الجامعة الإسلامية العالمية ماليزيا

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

يونيورسيتي إسلام انتارابغسا ملسيا

Project: Machine Vision

Image Classification

By

Mohamoud Hussein Mohamed

G1914041

Master of Science in

Mechatronics

Table of Contents

Abstract
Using machine vision for classification is one of the hot topics on the field of research, I this paper the famous methods used for image classification will be carried out and its results being compared. Along the test different aspects of the classification algorithm will be changed to visualize the effect that it has on the overall accuracy of the algorithm. A self-collected dataset of five different farm animals will be used in this image classification.

1.0 Introduction
Machine vision is one the latest technology that uses images as an input and analysis it to execute an action or provide results, we can consider machine vision as pair of eyes in a human being, as light enters our iris it produces an image in our retina which is processed in our brain. In machine vision the image input is always in numerical of each pixel in the image. The processor will be the computer that will process the image so that the output can be used for classification, detection etc.

Even though machine vision systems were available since 1950s, but recently the applications that uses machine vision has increased dramatically. Before the 21st century to possess a high resolution, camera was very unique but nowadays in every mobile is equipped with high resolution cameras. Some of the most common applications of machine vision are Final inspection of sub-assemblies, Engine part inspection, Label inspection on products, Checking medical devices for defects, Final inspection cells, Robot guidance, Verifying data matrix codes, Checking orientation of components, Traceability of manufactured products, Packaging Inspection, Checking laser marks and cuts, Medical vial inspection, Food pack checks, Reading bar codes.

In this paper compromise of 4 main sections, at section1 will include the introduction to machine, the second section will be describing the dataset that will be used to carry the classifications. In section3, the different algorithms that will be used will be briefly described, the section 4 will include the results of the classifications, finally the conclusion of the paper will be in the last section.

2.0 Dataset
Firstly, the dataset used in this project will be unique and its mad sure it to be the first time used for image classification. The dataset consists of 500 pictures for training the algorithm while another 200 images will used as test images to test the accuracy of the machine vision.

The image dataset that is used to carry out the classification will compromise of five different groups. As we are looking at the classifying farm animals, we picked the most common animal that can be found in a farm, the dataset is divide into five equal groups where each will consist of the same number of images, the groups are camels, cow, horse, goat, sheep.

The dataset was collected using the internet to download the images from different sources. This way was used to make sure that this dataset were not already compiled and used in image classification.

IMAGES EXTRA

3.0 CLASSIFICATION ALGORITHM
3.1 BAG-OF-WORDS
The bag of word is one the classical image classification, as the phrase says bag of words is using the image pixels in the image as the input and use a dictionary to classify the image, this technique is firstly introduced to be used for documents and search of words in a big file of written text and late it was used as an image classification method. Bag of visual word is a representation of unordered image patches is a bag, later the images patches will be classified using a visual dictionary. So, we can divide the bag of word in three main categories.

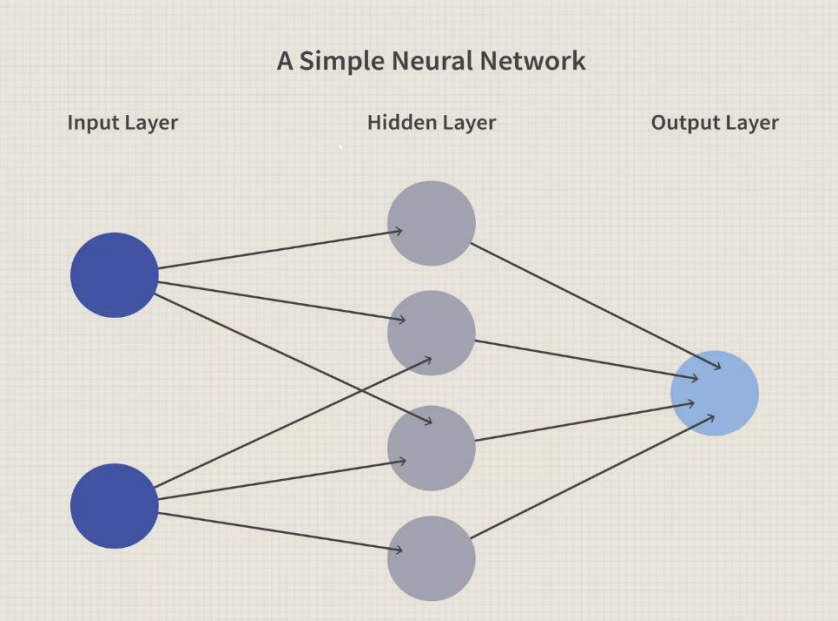
- Feature Extraction
- Codebook Construction
- Classification

In this project, for the local Feature extraction ORB (Oriented FAST and Rotated BRIEF) will be used rather than the most common SIFT as recently it was patented and cannot be used in the OpenCV library. For the visual word dictionary K-mean will be used and for the prediction and learning algorithm SVM (Support Vector Machine) will be used.

3.2 FULLY CONNECTED NEURAL NETWORK

Generally Neural networks are a series of algorithms that mimic the operations of a human brain to recognize relationships between vast amounts of data. They are used in a variety of applications. In other word its mostly used in deep learning.

Fully connected neural network can be said to be one of the easiest deep learning algorithms as its straightforward not very complicated as the modernized deep learning algorithm. It consists of a series of fully connected layers. As an example, a simple fully connected layer will consist of 3 layers and each layer will have several neurons. The 3 main layers are Input layer, Hidden Layer and Output Layer.



In this Project the layers of fully connected neural network that will be used to classify the dataset, will differ from test to test to compare the accuracy and the loss, the minimum number of layers used in the project is 3 layers fully connected neural network.

3.3 CONVULATION NEURAL NETWORK (CNN)

CNN being a neural network that being used in deep learning. Compare to the fully connected neural network CNN is more compilated and more used as it has higher accuracy. The general methods that CNN as deep learning algorithm uses the whole image pixels as it is instead of changing in into a single column array as its in the fully connected layer. Three main types of layers are used to build CNN architecture:

- Convolutional Layer:
  Convolution layers are layers that have filters, sets of kernels that are applied throughout the image.
- Pooling Layer: In pooling layers for example a kernel of 3 by 3 select the largest values on the feature maps and use these as inputs to subsequent layers.
- Fully-Connected Layer: this layer is always the last layer in CNN, after fully connected layer the image can be classified in the number of classes

In this project CNN architect used consist of 2 convolution layers,1 maxpooling layer and 3 fully connected layers.

3.4 TRANSFER LEARNING:

Transfer learning depends on pretrained networks, pretrained network are models that have been well trained on large scale dataset. Most of the pretrained networks are trained on ImageNet dataset. The usage of transfer learning image classification is one of latest techniques used as it was introduced only 6 years before. There are several pretrained model nowadays, they all differ in their neural network architecture.

To use Transfer learning to classify a person's specific dataset is by getting the pretrained network (in this project we use 2 type of pretrained network), and freezing the whole architecture and only modifying the last layer to by adding other layers or directly classifying the data in the groups that you require. Using transfer learning is faster than building a deep learning model from scratch and as it pretrained on large dataset the accuracy will be high.

3.5 FINE TUNING

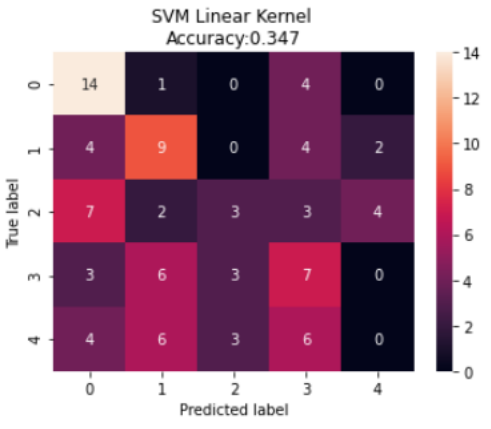
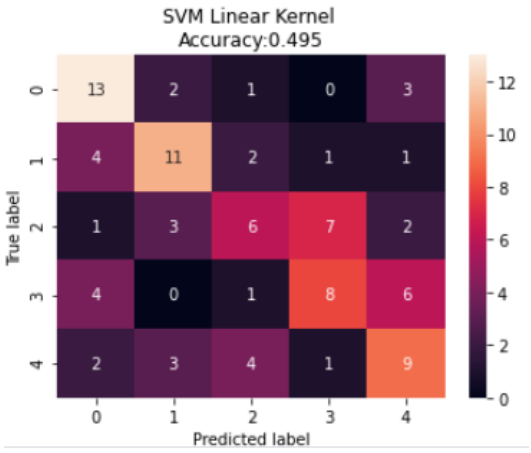
Fine tuning can be described as modified transfer learning, Its very similar as transfer learning, it uses the pretrained model in the model. the difference of fine tuning from transfer learning is that instead of freezing all the layers and keeping all the weight of the frozen layer's constant in the whole process, fine tuning changes the weight of the frozen layer through backpropagation. The rate of change in the weights depend on the learning rate used.

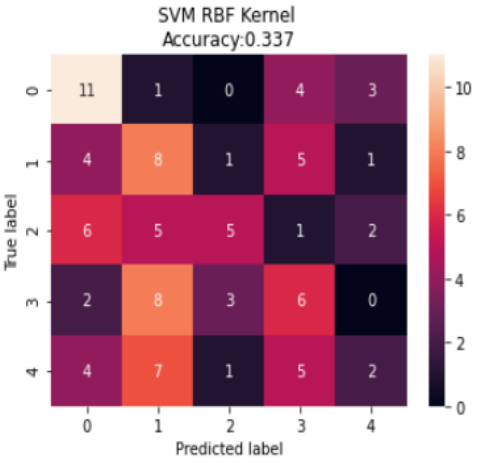
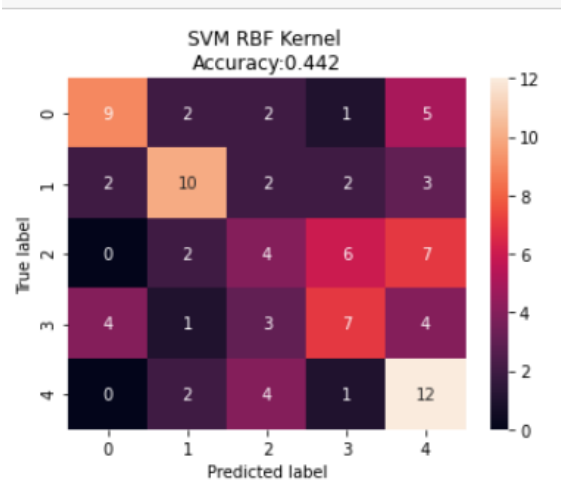
4.0 RESULTS

4.1 Classification accuracy using Bag-of-Words

4.1.1 Comparative analysis of different classifiers

Using Accuracy table, confusion matrix, Comparative analysis of ORB and HOG features has been carried out, Both SVM (linear &RBF) results have been collected.

SVM	ORB Features	HOG Features
Linear Kernel		

Radial Basis Function (RBF) Kernel		

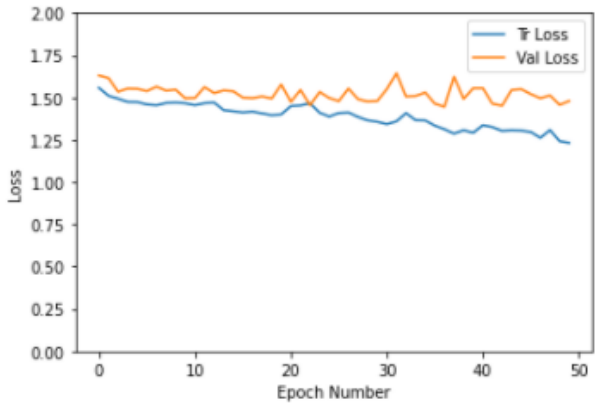
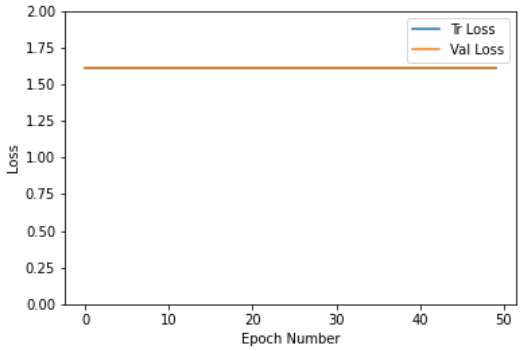
4.1.2 Comparative analysis of different number of clusters

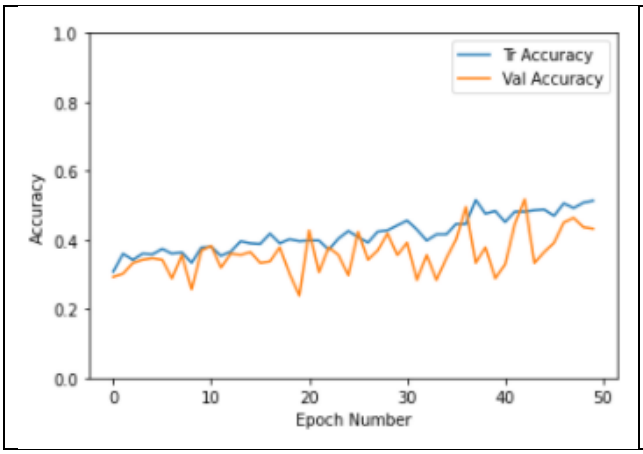
To check the effect of using different number of clusters the local features of ORB has been used.

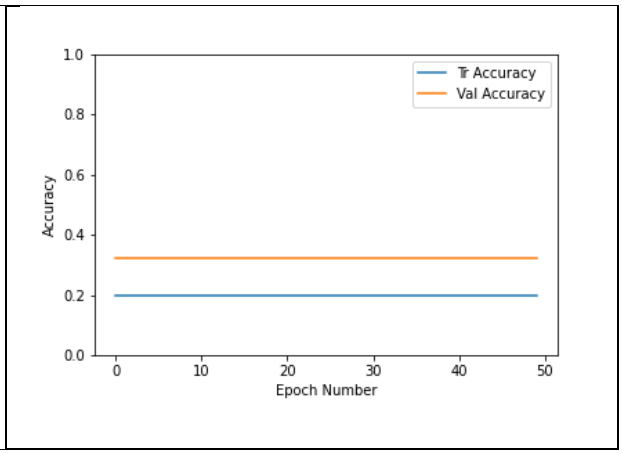
Number of k	SVM Linear Accuracy (%)	SVM RBF Accuracy (%)
100	28.4	38.9
200	29.5	44.2
500	31.6	40.0

4.2 Classification accuracy using Fully Connected Neural Networks

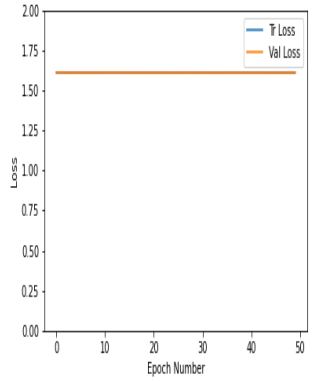
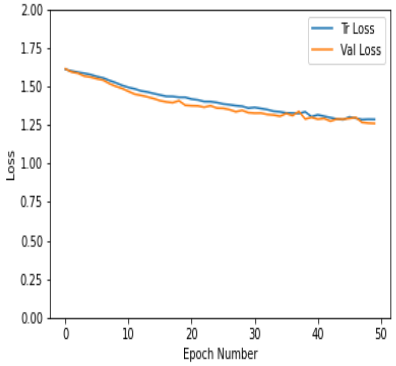
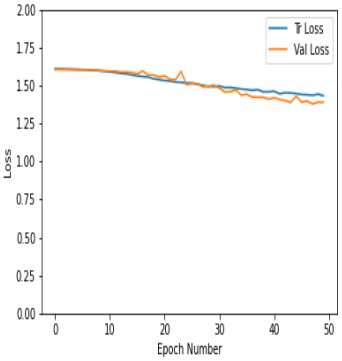
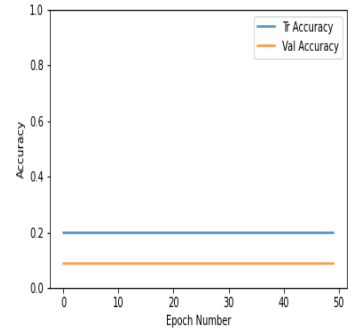
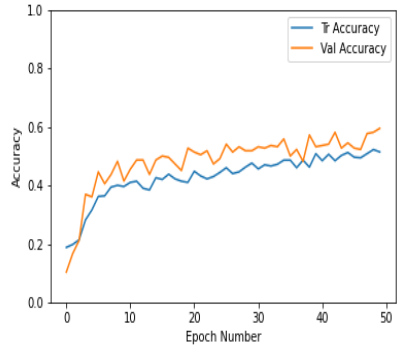
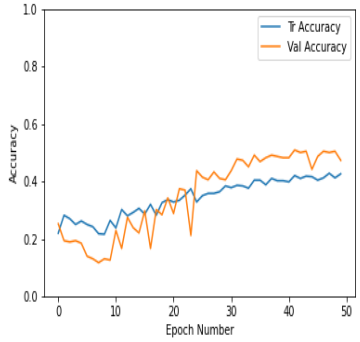
4.2.1 Comparative analysis of different architectures

Architecture1 : 3 layers of Fully Connected layers connected to the output layer	Archetucture2: 5 layers of Fully Connected layers connected to the output layer
Overall Loss:1.3105 Overall Accuracy:0.43	Overall Loss:1.609 Overall Accuracy:0.319
	



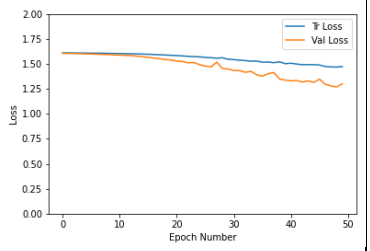
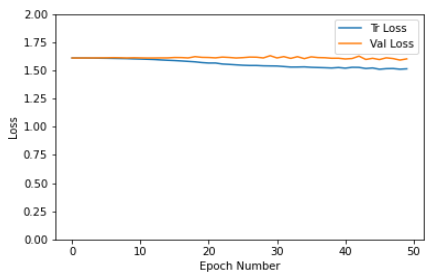
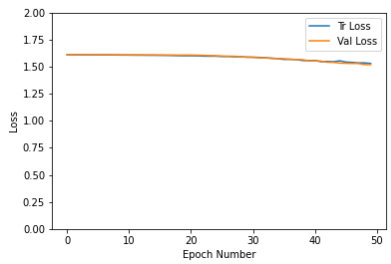
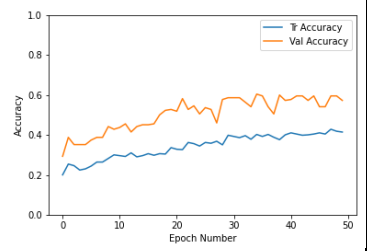
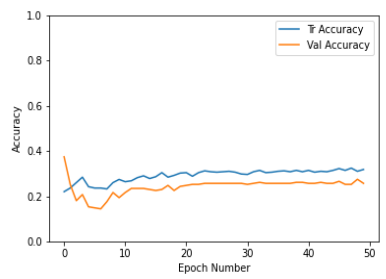
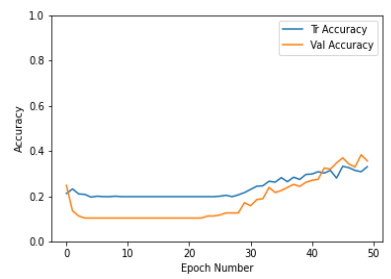


4.2.2 Comparative analysis between using Sigmoid, tanh, and ReLU as activation functions

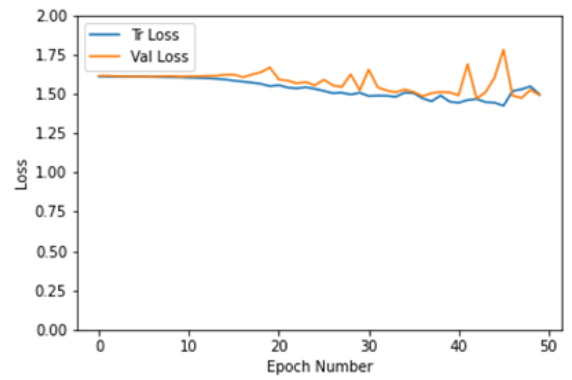
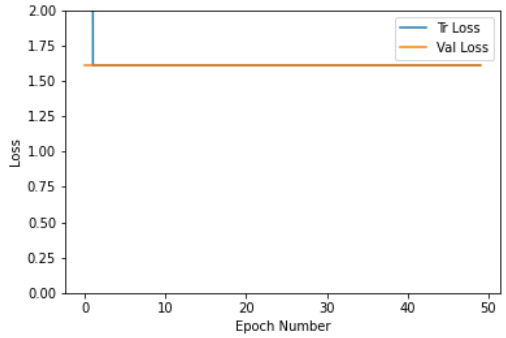
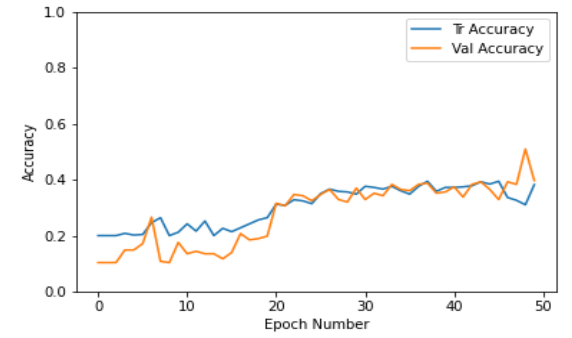
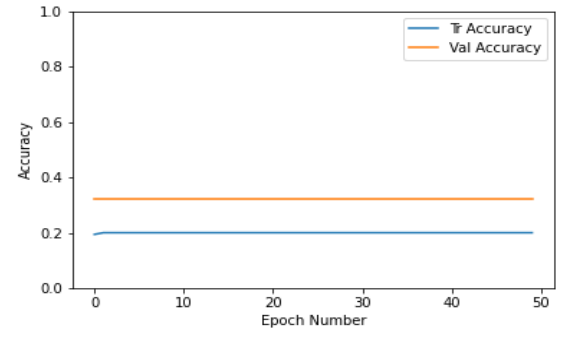
Sigmoid	Tanh	ReLU
Overall Loss:1.609 Overall Accuracy:8.5%	Overall Loss:1.25 Overall Accuracy:59.45%	Overall Loss:1.44 Overall Accuracy:47.3%
		
		

4.2.3 Effects of Dropout

To test the effect dropout Accuracy the dropout was implemented only in the 2nd layer of the fully connected neural network.

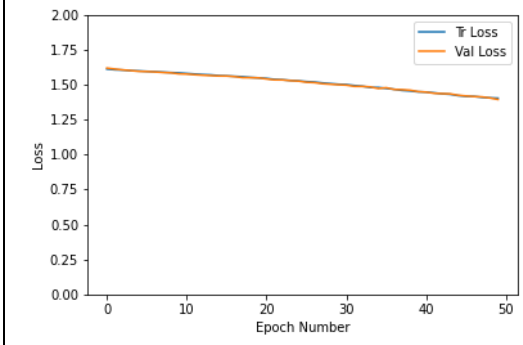
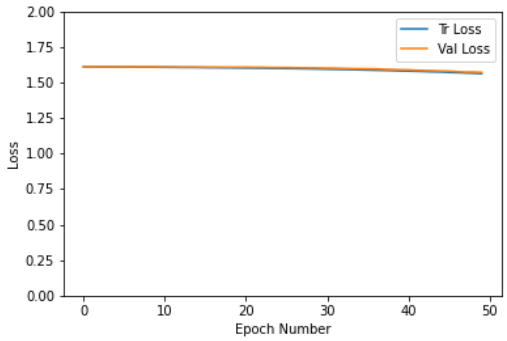
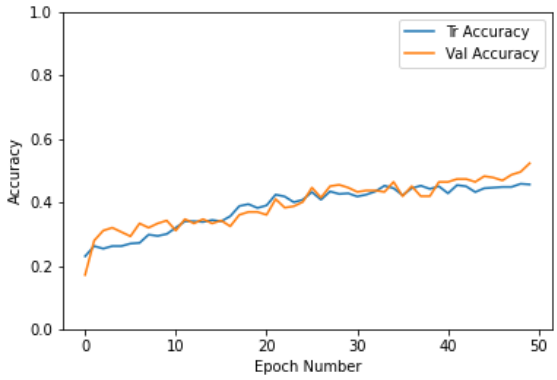
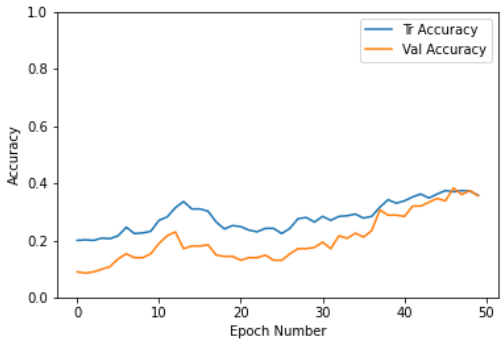
Dropout=0.2	Dropout=0.5	Dropout=0.8
		
		

4.2.4 Stochastic Gradient Descent VS ADAM

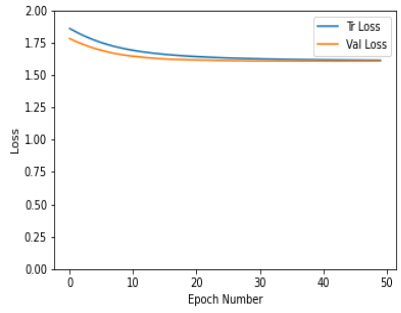
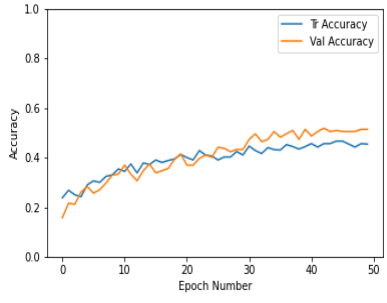
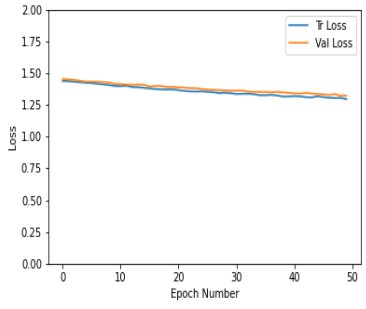
ADAM	Stochastic Gradient Descent
	
	

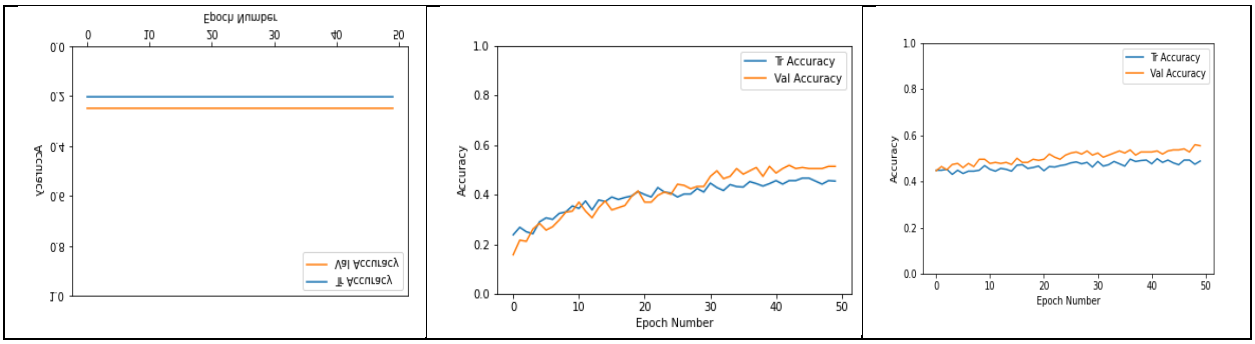
4.3 Classification accuracy using Convolutional Neural Networks

4.3.1 Classification accuracy using Convolutional Neural Networks

Architecture: COnvulation layer =2 Maxpool layer =1 Fully connected layer=3	Architecture: COnvulation layer =3 Maxpool layer =1 Fully connected layer=1
Overall Loss:1.393 Overall Accuracy:52.3%	Overall Loss:1.562 Overall Accuracy:35.5%
	
	

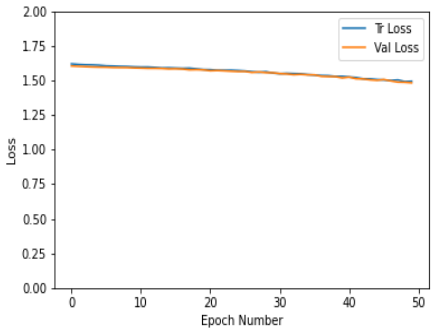
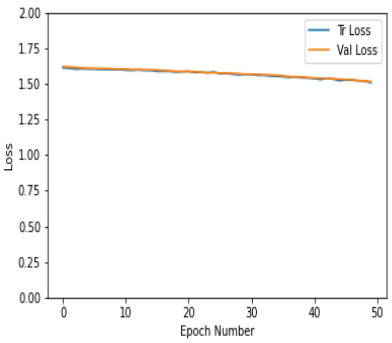
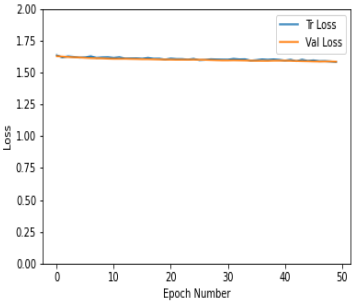
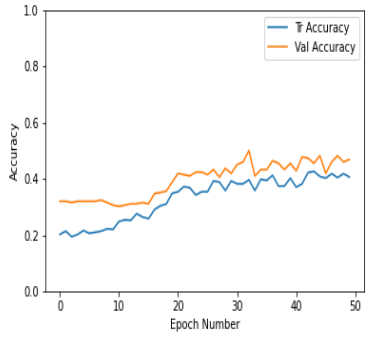
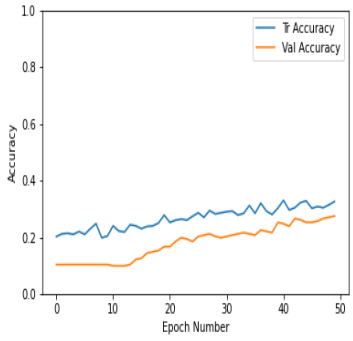
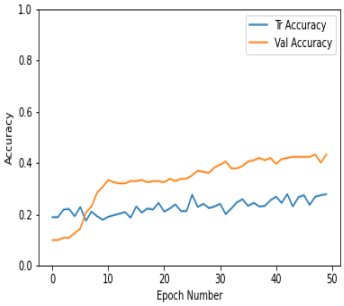
4.3.2 Comparative analysis between using Sigmoid, tanh, and ReLU as activation functions

Sigmoid	Tanh	ReLU
		

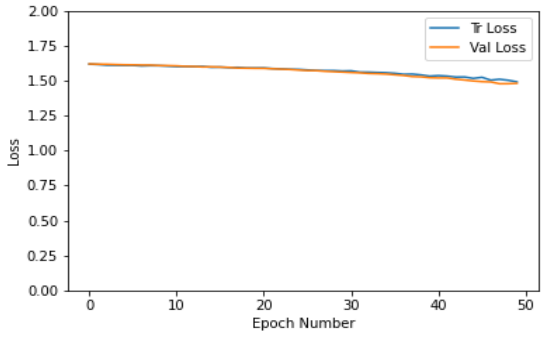
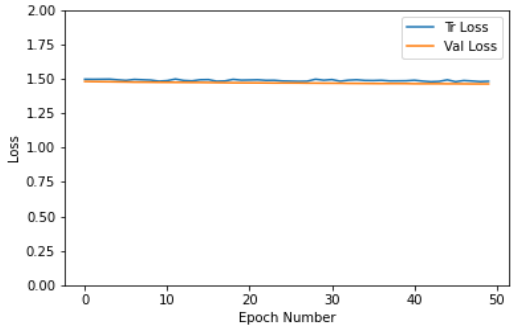
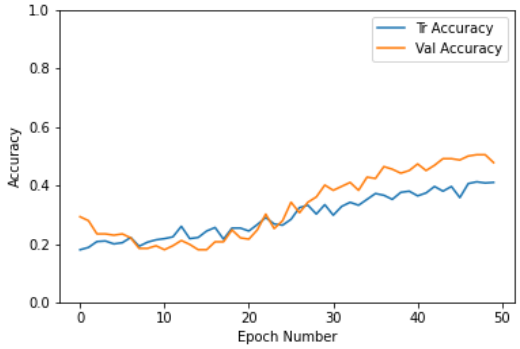
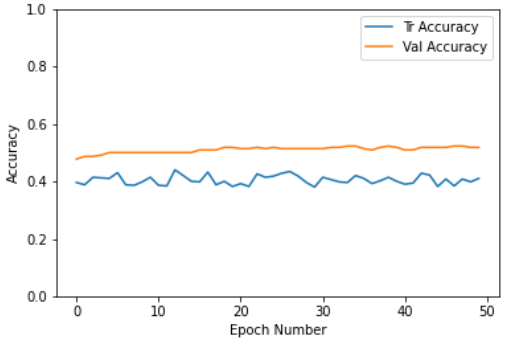


4.3.3 Effects of Dropout

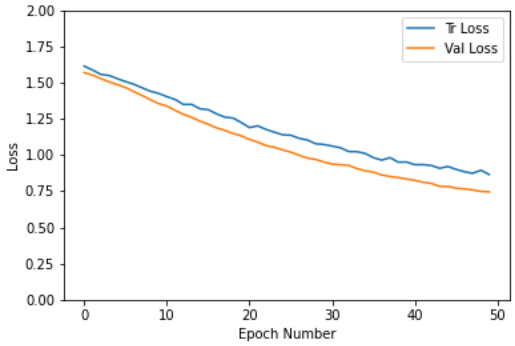
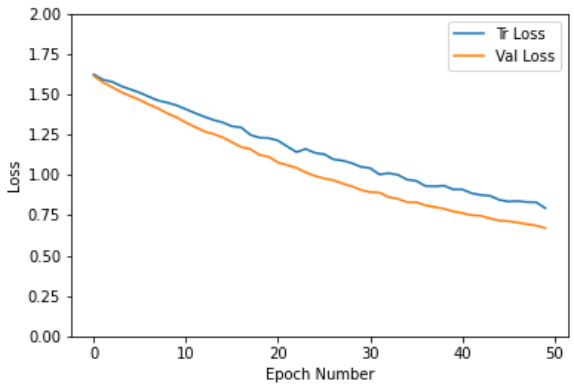
To test the effect dropout Accuracy the dropout was implemented only in the 2nd layer of the fully connected neural network

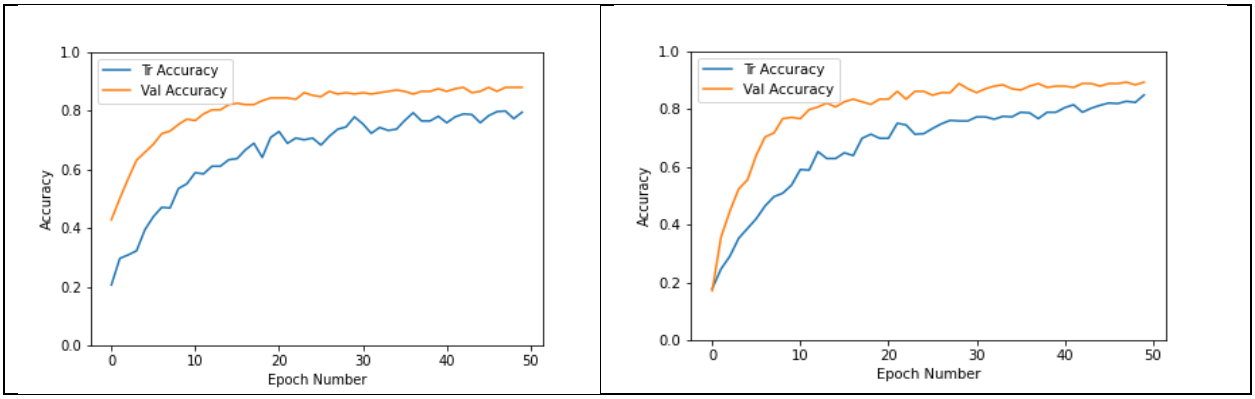
Dropout:0.1	Dropout:0.5	Dropout:0.8
		
		

4.3.4 Stochastic Gradient Descent VS ADAM

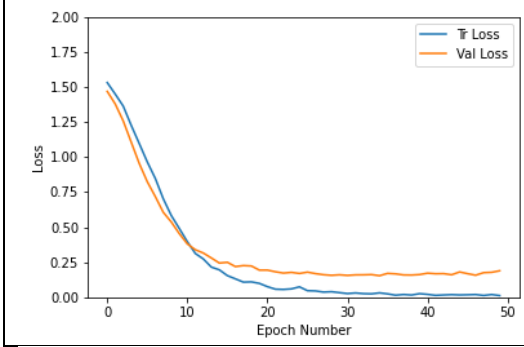
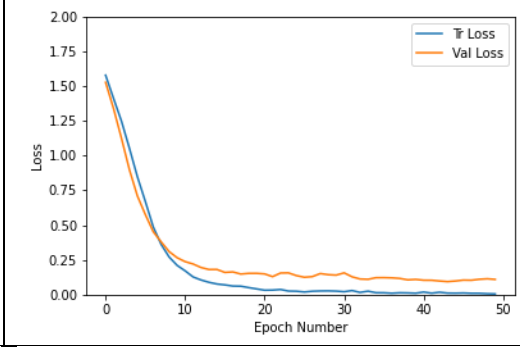
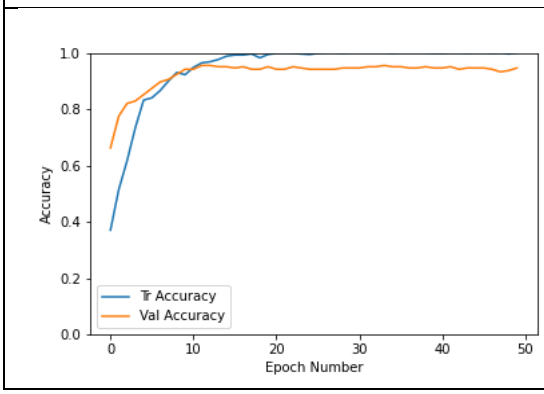
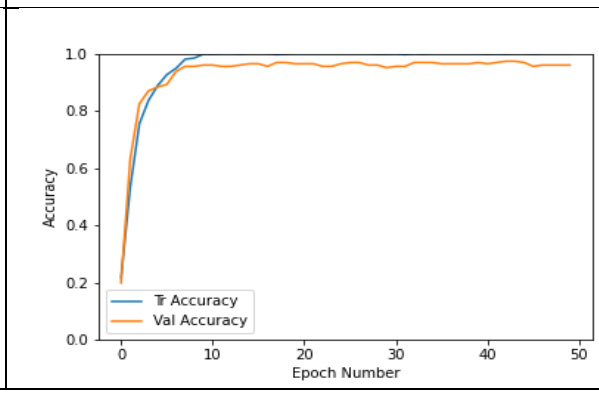
ADAM	SGD
	
	

4.4 Classification accuracy using Transfer Learning

Pretrained model:Resnet152	Pretrained model:Resnext101_32x8d
	



4.5 Classification accuracy using Fine-Tuning

Pretrained model:Resnet152	Pretrained model:Resnext101_32x8d
	
	

5.0 Github and YouTube Links

6.0 Conclusion

The dropout of the neural network image classification is important, as I t can be seen that if the dropout is low the accuracy of the training data and the validation data is big and they never intersect

while as we increase the dropout the gap close up. But if the dropout is too large like 0.8 as shown in the graphs in will intersect and continue to increase the gap after the intersection point.

The activation function that we tested ReLU showed the promising results however when we use sigmoid the accuracy and the loss of both datasets are seemed to be low constant and from there, we can conclude that relu and tanh worked on our neural networks.

As we change the optimization to SGD the accuracy dropped in most of our neural networks compare the adam optimization that we used in the rest of the image classification results. It seems that SGD optimizer depends highly on the high learning rate and doesn't function very well with low learning rate. In the tests above our learning rate was usually 0.00001.

In the overall we can visualize that the accuracy of the image classification increases as we down the list of methods, bag of words, fully Connected neural network, CNN, transfer learning, finetuning, respectively. The low accuracy in the early methods of classifier is probably due to low number of training samples. Usually in deep learning for image classifications the data set in always very large so that the machine learns and increase in its prediction accuracy. As our dataset is only 500 samples it is considered very small. However, as the usage of transfer learning and fine-tuning image classification we can observe that the accuracy is very high because it uses pretrained models that have been trained with ImageNet (massive dataset sample) so its perfectly suited to be used in a place where the quantity of the dataset is the limiting factor.