MNIST is a dataset related to the images. Microsoft cocoo dataset which is also an image dataset. Flicker is also another image dataset.

Imagenet, by prof. liphephe which has 14,197,122 images and is the first largest image dataset that had been curated. It has 21841 labels and with 1000 images illustrating each label. It is quality dataset which is human annotated.

Many of the images had been given a description or caption or annotation.

VQA is a visual question answering dataset which requires both computer vision as well as the natural language processing cross feeding task

Deep learning – basically a big neural network with many layers

The basic module of a neural network is a perceptron

For training a neural network, both feed forward and back propogation is done

To get the output for discrete labels, softmax activation function is used with multiple fully connected layers preceeded by flatten layers of input signals.

Convolution layers are used to extract features from the input images by following the sliding window approach.

The feature learning layers include convolution+ ReLU and pooling layers. The neural networks which comprise convolution layers are called convolution neural networks.

AI doesn’t prevail in the domain of computer vision in cars as technology doesn’t matter in long run. In the public policy point of view, trains are much cheaper, efficient transport than cars.