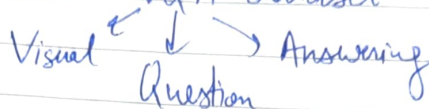


Today, we discuss multimodality and deep learning.

In 2013 - ImageNet & Google

Really popular dataset - VQA dataset



There was ~~no~~ lack of datasets involving both images and text.

VQA is asking questions from a visual content (images)

Deep Learning is basically when you more things than just fully connected NNs.

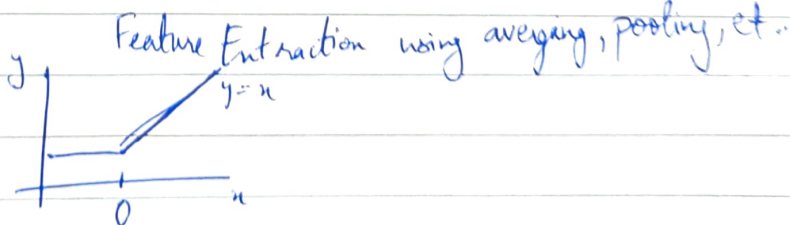
Softmax discretizes the output.

HW What is it called when not discrete?

Layers before NNs are feature extraction layers.
↳ Convolution is a major one

A layer which does Convolution is called Convolution layer.

ReLU layer →



Giving someone a tool with inaccuracy is not really good.
They might rely on it but it can't tell it very wrong.

Professor declares he's a web and shows an image of crowded Japanese street because AI might have failed even works on crowded Connaught Place street.

2027 - Computer Vision bubble will burst.

The world alters with time. Can't predict future.

Trains are the most efficient way of transport people.

Some European cities have banned cars

Plane > Train > Car = Safety

Why do we own cars? जीडी मारुति

Complex Use Case of Deep learning → Found Comic Panels where © have expired
OCR to find panel
LSTM on text + image
Predict next panel from earlier panels

Flicker scraped images → Copyright concerns

ML & AI depends on unpaid labour of all those who have uploaded & clicked pictures.

CV is not a solved problem → esp. in non-natural images

Keras, TensorFlow, Theano, Lasagna