Neural Networks and Deep Learning (ECS659P/ECS7026P)
Assignment

Student Name: Vishal Choudhary
Student ID: 230798656

Introduction:

Image classification is one of the foundation part of computer vision and ability to classify image accurately plays crucial role in many field. This report is about using neural network techniques to classify CIFAR-10 dataset. The CIFAR-10 dataset has 10 classes and contain 60000 colour images (3 x 32 x 32) that's been divided into 50000 for training and 10000 for testing purpose. The aim of this study is to follow a specific architecture design and apply hyperparamter tuning on the model to generalise and predict unseen images accurately.

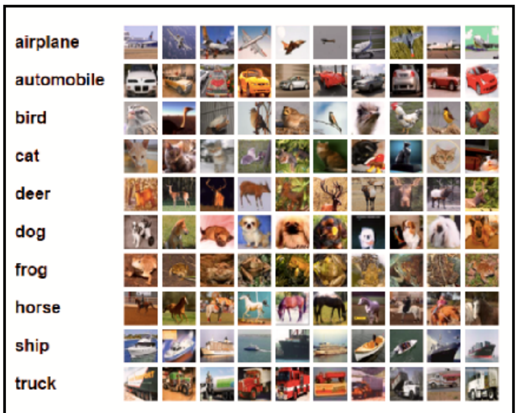


Fig. 1. CIFAR10 dataset

Basic architecture:

The architecture of basic model, referred as **Model_0,** is designed to process image of CIFAR-10 dataset by passing it through number of intermediate blocks and then passing it through an output block to make classification of image. The basic component and there working of this architecture is explained below:

- **Intermediate block:** The `IntermediateBlock` iscore fundamental processing unit of architecture. This block consist of convolutional and fully connected layers that perform two task:
  Task-1: Vector `m` computation:
  - It takes image as input having `**C**` colour channels and then it computes `C` dimensional vector `**m**` by finding the average of each colour channel separately.
  - Passes `m` through a fully connected layer which gives a vector `**a**`.
  - The vector `a` has length equal to the number of convolutional layer `L` within the block. i.e., a = [a1, a2, ..., aL]. This computation follows:
  $$a = f(Wm + b)$$
  , where 'f' is activation function, 'W' is weight matrix, and 'b' is bias vector.

  Task-2: Weighted Sum of Convolved Images:
  - It passes same image through `L` convolutional layer within block denoted as C1,C2…CL.
  - Each convolutional layer apply learned filter to image that detect features in the image, such as edges, textures and patterns. Each convolutional is followed by Batch normalization to improve stability and speed and ReLU activation adds non-linearity for complex pattern learning. Max pooling is used reduce spatial dimensionality whereas dropout regulation is used to prevent overfitting. Each convolutional layer generate its own output image.
  - Next step is to perform weighted sum of these image with weighs derived from vector `a` while ensuring output image of each convolutional layer maintain same shape as follow: **X_prime = a1*C1(X) + a2*C2(X) + ... + aL*CL(X)**
- **Output block:**
  - The output image from final intermediate block `BK` is taken as input in **`OutputBlock`.**
  - Similar to intermediate block, it computes a `C` dimensional vector `m` by averaging each colour channel of the input image.
  - Then the vector `m` is passed through multiple fully connected layer in sequence to transform into final output vector `**O**`. Vector `O` is 10 dimensional logit vector corresponding to 10 classes of CIFAR-10 dataset.

Improved architecture:

The architecture of improved model, refereed as **`CustomWideResNet16x8`** which is based on a variant on ResNet (reference). This architecture is improved version of basic architecture while ensuring following same format. Here are the key improvements implemented in the CustomWideResNet16x8:

- **DropBlock Regularization:** DropBlock is advanced regularization that is integrated inside the convolutional layer which zeros out contiguous regions of the feature map and encourage model to learn more robust feature and not rely just on specific local patterns.
- **Expanded Model Capacity**: This model uses approach inspired fromWideResNet and increased the number of channels (8 times the basic model) inside the convolutional layer. This increase allows model to capture a broader range of feature from the input image which result in better performance.
- **Deeper Convolutional Layers**: In this improved model, the number of convolutional layer within the model is configurable, which result deeper understanding of features in the image.

Methodology and Results:

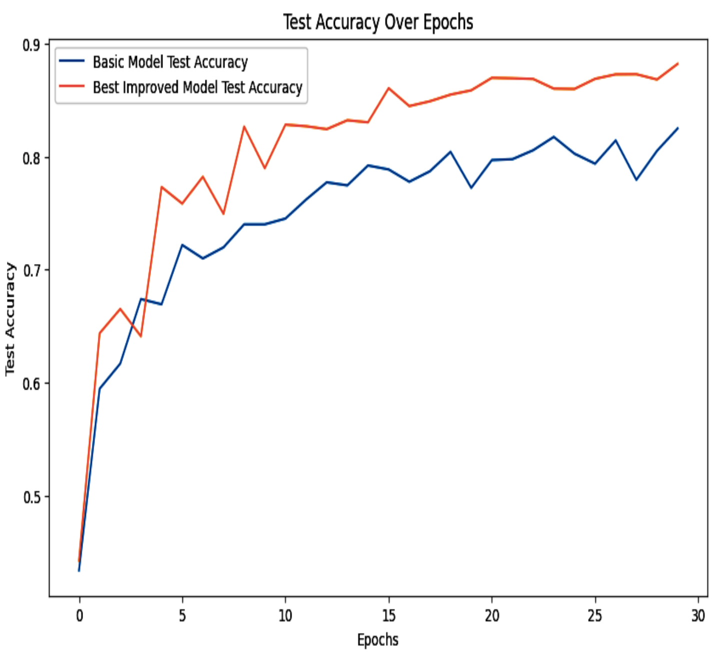
The methodology used to enhance the neural network architecture was iterative, where we have used improved architecture inspired from WideResNet model and the application of dropblock regularisation. List of **hyperparamter** and training techniques employed are:

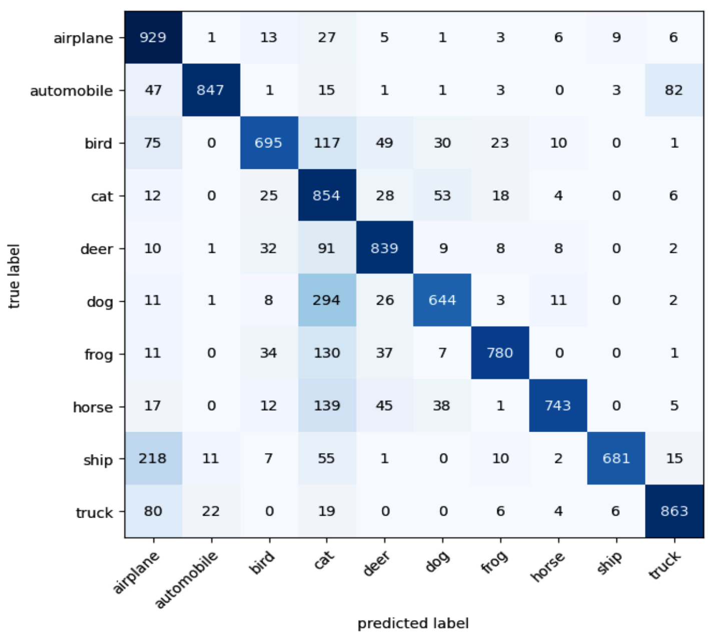
- Learning rate: 0.001, 0.01
- Batch Size: 64,128
- Number of block: 7,10
- Number of convolutional layer: 3,5
- Optimizer: Adam optimizer

The enhanced neural network showed improved performance in compare to the basic model. It yield a test accuracy of 88.18% showing a significant improvement. The table below shows the comparison between the best hyperparamter performance of improved model with the basic model.

Model	Final Train Loss	Final Train Accuracy	Final Test Loss	Final Test Accuracy
Basic Model	0.594021	0.804802	0.530998	0.824780
Best Improved Model	0.241319	0.916061	0.368046	0.881867

The graph below shows comparison between Test accuracy over epochs for both the model along with confusion matrix showing **`CustomWideResNet16x8`** proficiency in classifying images.





Discussion:

The confusion matrix along with performance metrics highlights `**CustomWideResNet16x8**` proficiency in classifying CIFAR-10 images. There is significant improvement overall in distinguishing between categories of images. The integration of using DropBlock, wider convolutional layer and deeper network has made model to learn more generalise feature rather than overfitting the training data.

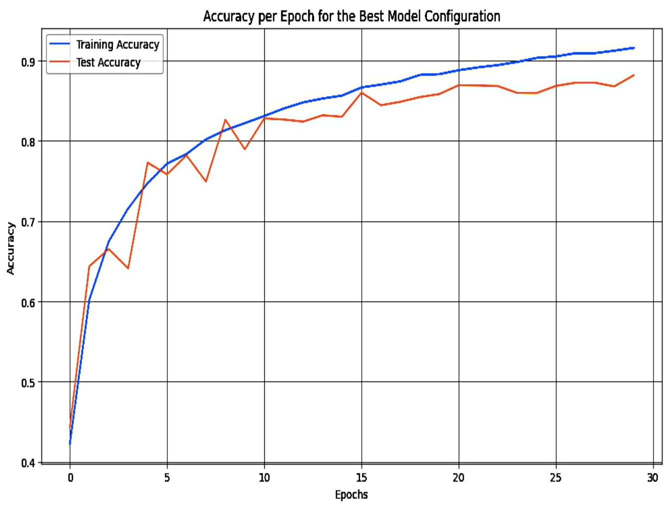
Notable, the model excels in identifying classes like airplane, truck and cats with high accuracy with 929, 863 and 854 correct predictions showing model able to capture the distinguish feature of the classes. However, confusion matrix also shows that model is confuses between few classes. For instance, model predict dog as cat for 294 instances and ship as airplane for 218 instances which could be due to visual similarity and colour similarity between those objects. Such instances shows that there is still potential space for improvement in the model.

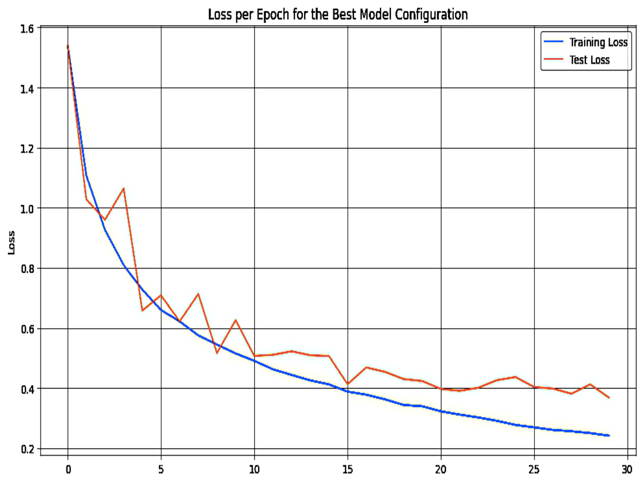
Conclusion:

The enhancement done on the basic architecture which follows a specific format delivers a superior performance on the CIFAR-10 dataset. The model showed a notable test accuracy of **88.18%**with a significant drop in losses of **0.368**. This shows that right choice of using strategic architecture along with right choice of hyperparamter selection for it.

Appendix:

Attached are plots and prediction output that are been carried out through model building process:





Prediction done by Improved Model