In 2013, image-net and google were competing to advance computer vision

VQA datasets

Visual Question Answering dataset

NLP stuff – Question answering is when you ask a machine a question and expect an answer

VQA is when you show the computer a photo and ask the computer about the photo.

VQA is the first dataset to focus on that specific multi modal task

What is Deep Learning?

➔ We all know about neural networks

We have nodes and edges [4-2-2-4] fully connected

We get an output

We see the error, and then we back propagate to change the weights

We run the model on another vector, change the weights again, and this continues till the output gives us smaller errors

You need at least one intermediate layer, for some mathematical functions

If you need the output to be discrete, then the final layer is usually softmax layer

What is the output layer when we do not need discrete output?

Deep Learning -> Multiple layers of intermediate layers, all of which are not fully connected. There is a particular form of layer, called Convolutional layer which is used on images. Earlier, feature extraction used to be manual, but now it is done completely by using these layers

Convolution Layer, ReLU Layer, and Pooling layer

Convulation Layer

➔ Show the diagram of the window passing over the image

The window has some weights, which get multiplied to the image and gives you the output

The layer is represented as a box, because the image has 3 colours

RELU

Diagram of _/ function

For Pooling, show a max pool layer

Object Recognition Demo

→ From around 2010, this became an actual goal for researchers
→ Discussion of object detection on an crowded Japanese street, the model picked out most people and objects; however, it significantly missed a stop sign and a person, which poses challenges to real world application
→ In around 2015, the whole Idea of Autonomous Cars became viral, with projections till 2020, however, the professor believes it's a bubble, because it doesn't matter what the accuracy is, the world will change. Another challenge is that traffic laws are not followed in majority of the world. Then the policy considerations and the fundamental question of "Why do you want Automated Cars?". -> The top reason being people want to be lazy. Most people in the world do not have the money to buy a car, let alone an automated car. There is finite space on planet earth, and it is not possible to get everyone a car just because the space considerations of the road.
→ ban cars from cities, there are cities in Europe that have banned cars from city centres
→ Statistically a plane is safer than trains, and trains safer than cars. Why don't we make trains all over the place?
→ One of the first few datasets were based out of downloading pictures from flicker, without thinking about the privacy concerns, and is unethical and legally dubious

The effort of taking photos is still effort, and the users are not getting paid for it. More than any other AI, CV lies on unpaid labours of thousands/millions of people.

Computer Vision is much harder at representations, which are not natural representations [e.g. comic books].

The face detection software of delhi police doesn't work 1/3 of the time

Keras, Theano, tensorflow, lasagne -> Deep Learning Frameworks

Demonstation of code from AMNIST

→ Failed because of package installation issues