COMPUTER VISION

Cognitive Science	NLP
Machines which could think like humans	Machines act like humans

- When we look at world around, we
  1. Recognize things
  2. Analyse objects and actions
  3. Match moving points (not possible for a 3-year infant

- Computers were created to visualize
  1. Data just like humans do
  2. Digitize data

- Optical character recognition: written to typed text -> deal with data in images
- Geometry: object at some distance is recognized by us. Why not by machines?
- Image processing: photos smoothening using many techniques (helped in ML)
- Eg: Mosaic photography – capture photo of a room from different angles and directions, find common features and tie the images
- We look and reason images and try to connect them to some instant.
- There is no start and end date of CV.

History of CV

- 1960: digital cameras -> digital images(numbers)
- Blocks world model
- 1975: optical flow – track object motion
- 1985: Markov Random field - Data structures store data such that Pr(pixel)=Pr(neighbouring pixels)
- Graph cuts: image = points and edges -delete background in images
- 1990: Segmentation algorithm - break image into parts
- 1995: Face detection (not Deep Learning) -> FRT
- 2010: ML on various techniques

Standard tasks

I. Processing: get basic data from image and then
   a. Convert colour image to grey-scale
   b. Normalize pixel value
   c. Find edge of image (light has a sudden shift)
II. Feature extraction: illumination, colour, texture
III. Segmentation: unsupervised and Semantic (supervised = human annotated)
IV. Stitching: Mosaic
V. Recognition: what is the object in the image
VI. Detection: what is in the image + a rectangular bound around it

VII. Captioning: generate text output/description from an image. Its reverse is Stable diffusion.

Edge detection

- Find a point where intensity suffers a sudden shift
- It's just like the **Differentiation** in Maths
- An image is just a box of numbers
- Canny filter: pass over every pixel and give difference wrt surroundings. Longest edge = longest 0s or 1s = contours of the image

Open-source software for Image processing

I. Open CV: largest library for CV content, can run on various languages
II. Google colab: free GPU access
III. GIMP: free image editor, substitute for Adobe

Data resources

- MNIST: image (128*128 pixel) of hand-written digits
- Imagenet: 14 million images, each label had more than 1000 images
- Flicker: before Instagram. Scientists downloaded images of people without their consent. What about © violations? People were not aware that their data could be used to make money in future. Their data was being distributed like **free lunch.** The owner does not get anything.
- Microsoft SCOCO dataset
- VQA: It was a Q&A dataset with focus on multi-modal task (CV+NLP)

DEEP NN = multiple CNN: multiple fully connected layers with specific functions

- Convolution -> ReLU -> Pooling: Edge detection
- Convolution layer summarizes weight of a cell wrt surroundings.

Simple use case

- FRT for blind people to navigate. But an error might occur which will be very deadly.
- There are pretty visible gaps like the stop sign was not detected in the image.

Why automated cars?

- No work for us -> we want to be lazy. But companies say that they will increase the safety of the passenger.
- When we drive, world is altered and ML cannot predict future. Eg: It can 20-30 people in India in 1 day.
- Automated cars are costlier than normal cars. We have finite land on earth. Thus, cars for all is never possible.
- Car transport is unscientific and a wasteful use of resources. Buses and trains are cheaper.
- Going everyday by Uber is cheaper than buying a car. People buy it just to satisfy their ego.
- A car causes the maximum inconvenience to humans. But people make scenarios that they want a car for saving a dying person or a pregnant woman.

- Car is a luxury which harms the planet.

Complex use case in comics

- Models are made to predict the future/text in the last panel: accuracy = 60%
- But humans do so with 85-90% accuracy.
- CV is harder with unseen natural representations.
- Humans can make out meaning from a badly-made sketch but not AI.

Keras (simplification of Tensorflow) and Lasagna are very useful libraries to make large models. They make coding easy.