# Report of assignment 1

Introduction

This assignment is designed for classifying the ten classes ('airplane', 'automobile', 'bird', 'cat', 'deer','dog', 'frog', 'horse', 'ship', 'truck') of dataset CIFAR10. In our daily life, classifying the category of a given object is a common task. But when we want machine to do the classification, we need to find some specified architectures and train the model based on a large number of processed images to attain the ability to generate related features of each class. Till now, people have achieved great classification accuracy on CIFAR10 dataset. Around 2012, MCDNN got 88.8 percentage correct to this dataset, and with the development of machine learning large model, people attained 96.7 percentage correct through DenseNet in 2017 and finally achieved 99.6 percentage correct to this dataset through efficient adaptive ensembling in 2022. Hence, we can consider that this classification task is nearly solved by some of the large models.

Method

To finish this classification task, we plan to choose three different architectures. The first model we choose is the simplest linear model, and the second model is based on the basic CNN and will do batchnorm2d improvements about it, and the third model is through training a LoRA over a pretrained basic Vision Transformer with kernel size 16 on dataset laion400m. We plan to choose CrossEntropy loss as our loss function and SGD as our optimization function. Each method will be trained at least for 5 epochs, if the loss tends to be flat, we will stop training, otherwise, we will continually train the model till its loss is consistent.
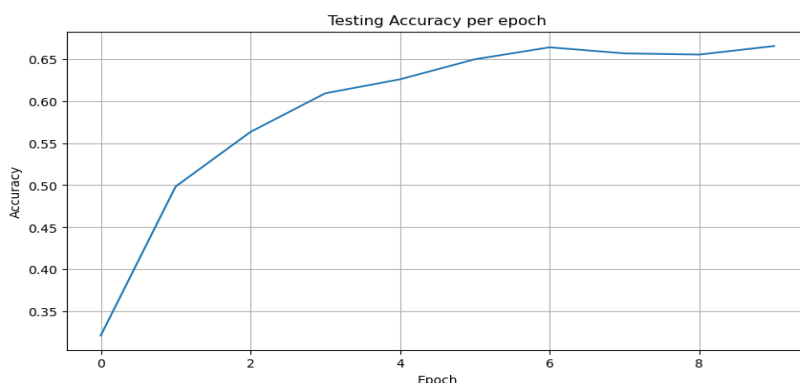
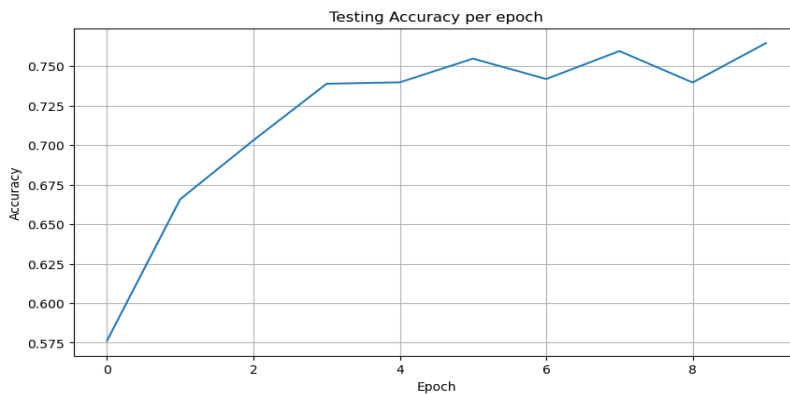### Result

Linear net (MLP)
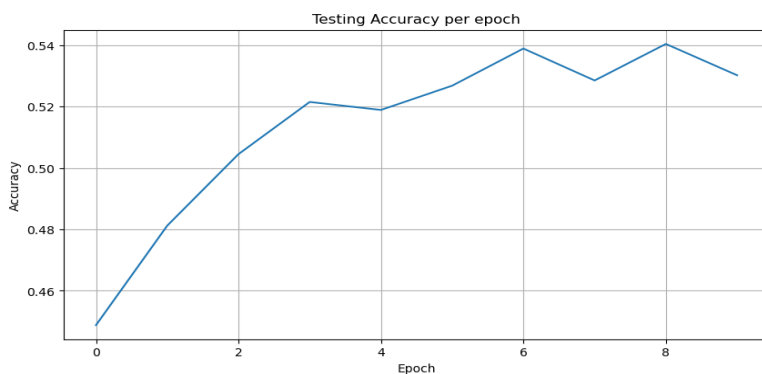

```
Accuracy for class plane is: 61.4 %
Accuracy for class car   is: 59.6 %
Accuracy for class bird  is: 43.2 %
Accuracy for class cat   is: 31.5 %
Accuracy for class deer  is: 43.5 %
Accuracy for class dog   is: 46.6 %
Accuracy for class frog  is: 65.3 %
Accuracy for class horse is: 57.4 %
Accuracy for class ship  is: 72.7 %
Accuracy for class truck is: 49.0 %
```
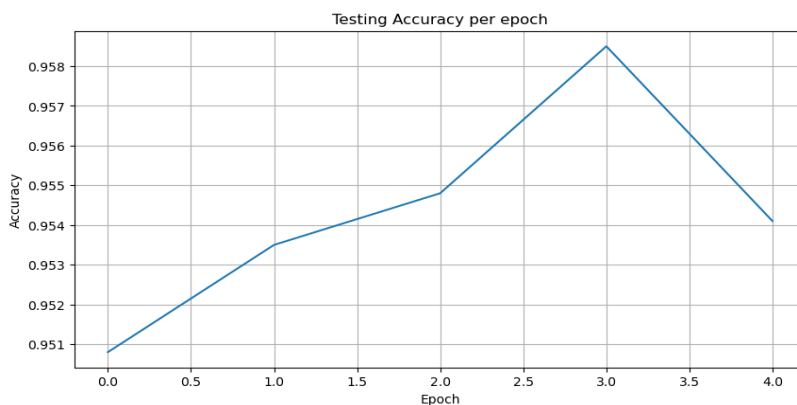
CNN


```
Accuracy for class plane is: 71.0 %
Accuracy for class car   is: 81.9 %
Accuracy for class bird  is: 52.6 %
Accuracy for class cat   is: 51.2 %
Accuracy for class deer  is: 62.3 %
Accuracy for class dog   is: 46.6 %
Accuracy for class frog  is: 69.9 %
Accuracy for class horse is: 78.4 %
Accuracy for class ship  is: 81.2 %
Accuracy for class truck is: 70.1 %
```

CNN with BN


```
Accuracy for class plane is: 70.7 %
Accuracy for class car   is: 84.9 %
Accuracy for class bird  is: 69.4 %
Accuracy for class cat   is: 71.6 %
Accuracy for class deer  is: 70.7 %
Accuracy for class dog   is: 61.2 %
Accuracy for class frog  is: 79.9 %
Accuracy for class horse is: 81.4 %
Accuracy for class ship  is: 89.5 %
Accuracy for class truck is: 85.3 %
```

Lora on ViT-16-B


```
Accuracy for class plane is: 97.1 %
Accuracy for class car   is: 98.9 %
Accuracy for class bird  is: 94.0 %
Accuracy for class cat   is: 84.9 %
Accuracy for class deer  is: 97.3 %
Accuracy for class dog   is: 94.7 %
Accuracy for class frog  is: 95.9 %
Accuracy for class horse is: 96.3 %
Accuracy for class ship  is: 99.4 %
Accuracy for class truck is: 97.4 %
```

Analysis

- Based on the results, we can find that each model has its own upper accuracy limit, start from 50% of the simplest linear model, to around 70% of the normal CNN and finally around 95% of the LoRA vision transformer model. Hence, we can find that the higher the dimension of parameters is, the better the model will perform.
- Also, we can find that batchnorm2d can greatly improve the original CNN model, from 65% to 75%. We can deduce that batch norm can efficiently help model catch the features of different classes.
- Linear model designed in this experiment somehow cannot classify the classes effectively. We guess that it is because a linear model will not process images like CNN or ViT, it can only use single pixel to calculate in one node, therefore it will not or hardly consider the relevant pixels and lose some of the information. And we infer that linear model is hardly to be improved by just adding more layers and this action may lead to overfit problem.
- Each model has its own converge speed, the simpler the task is, the more complex the model is, which will both lead to faster conversion. Surprisingly, our ViT-B-16 model with LoRA seems to converge to the finally accurate performance by consuming just a single epoch.
- Interestingly, we can find that the class cat attains the lowest class accuracy in most of the models, and the class ship achieves the highest class accuracy in most of the models. We guess this is because the features of class cat are more general, such as four legs, and two eyes, and meanwhile class ship has some unique features like sharp heads. However, the true reason for this result is unknown, it can do other further experiments to determine the reason.