In order to implement a linear classifier, we designed a feature extractor. The feature extractor is used to sample and normalise 8x8 patches every 4 pixels in the x and y directions. That means bag-of-visual-words features based on fixed size densely-sampled pixel patches are extracted. These bag-of-visual-words features constitute a feature matrix. The feature matrix by the algorithm K-Means is used to learn a vocabulary. Each vocabulary has a closest centre. When vocabularies of a patch are mapped to the clusters, machines automatically match one of the vocabulary clusters and recognize the image.

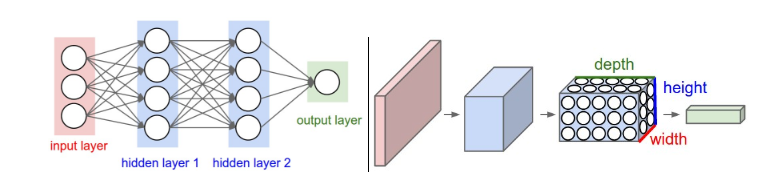
Note: Initially, the suggested number of $\sim500$ clusters was used, however after testing the classifier with different numbers of clusters, it was found that using $\sim600$ clusters maximized the accuracy.

A set of linear classifiers are trained by a subset of the images from training.zip data set. Internally, we represent the data set to a matrix containing exactly one reference to each image in the data set with multiple features.

After the algorithm K-Means, we get the classes. By comparing the classes with the label, we can calculate the accuracy and error of the classifier. When we test the accuracy of our classifier, we choose 90% of training.zip data set as training data and another 10% as testing data.

To achieve more higher accuracy, we compare the test result of 500, 600, 800 and 1000 clusters. It is appeared that more than 1000 clusters provided no valuable contribution to the accuracy.

Convolutional neural network is a class of deep neural networks, which most commonly applied to analyzing visual imagery. CNN are made up of neurons that have learnable weights and biases. A simple CNN is a sequence of layers that transforms volumes of activations to another through differentiable functions. We use three main types of layers that is Convolutional Layer, Pooling Layer, and Fully-Connected Layer to build CNN architectures.



In this coursework, we use CNN model named Alexnet to realize image recognition. Transfer learning as a tool by change specific parameter of layer of Alexnet realize the demand of this coursework. In AlexNet model, the convolutional layer has 5 layers, which function is abstraction and extraction of features. The function of pooling Layer is feature fusion and dimensionality reduction. AlexNet model has 3 fully connected layers. Transfer Learning is the reuse of a pre-trained model on a new problem. It is more popular in the field of deep learning because the knowledge of an already trained Machine Learning model is applied to a different but related problem. In transfer learning, we exploit what has been learned in one task to apply in another. For instance, in this coursework, the training.zip data set has 15 scenes. That means we will replace 1000 fully-connected layers with 15.

To match the CNN architecture, we need to transfer the grey scale image to colourful image. Meanwhile, the image size is adjusted to 227 by 227.

