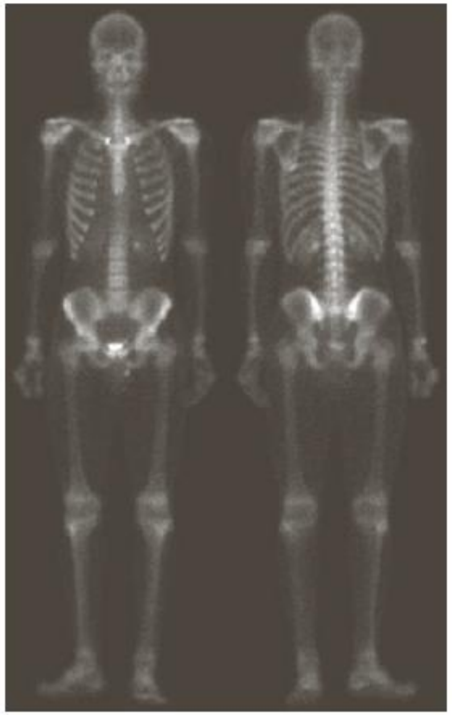


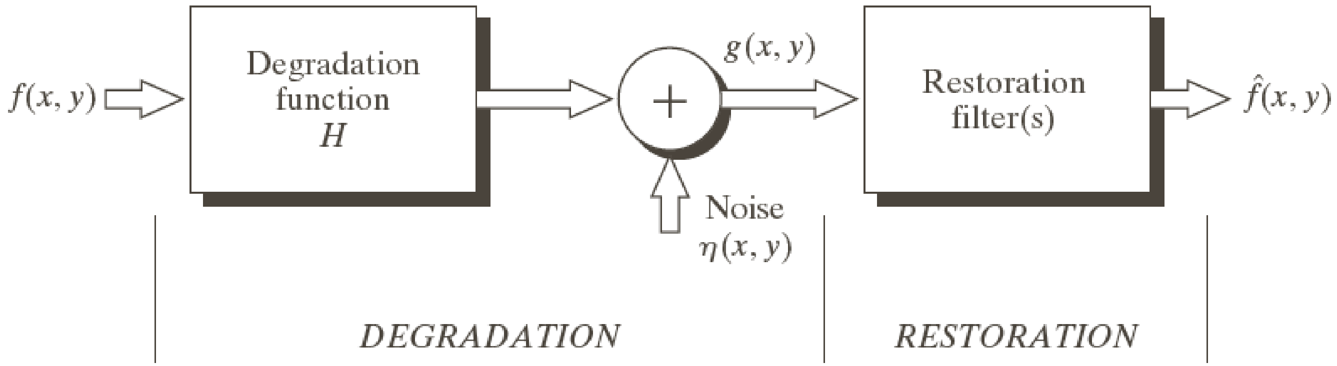
FIGURE 3.43 (a) Image of whole body bone scan. (b) Laplacian of (a). (c) Sharpened image obtained by adding (a) and (b). (d) Sobel gradient of (a).

Image Restoration

- In contrast to image enhancement, in image restoration, the degradation is modelled, this enables the effects of the degradation to be (largely) removed.
- Image degradations
  - Motion blur
  - Defocus blur
  - Resolution
  - Noise
  - Haze
  - Poor lighting

Linear, Position-Invariant Degradations

FIGURE 5.1
A model of the image degradation/restoration process.



$$g(x, y) = H\left[f(x, y) \right] + \eta(x, y)$$

Linear, Position-Invariant Degradations

In the presence of additive noise,

if H is a linear operator and position invariant,

$$g(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha,\beta)h(x-\alpha,y-\beta)\,d\alpha\,d\beta+\eta(x,y)$$

$$=h(x,y)\bigstar f(x,y)+\eta(x,y)$$

$$G(u,v)=H(u,v)F(u,v)+N(u,v)$$

Conventional Solutions

- The Inverse Filter
- The Wiener filter
- MAP formulation
- Wavelets and other transform domain approach.

Inverse Filtering

An estimate of the transform of the original image

$$F(u,v) = \frac{G(u,v)}{H(u,v)}$$

$$F(u,v) = \frac{F(u,v)H(u,v) + N(u,v)}{H(u,v)}$$

$$= F(u,v) + \frac{N(u,v)}{H(u,v)}$$

Minimum Mean Square Error (Wiener) Filtering

- **N. Wiener (1942)**
- **Objective**

  Find an estimate of the uncorrupted image such that the mean square error between them is minimized

$$e^2 = E\left\{(f - f)^2\right\}$$

Minimum Mean Square Error (Wiener) Filtering

$$F(u,v) = \left[\frac{1}{H(u,v)} \frac{|H(u,v)|^2}{|H(u,v)|^2 + K} \right] G(u,v)$$

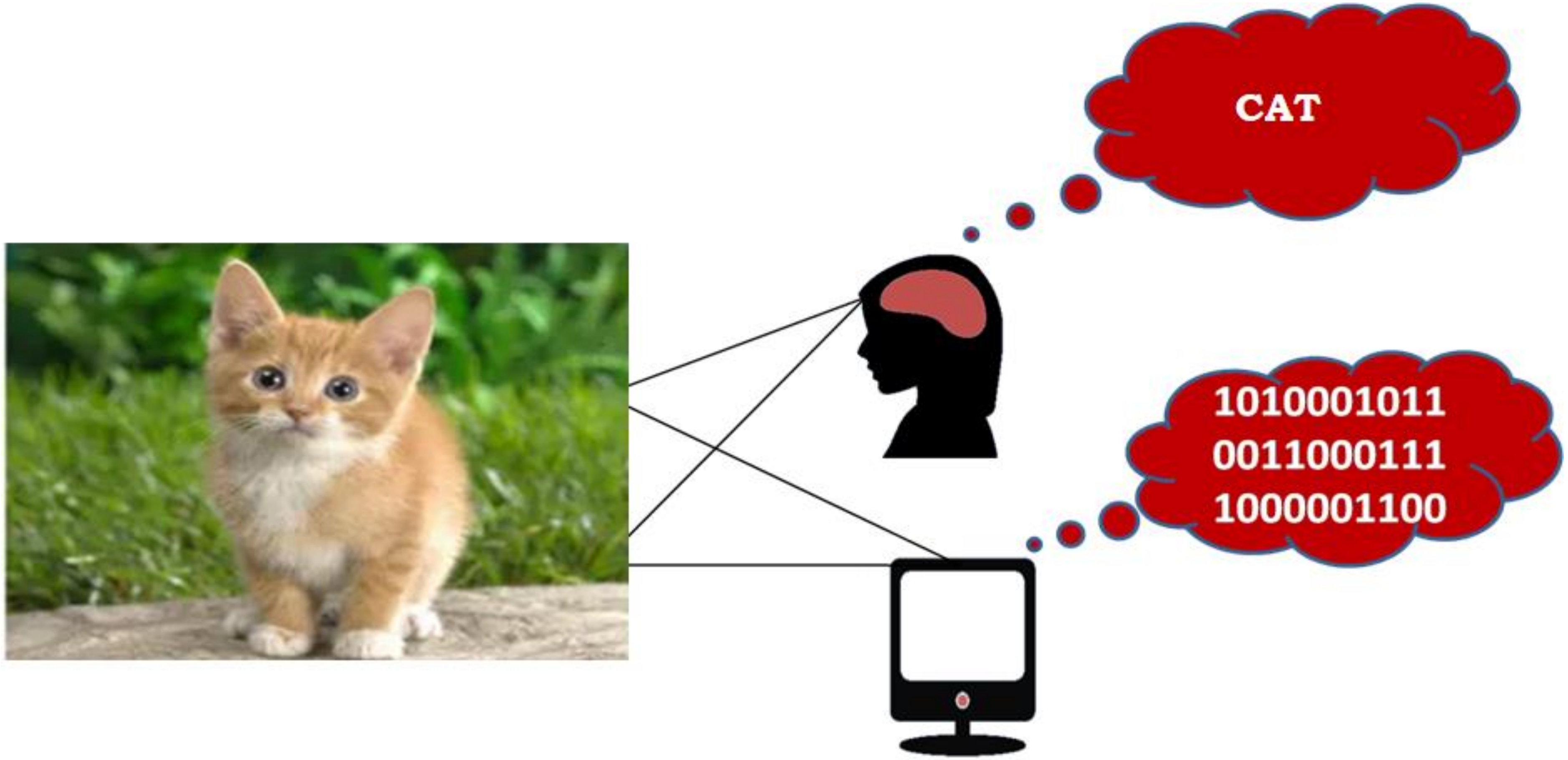
K is a specified constant. Generally, the value of K is chosen interactively to yield the best visual results.

Maximum a posteriori (MAP) estimation

- For an image with n pixels, write this process as ˆg = A f + n where ˆg and f are n-vectors, and A is an n × n matrix.
- Estimate f(x,y) by optimizing a cost function:
- $F = \arg\min\,(g - Af)^2 + \lambda p\,(f)$
  - Likelihood function and prior

Computer Vision

Bridge gap between human and machine understanding

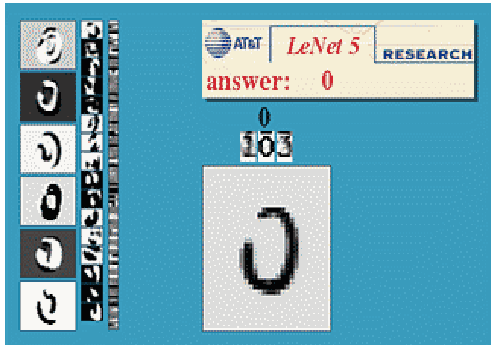


Computer Vision-Major Milestones

1959 | 1963 | 1974 | 1982 | 1991 | 2001 | 2012 | 2014

Image Scanning

AI emerges to solve computer vision



Optical Character Recognition

Detect edges Corners

Face Recognition: Eigen Faces

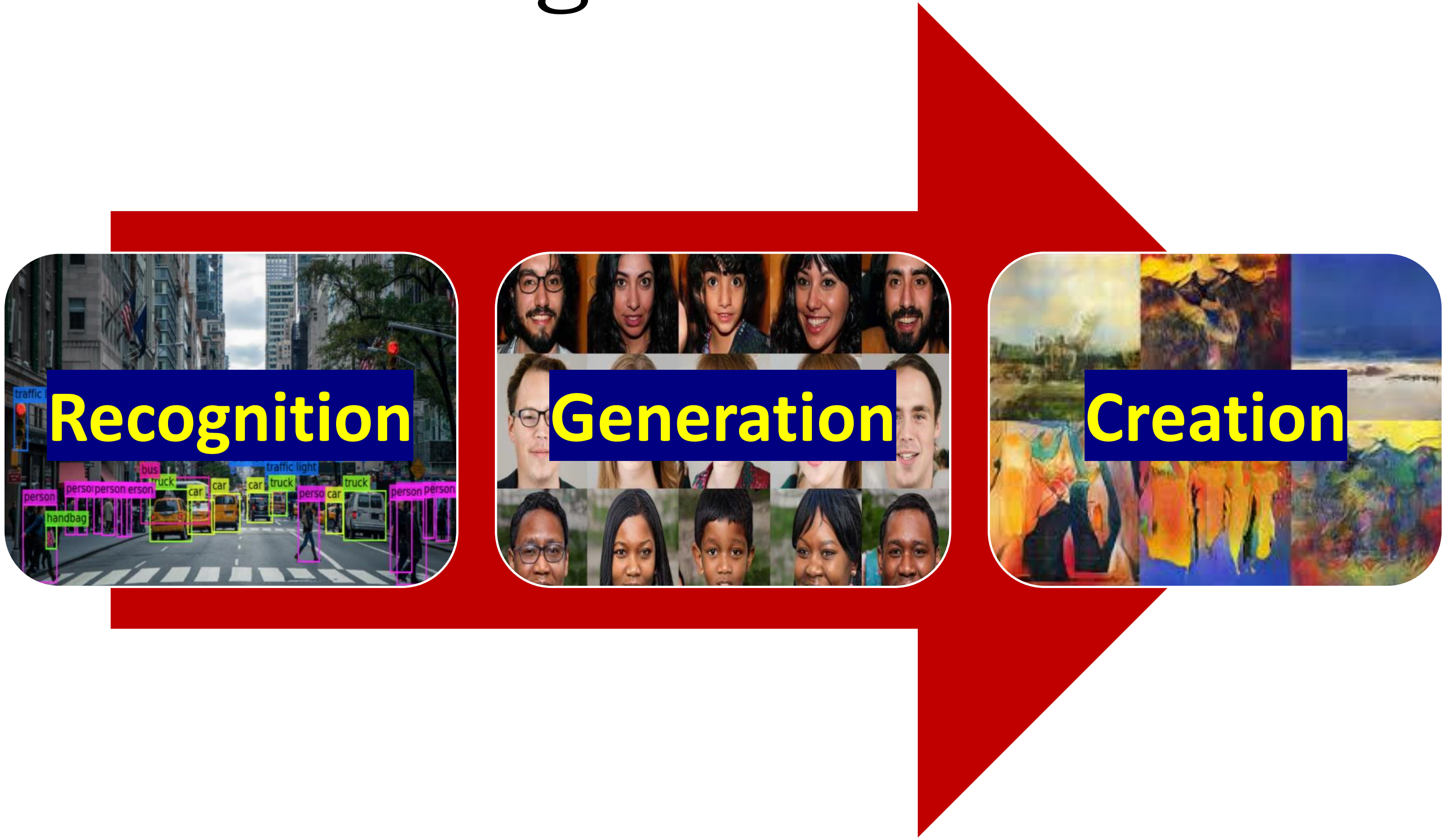
Viola Jones Face Detection

AlexNet for Image Classification

GAN for Image Generation

Scanning an image into a computer started in 1959 and in 2014 computers started created images by themselves

A Move towards Higher Form of Intelligence



A human can recognise things, generate new ideas based on existing and create total new ideas. Computer vision researchers are making computers do the same

Recognition: Image Classification

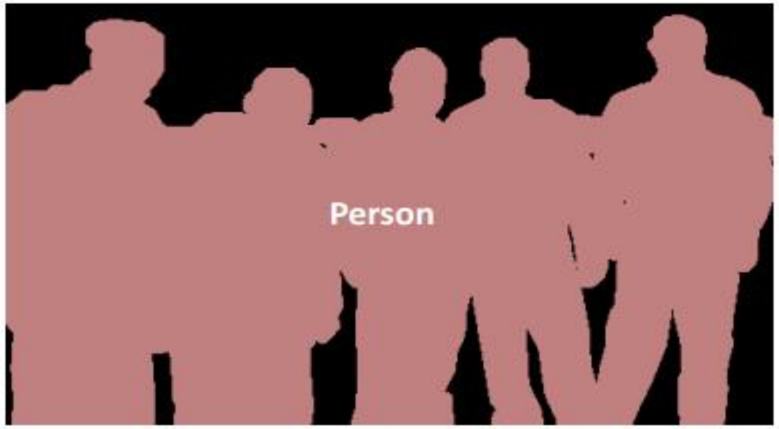
CIFAR 10
10 object classes
50000 training images
10,000 test images
Accuracy: 99.5%

IMAGENET
1000 object classes
1,281,167 training images
100,000 test images
Top 1 Accuracy: 90.88%

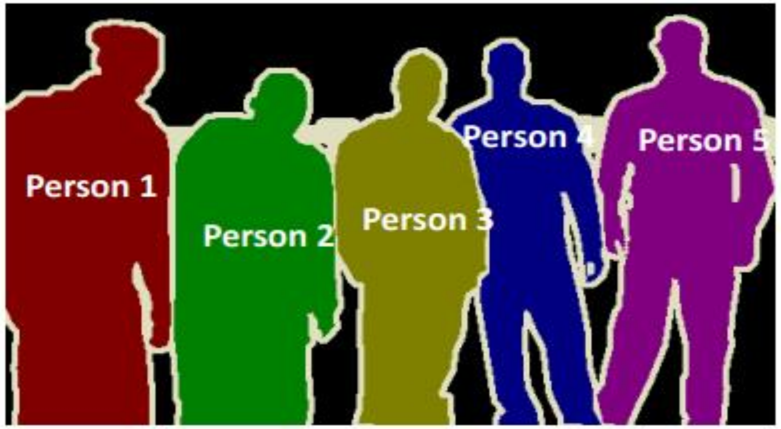
Image classification on standard datasets have achieved good Accuracy

Recognition: Object Detection

Object Detection



Semantic Segmentation

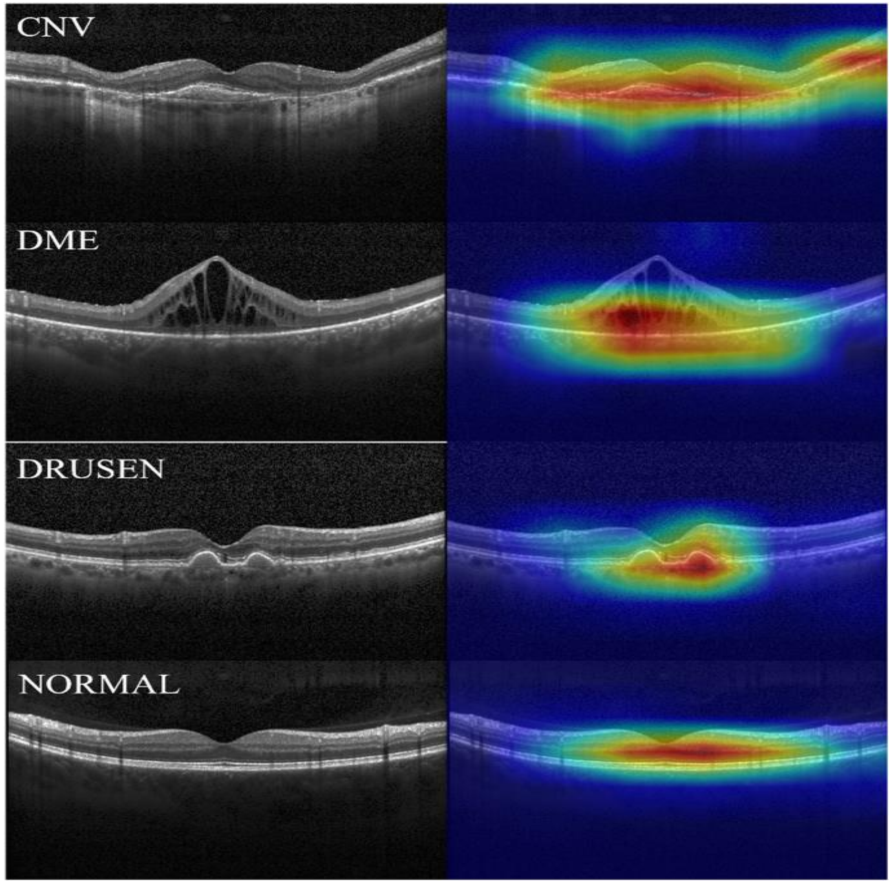


Instance Segmentation

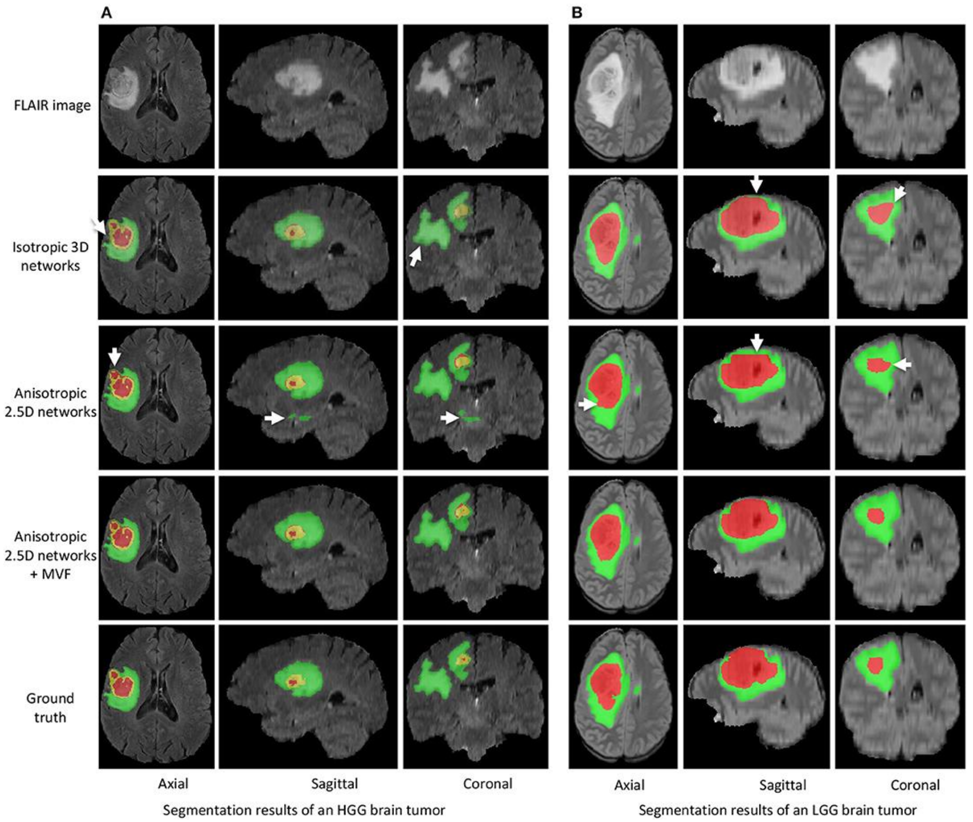
Task varies from simple object detection to semantic segmentation to instance segmentation

- Object Detection: draws bounding boxes around objects
- Semantic segmentation: pixelwise segmentation of object
- instance segmentation: segments multiple objects of the same class as distinct individual instances

Recognition: Medical Image Classification and Segmentation

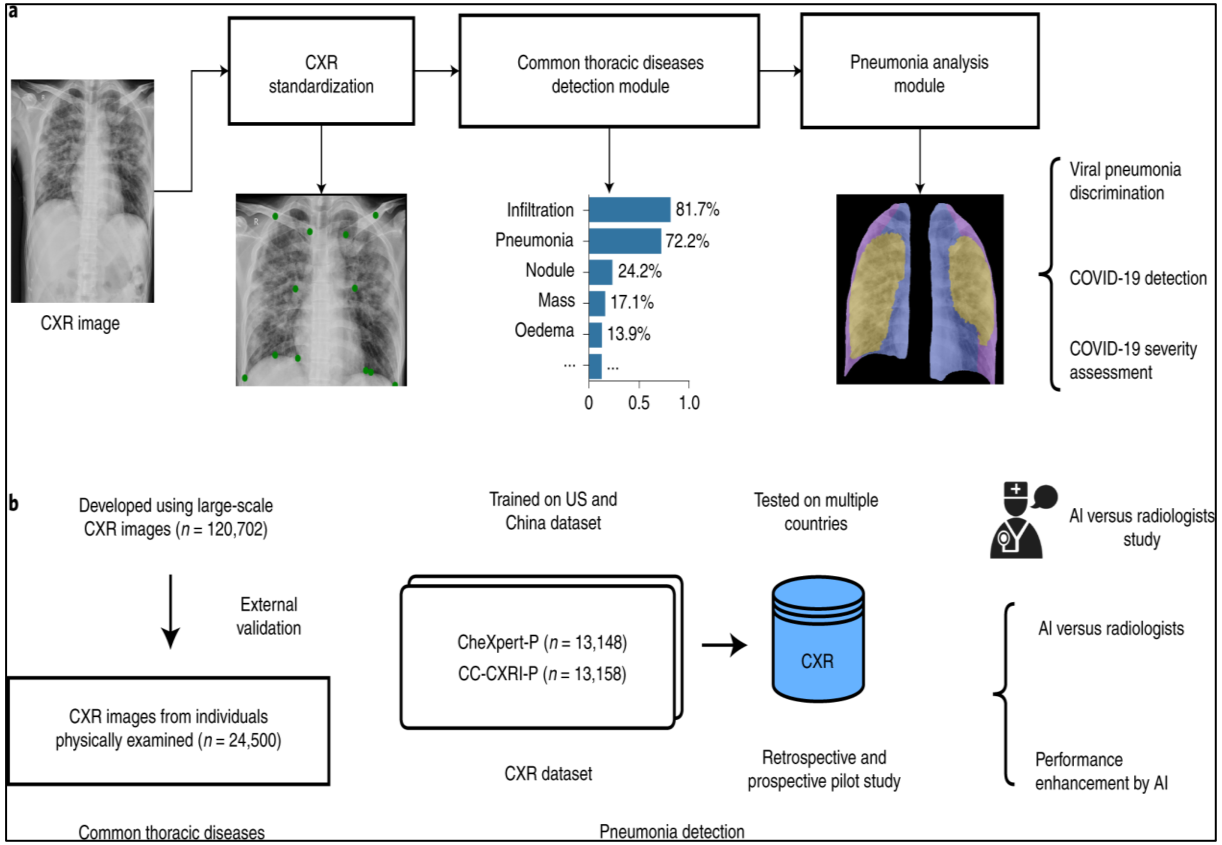


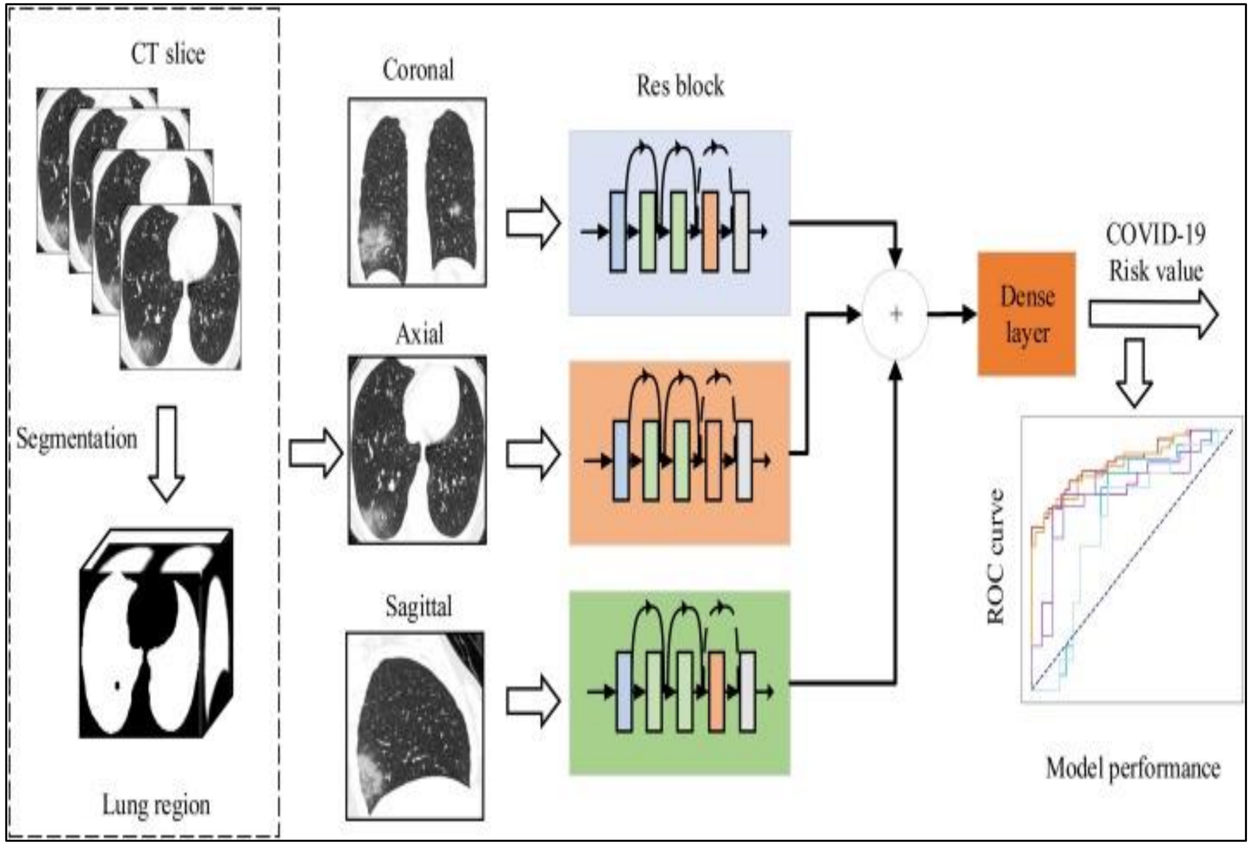
OCT Image Classification



Tumor Segmentation

Recognition: Medical Image Classification and Segmentation





Chest XRay

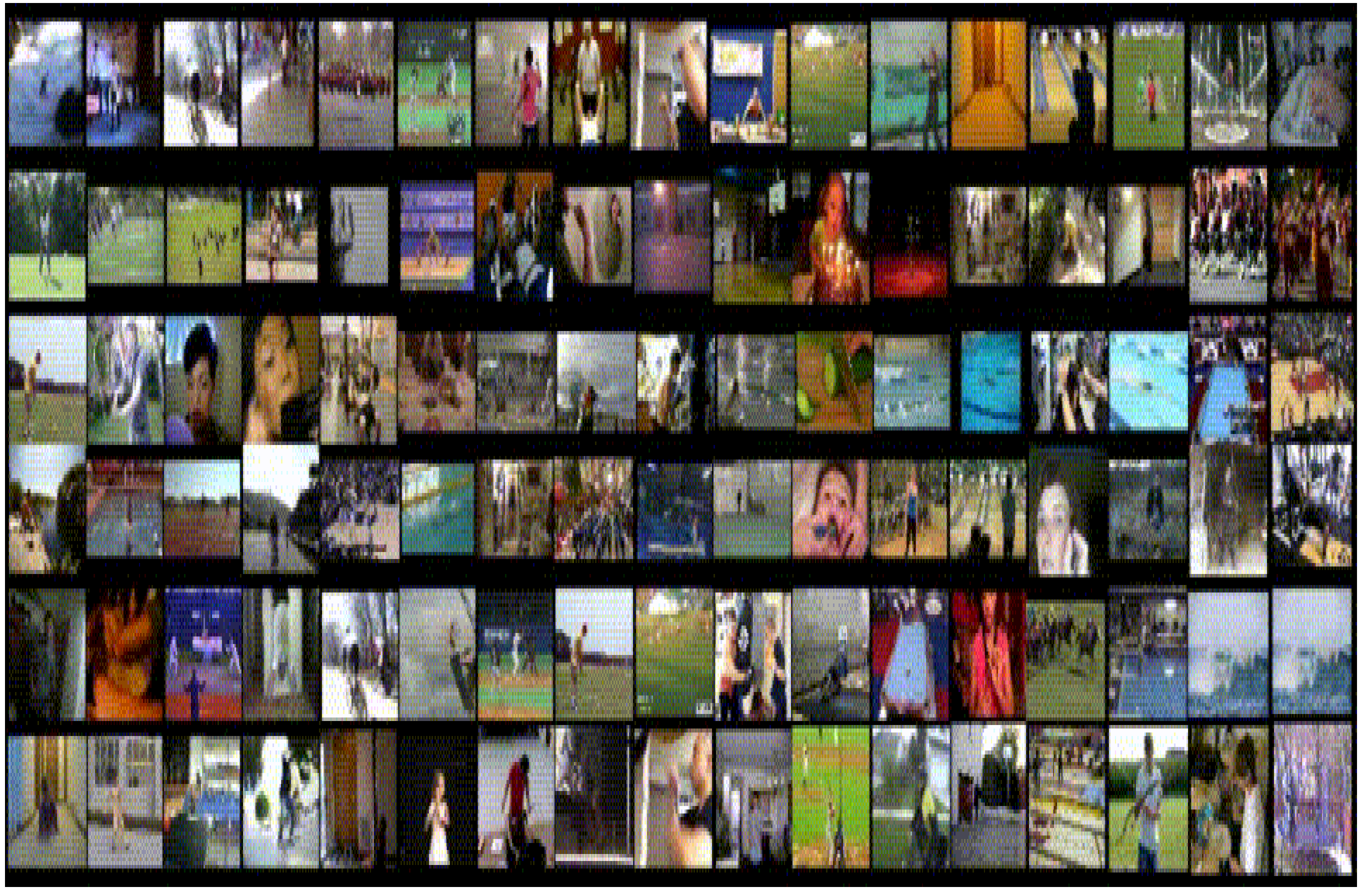
CT Scan

A number of works on Covid Detection were developed in short span of time

Recognition: Video Understanding

Involves labeling events in a given video

Recognition: Large scale activity recognition



Involves labeling large number of classes of actions using a machine learning model

Recognition: Video Understanding

Automated surveillance

Crowd behavior analysis

Motion capture and animation

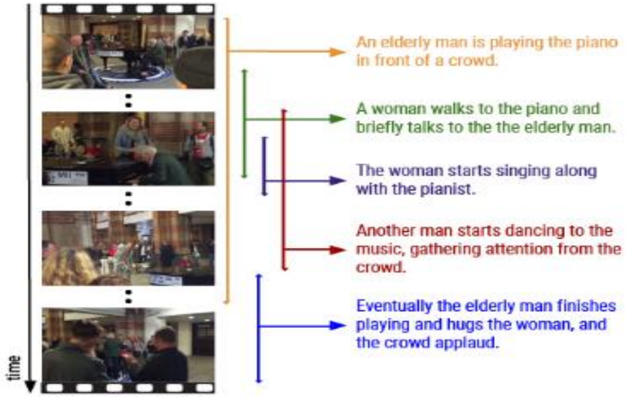
Unconstrained Video Search

Human computer interaction

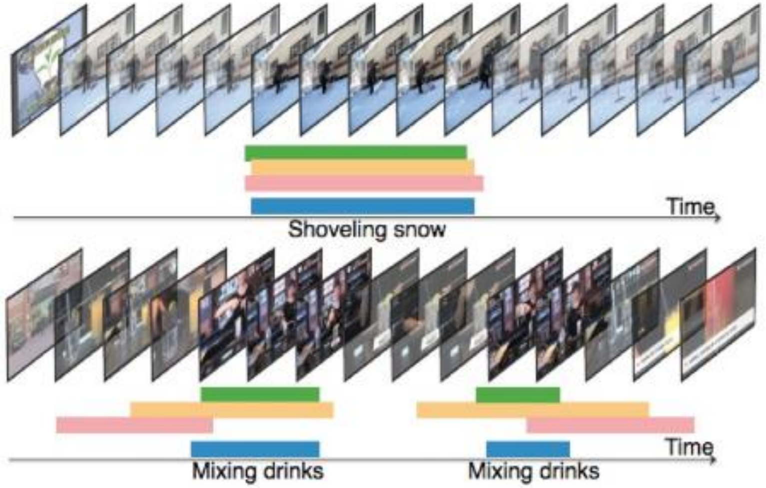
Wearable devices

Applications of video understanding

Recognition: Video Understanding



Left: Stand, Carry/Hold, Listen to; Middle: Stand, Carry/Hold, Talk to; Right: Sit, Write



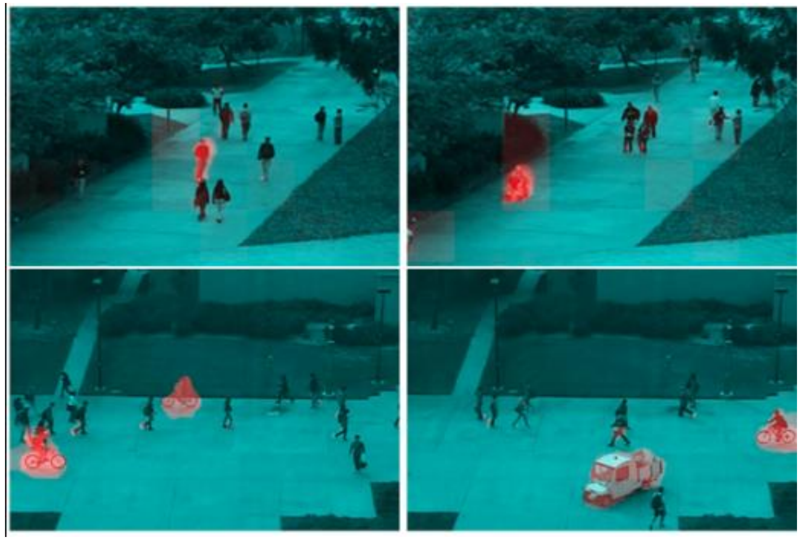
Dense Captioning

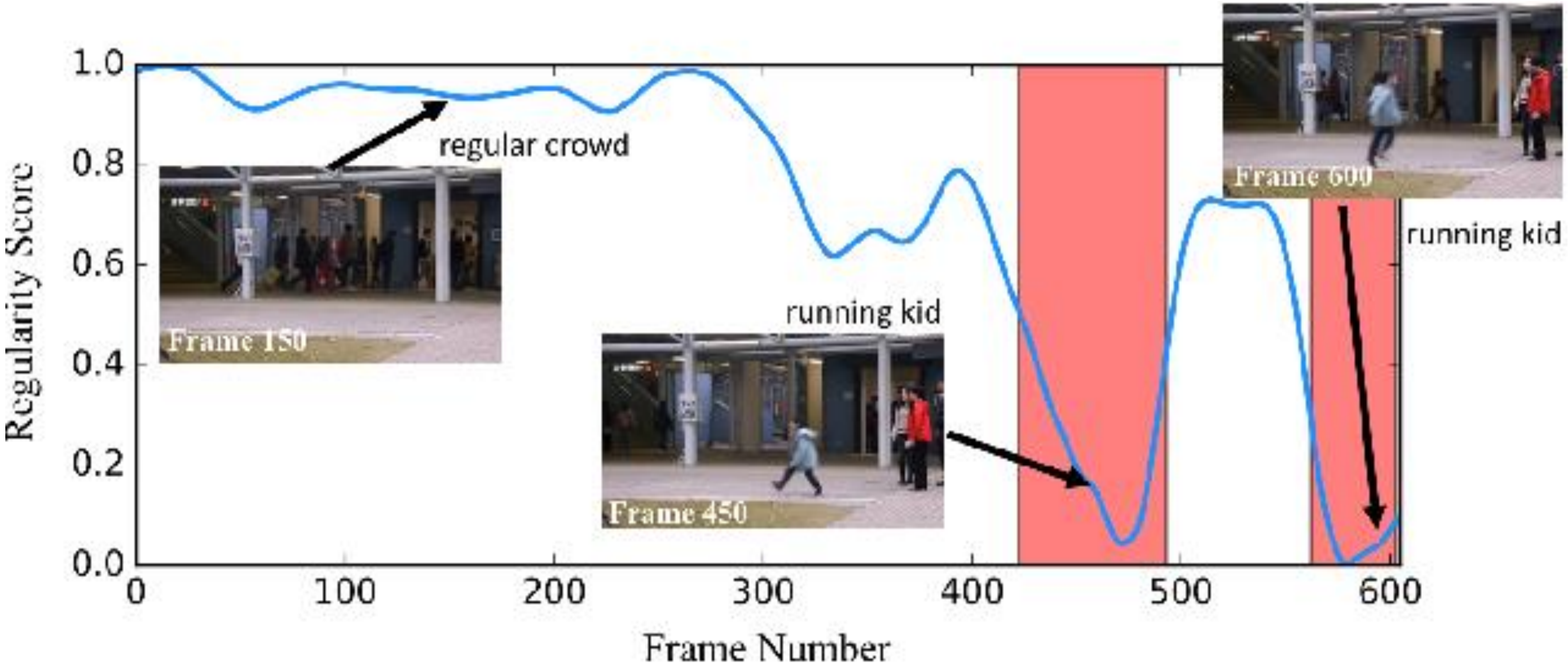
Spatio-temporal action localization

Temporal action localization

Other tasks involve generating caption for a video, temporal annotation of actions and spatio temporal localization of actions

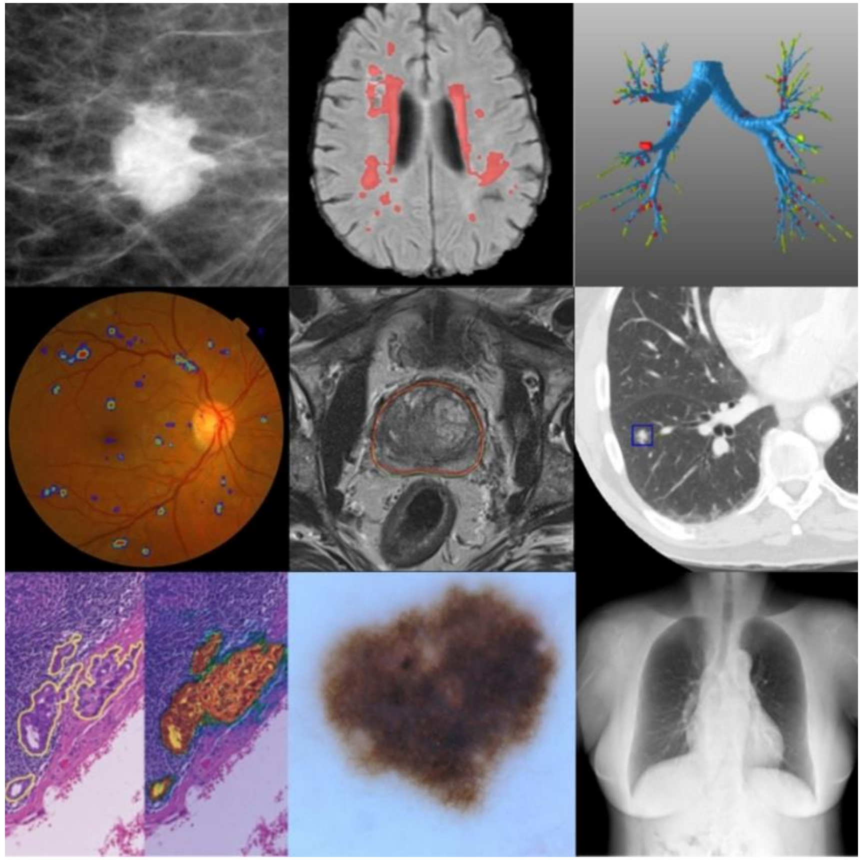
Recognition: Abnormal Event Detection



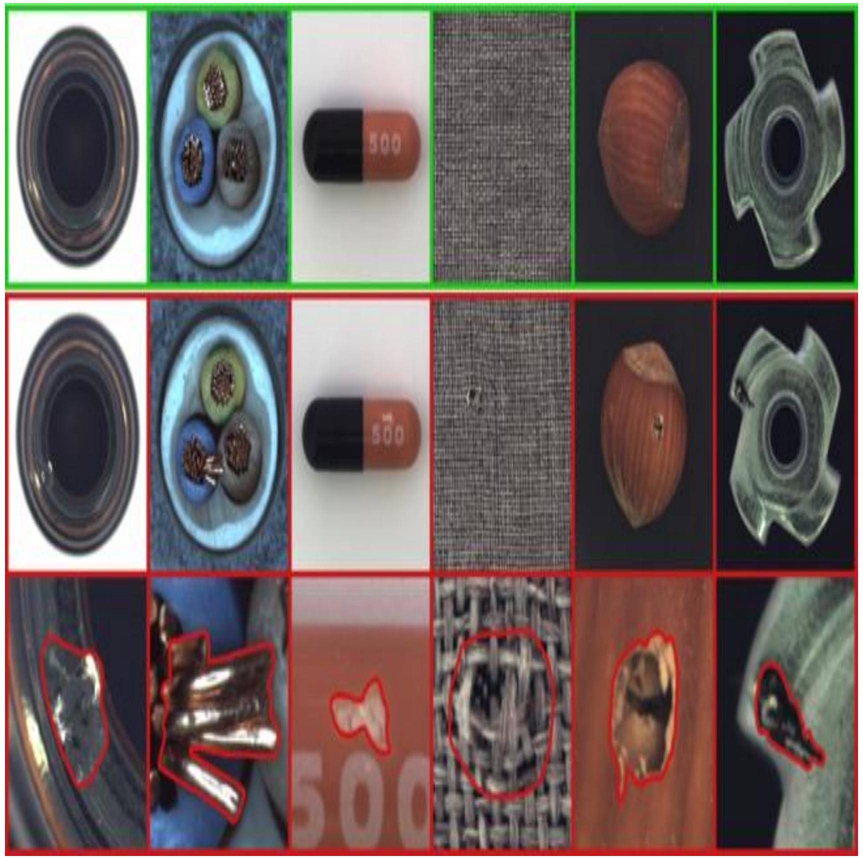


- Learn normal patterns
- Detect any thing deviating from normal as abnormal

Recognition: Medical Anomaly Detection



Medical Image Anomaly Detection



Industrial Image Anomaly Detection

Generation: Image Synthesis

- Use GANs for image generation
- DC GAN

Generation: Image Generation based on an attribute

- GANS generate new inputs based on attribute/pose given a input
- Style GAN

Generation: Other Applications

Super Resolution

Haze Removal

Image Artifacts

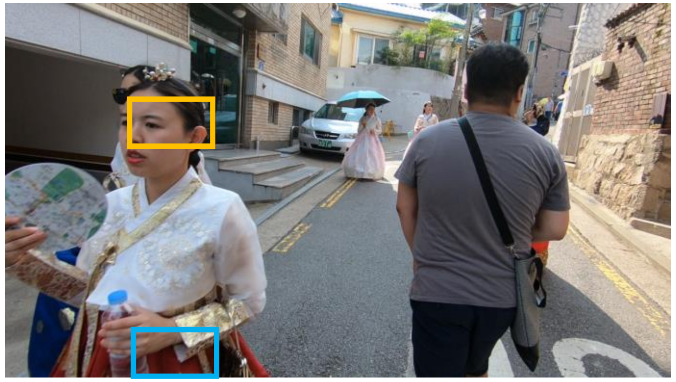

Clean


Low-Resolution

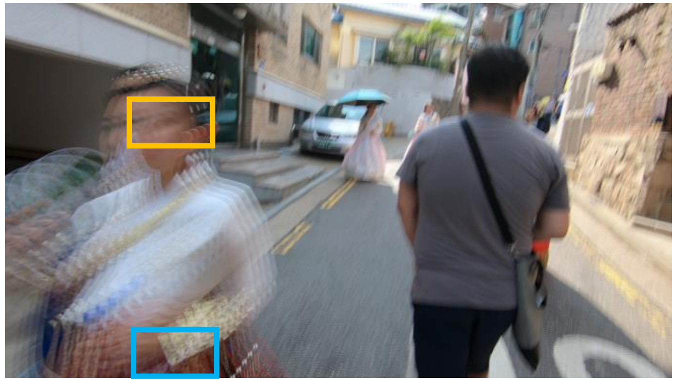

Motion Blur


Blur & Compression

Blur Examples

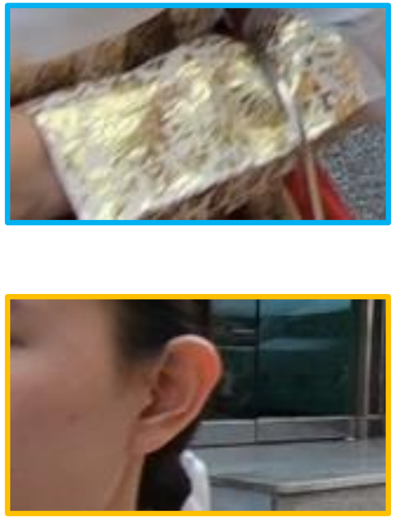


(a) Sharp

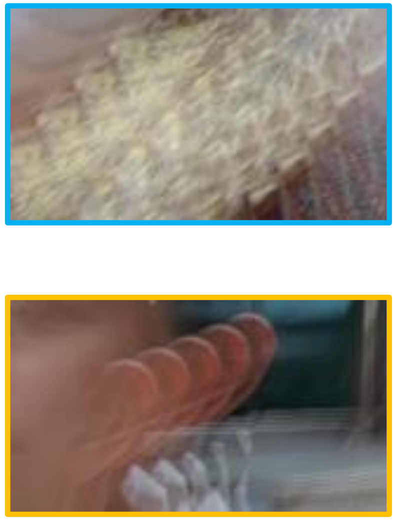


(b) Blur (120 fps)

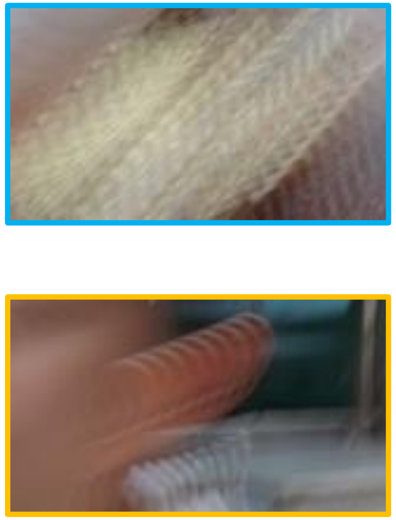
(c) Blur (1920 fps)



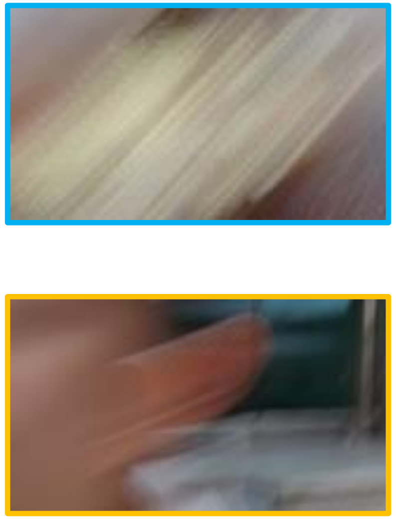
(d) Sharp



(e) Blur (120 fps)



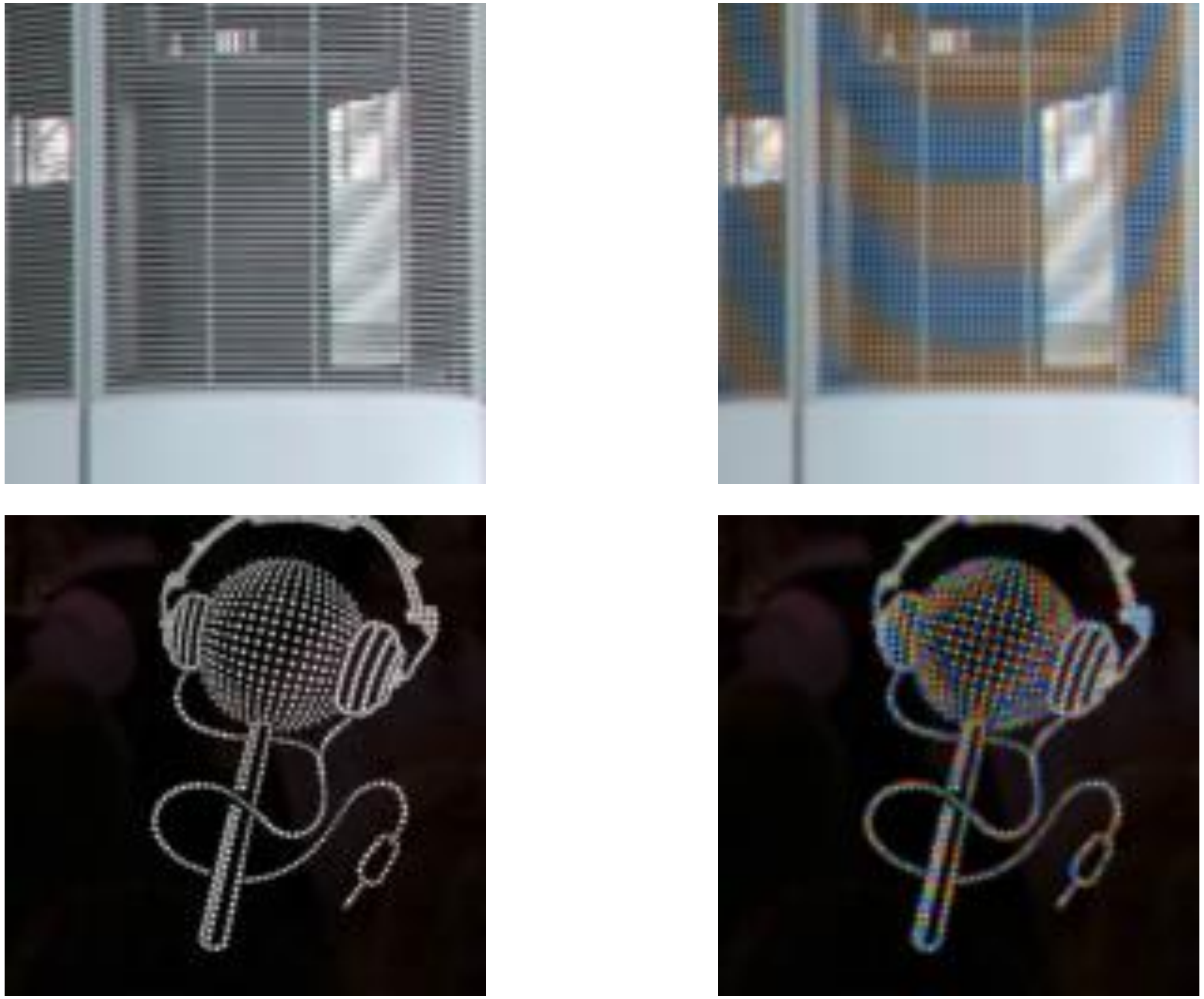
(f) Blur (240 fps)



(g) Blur (480 fps)

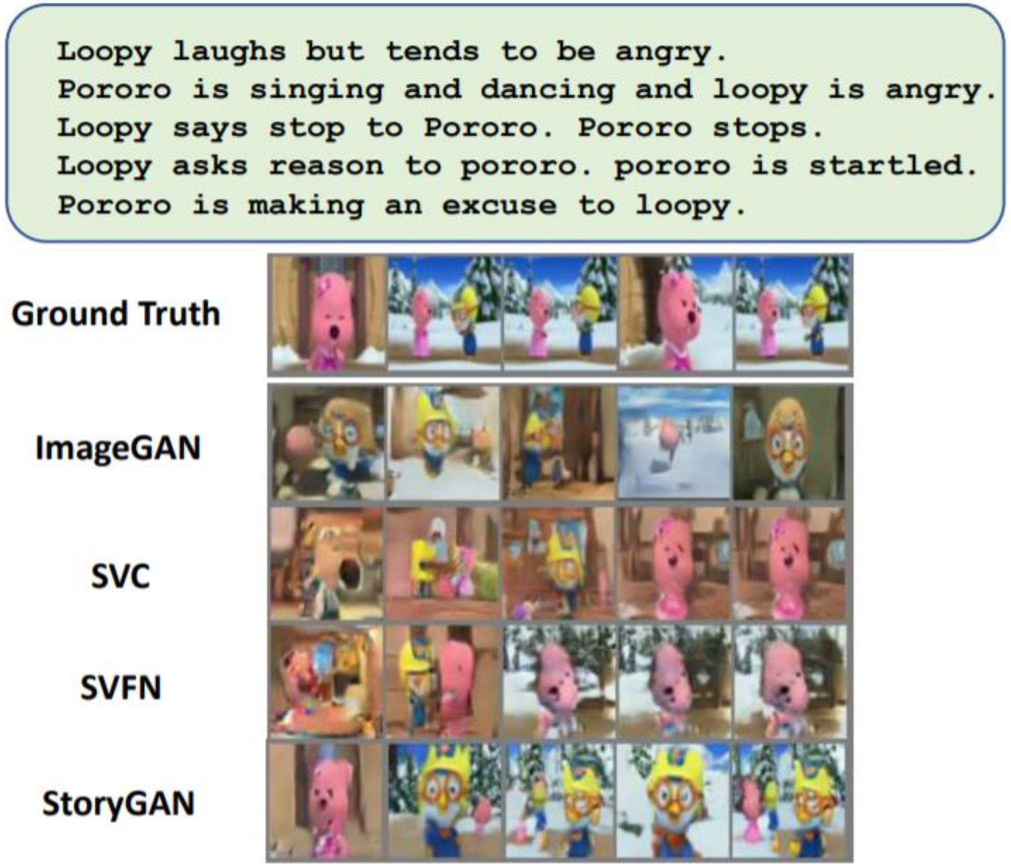
(h) Blur (1920 fps)

Moire Examples



Sharp Moire

Creativity



Art Work by GAN

Generate images from captions

GANs are also used for creative applications like generating images based on a caption and creating new artwork

Computer Vision: Other Research

- 3D Reconstruction
- Embedded Vision
- Explainable AI in computer vision
- Single Shot and Zero Shot Recognition
- Fake Image/Video Detection

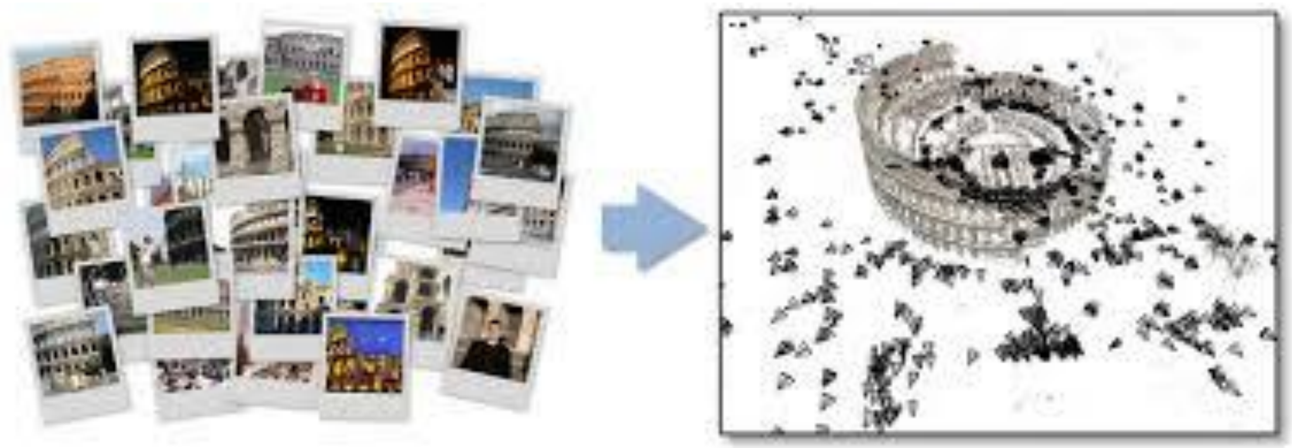
3D reconstruction

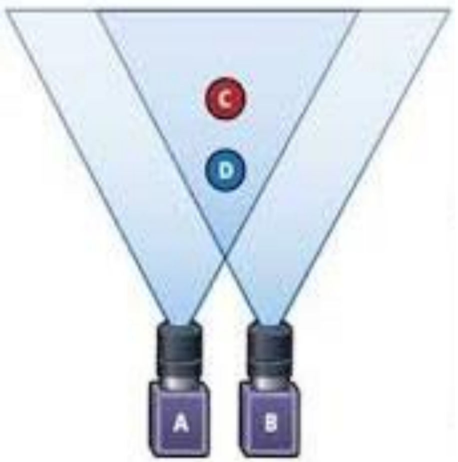
Recover the 3D properties of a geometric entity from its 2D images

3D reconstruction

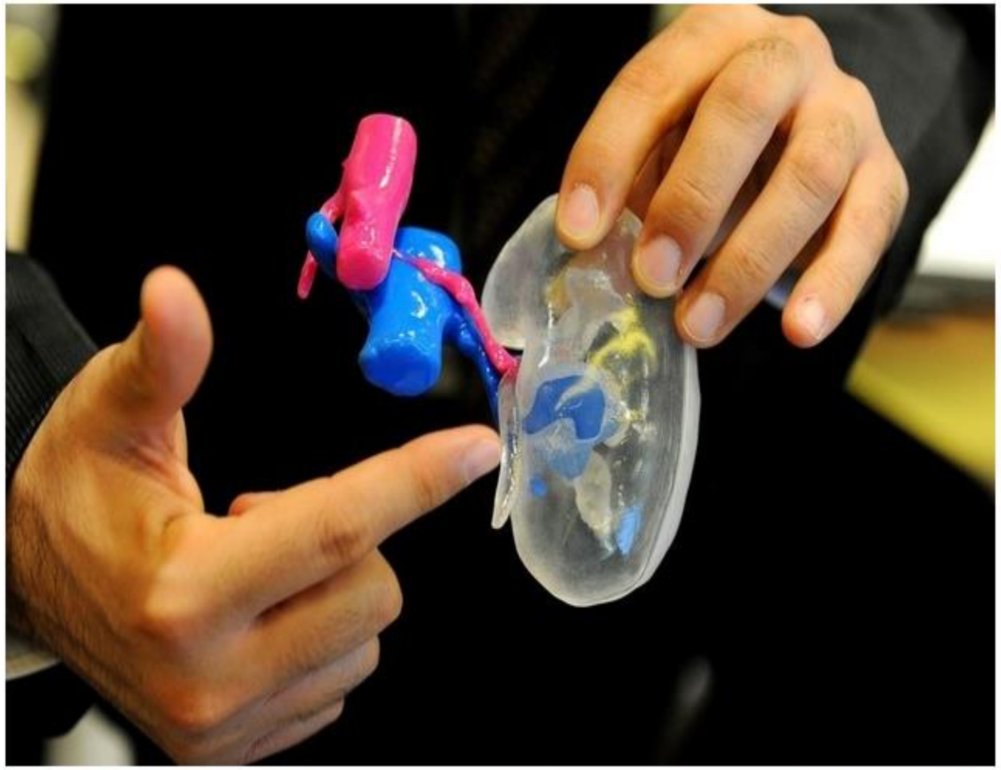
3D reconstruction can be done from

- Single image - monocular cue:shade, texture
- Two images - computational stereo
- Multiple images - Structure From Motion





3D reconstruction: Application in Medical Imaging

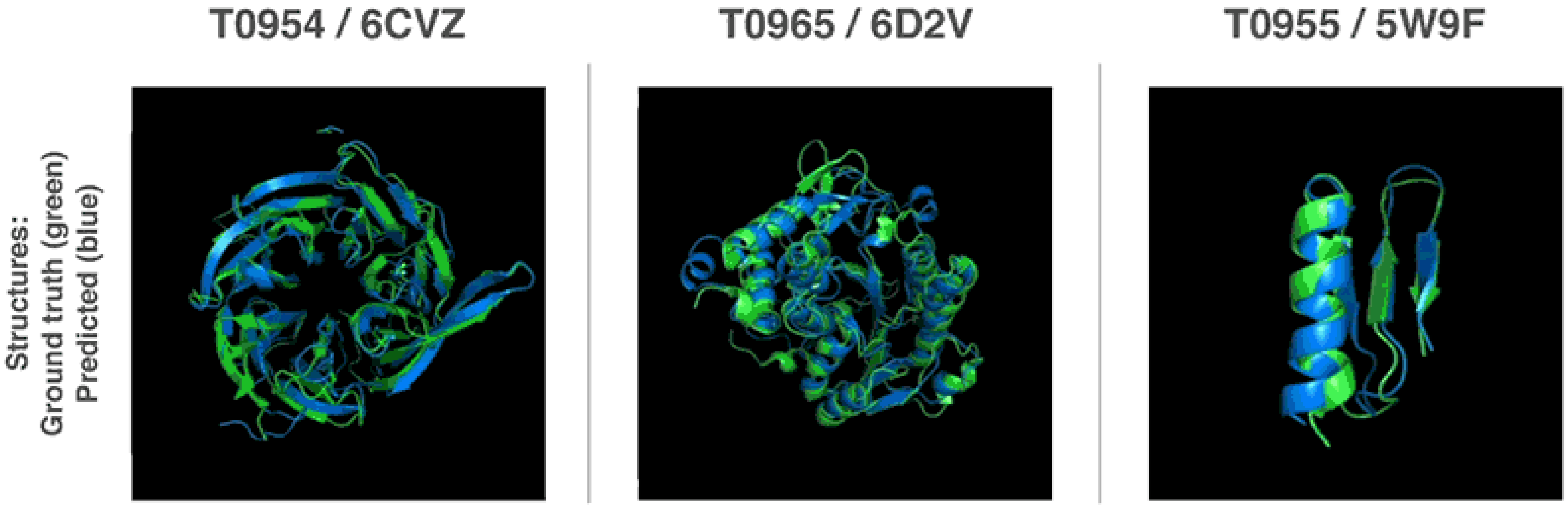


Doctors at Dubai Hospital performed a kidney surgery using 3-D technology, making the hospital the first in the MENA region to use this technology for kidney surgery. Courtesy Dubai Health Authority

Dubai doctors use 3-D modelling for kidney operation

Examples of using 3D reconstruction for medical applications

3D reconstruction: Modelling Structure of Protein



What any given protein can do depends on its unique 3D structure.

The ability to predict a protein's shape is useful to scientists because it is fundamental to understanding its role within the body, as well as diagnosing and treating diseases believed to be caused by misfolded proteins, such as Alzheimer's, Parkinson's, Huntington's and cystic fibrosis.

3D reconstruction: Simultaneous Localisation and Mapping

Mapping : building a map of the environment which the robot is in.
Localization : navigating this environment using the map while keeping track of the robot's relative position and orientation

Edge Revolution: Embedded Vision

- CNNs are developed for clusters of CPUs and GPUs which consume about 80 and 200W respectively,

- Real processing need to be done in limited power environment: mobile platforms : 5-10W, wearable devices: 0.1-1W

Edge Revolution: Embedded Vision

- Embedded Vision: Integrates camera and processing board
- Small computers on the "edge": on devices ranging from wearables and mobile phones to embedded processors in car
- Achieved by
  - Using specialized hardware
  - Optimization of algorithm: Weights of CNN

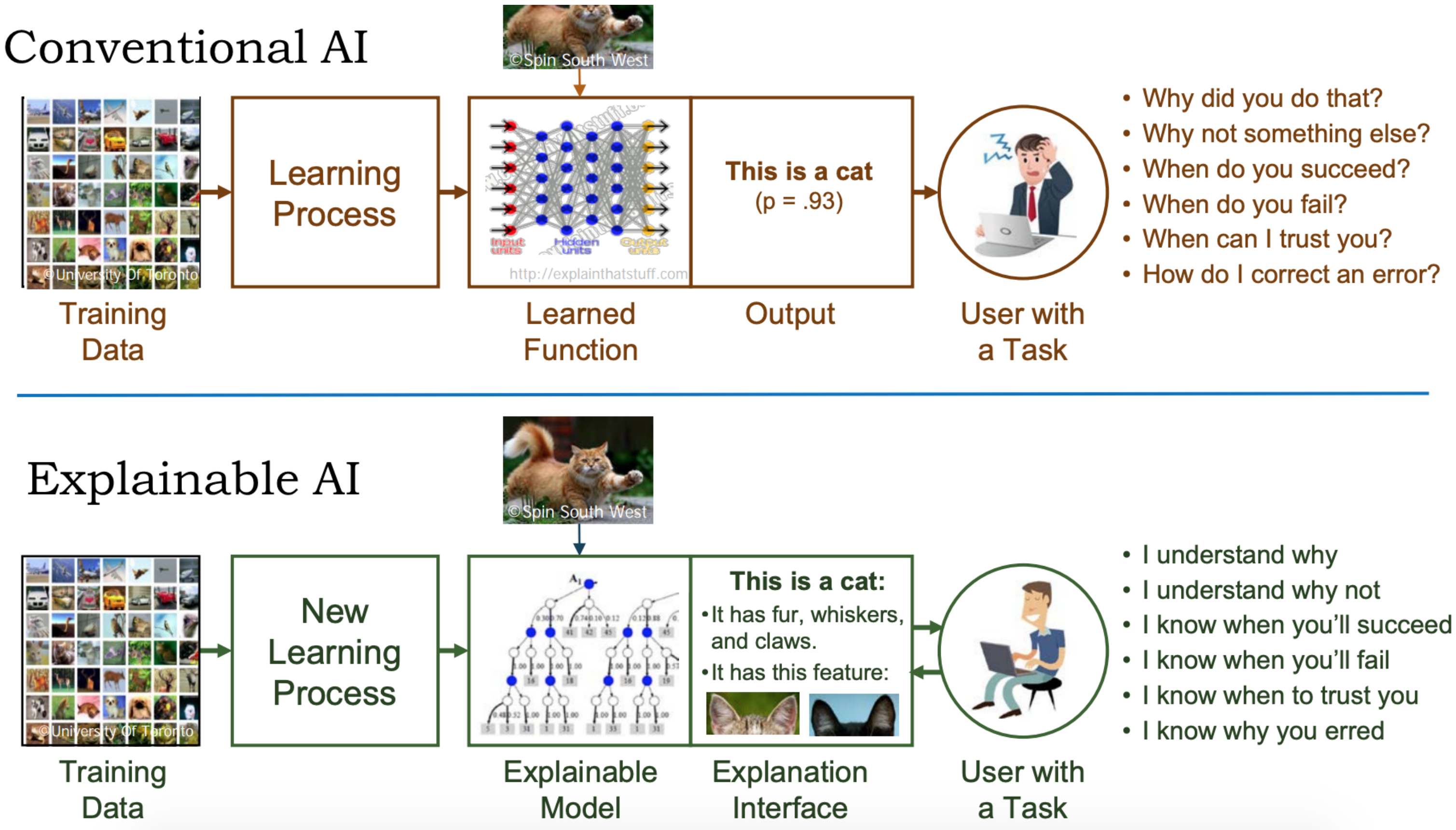
Edge Revolution: Embedded Vision

Optimizing algorithm

- Choose right CNN architecture
  - top performing with smallest number of parameters.
- Setting some weights to 0
  - Set less important weights to 0 with moderate effect on accuracy
- Quantizing weights to fewer bits
  - Reduce computation using smaller number of bits to represent weights
- Weights Deduplication
  - Eliminate duplicate or redundant information
- Computation increase can be offset by
  - Integer/Fixed point arithmetic instead of floats
  - Approximate/Imprecise computation
  - Use less precision / fewer bits
  - Do convolutions in frequency domain

Explainable AI

Novel AI-models explaining why image classification or segmentation was done so

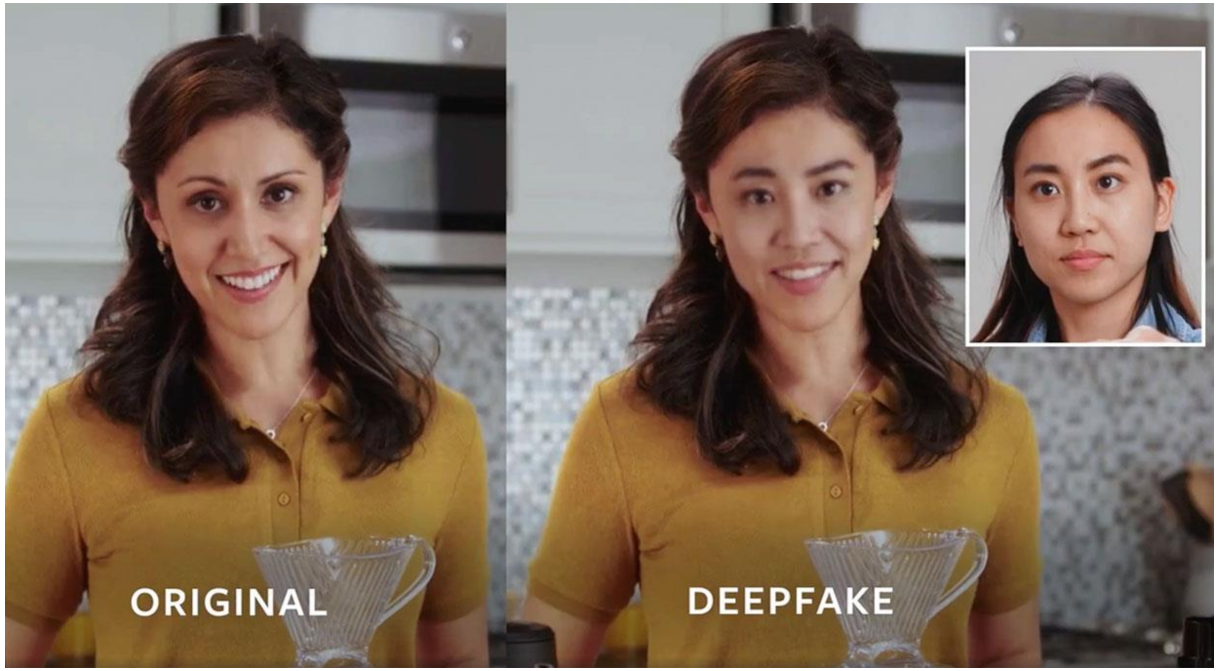


Single Shot and Zero Shot Recognition

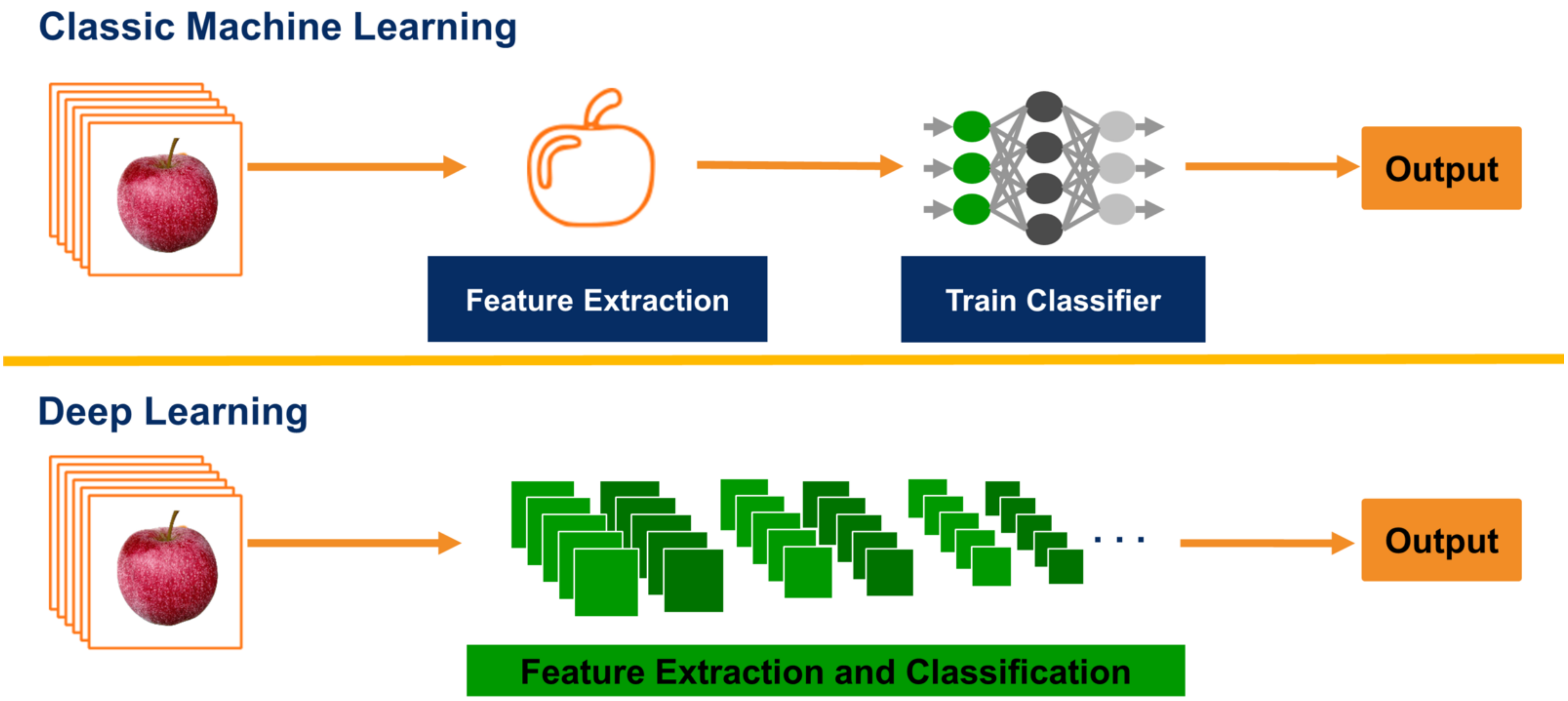
Training from minimum number of data for recognition is an evolving area of research

Fake Image/Video Detection

GANs can generate realistic fake images which can be misused. Identifying fake image and deep model which generated it forms another area of research



Classical Computer Vision



Classical Computer Vision

- feature based
  - Edges
  - Colour
  - HOG
  - SIFT
  - SURF

Recognition: Image Classification

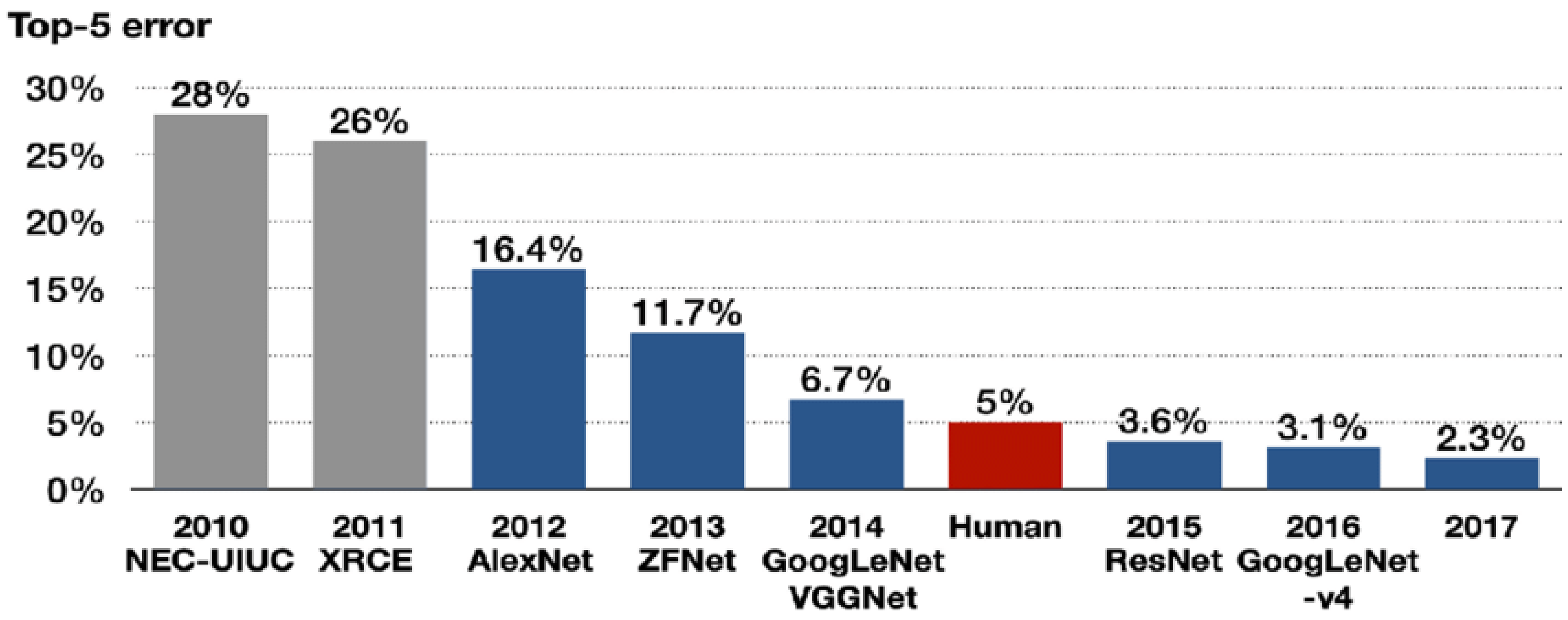
CIFAR 10
10 object classes
50000 training images
10,000 test images
Accuracy: 99.5%

IMAGENET
1000 object classes
1,281,167 training images
100,000 test images
Top 1 Accuracy: 90.88%

Image classification on standard datasets have achieved good Accuracy

DL for Computer Vision : from 2012 onwards

Results on ILSVRC dataset



What all we cover in this course

- Convolutional neural Networks
  - Alexnet
  - VGGnet
  - Resnet

- Generative Adversarial Networks (GAN)
- Transformers

Top Vision/ML conferences

Computer Vision

- IEEE International Conference on CVPR(Computer Vision and Pattern Recognition)
- ICCV (International Conference on Computer Vision)
- ECCV(European Conference on Computer Vision)
- WACV(Winter Conference on Applications of Computer Vision)

Top Vision/ML conferences

Machine Learning

- NeurIPS (Conference on Neural Information Processing Systems)
- ICML(International Conference on Machine Learning)
- ICLR(International Conference on Learning Representations)
- AAAI(Association for the Advancement of Artificial Intelligence)

Evaluation Policy

- 1. Quizzes (2): 10%
- 2. Homework Assignments (2): 10%
- 3. Programming Assignments (2): 20%
- 4. Term Project: 40%
- 5. Final Exam: **20%**

A Desired Course Outcome

- ICCV 2023 (https://iccv2023.thecvf.com/)
- ICCV Workshops and Challenges
- 56 CV/ML related workshops!
- Paper Submission Deadlines: August/September/October 2023
- Individual/Group projects encouraged.

Relevant papers

- [Imagenet classification with deep convolutional neural networks](#)
  A Krizhevsky, I Sutskever, GE Hinton

 Proceedings of the 25th International Conference on Neural

 Information Processing Systems (Lake Tahoe, NV, Dec. 2012), 1097 –
1105 (2012): 120054 citations

- [Very deep convolutional networks for large-scale image recognition](#)K Simonyan, A Zisserman
- arXiv preprint arXiv:1409.1556 (2014) : 91005 citations

Relevant papers

- Deep Residual Learning for Image RecognitionK He, X Zhang, S Ren, J Sun

Computer Vision and Pattern Recognition (CVPR), 2016 : 147081 citations

- Generative adversarial networksI Goodfellow, J Pouget-Abadie, M Mirza, B Xu, D Warde-Farley, S Ozair, ... (NIPS 2014) : 52737 citations
- Attention is all you needA Vaswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, AN Gomez, ...

Advances in neural information processing systems 30 (2017): 59227 citations

References

- Images taken from different sources from the internet