NATIONAL UNIVERSITY
of Computer & Emerging Sciences

Scratch CNN Experimentation Report

Sharoz Haseeb
21-7226

Dataset used: CIFAR10

Framework used: torch `2.1.0+cu118` torchvision `0.16.0+cu118`

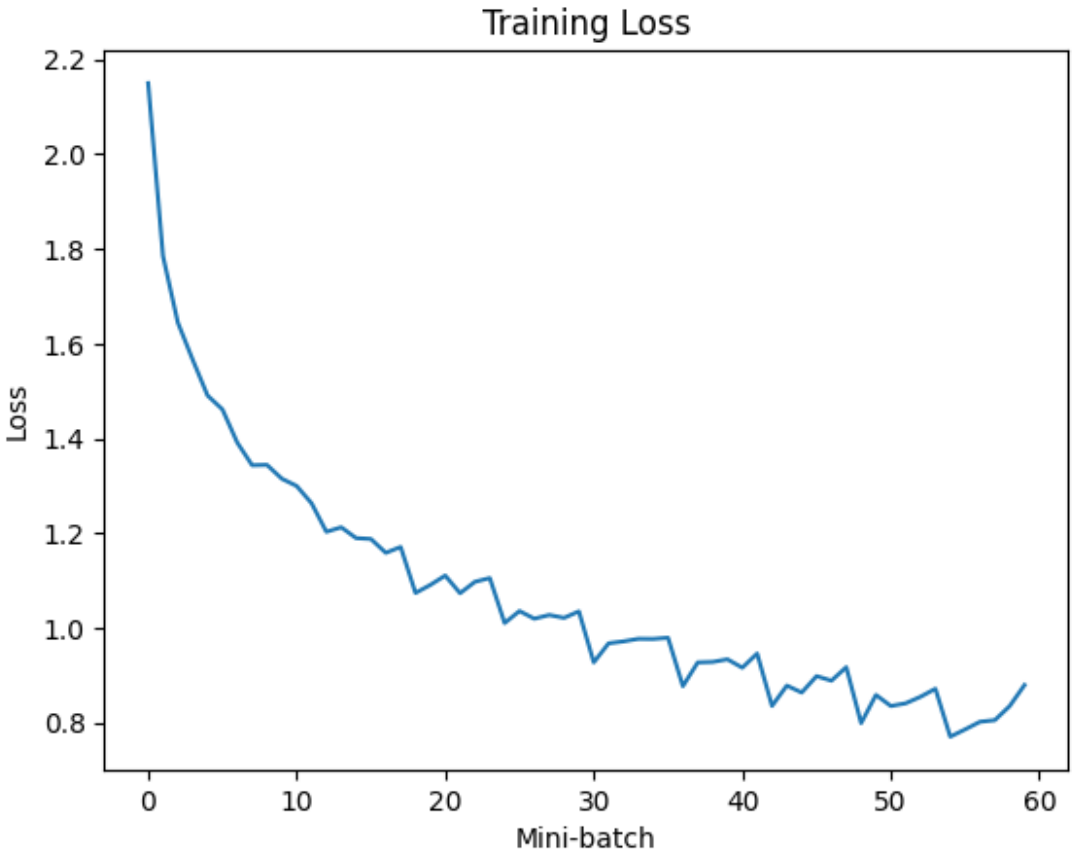
Experiment 1

	Layer type	dimension
1	Conv2d	(3, 6, 5)
2	Maxpool2d	(2, 2)
3	Conv2d	(6, 16, 5)
4	Maxpool2d	(2, 2)
5	Linear	(16*5*5, 120)
6	Linear	(120, 84)
7	Linear	(84, 10)

Note: Stochastic Gradient Descent with Momentum was used as an Optimizer

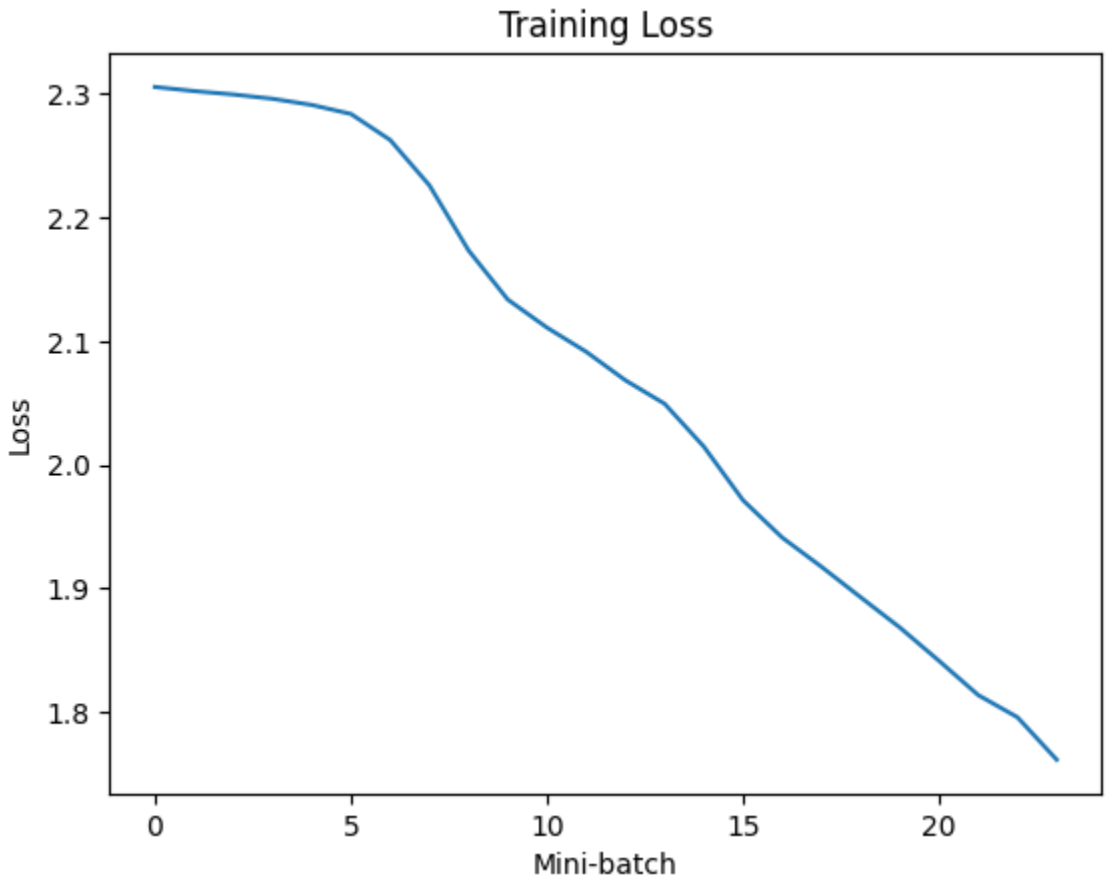
Hyperparamaters

	Param Type	Value
1	Batch size	4
2	Epochs	10
3	Lr	0.001
4	Momentum	0.9




```
Accuracy of the network on the test data: 65.50%
Precision: 0.65
Recall: 0.66
F1 Score: 0.65
```

	Param Type	Value
1	Batch size	8
2	Epochs	8
3	Lr	0.0001
4	Momentum	0.8




```
Accuracy of the network on the test data: 37.17%
Precision: 0.36
Recall: 0.37
F1 Score: 0.35
```

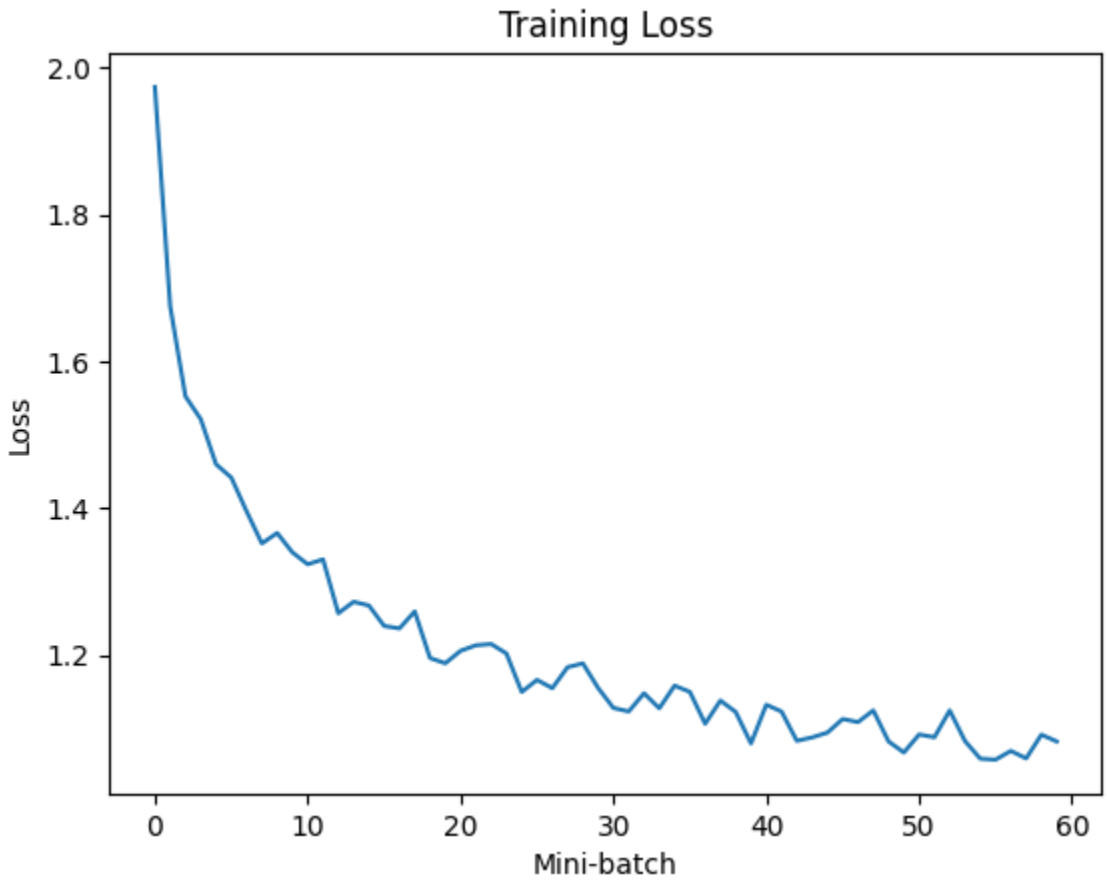
Experiment 2

	Layer type	dimension
1	Conv2d	(3, 6, 5)
2	Maxpool2d	(2, 2)
3	Conv2d	(6, 16, 5)
4	Maxpool2d	(2, 2)
5	Conv2d	(16, 32, 5)
6	Linear	(32, 120)
7	Linear	(120, 84)
8	Linear	(84, 10)

Note: ADAM was used as an Optimizer

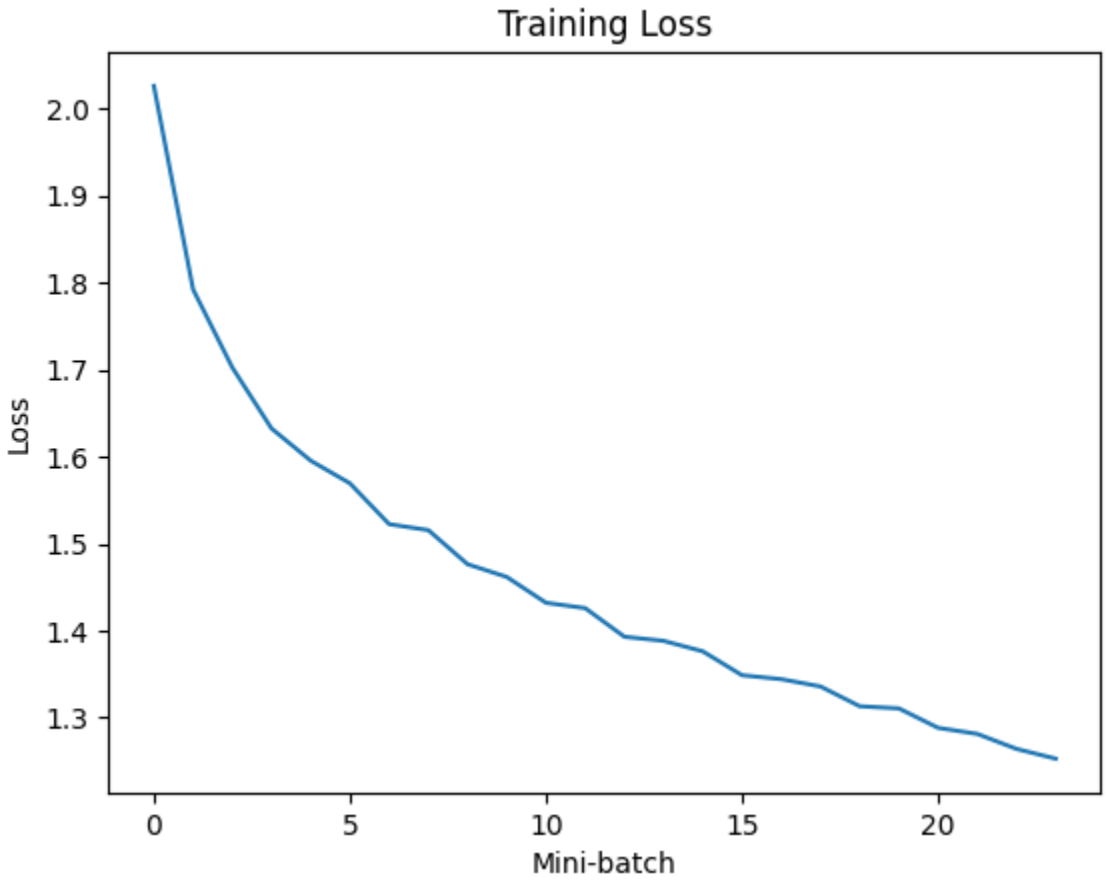
Hyperparamaters

	Param Type	Value
1	Batch size	4
2	Epochs	10
3	Lr	0.001




```
Accuracy of the network on the test data: 60.81%
Precision: 0.62
Recall: 0.61
F1 Score: 0.61
```

	Param Type	Value
1	Batch size	8
2	Epochs	8
3	Lr	0.0001




```
Accuracy of the network on the test data: 54.57%
Precision: 0.55
Recall: 0.55
F1 Score: 0.55
```

Discussion

Data Preprocessing:

The Cifar10 dataset was used and was normalized in both of the experiments to help the model converge faster.
Data Loaders were created using torch vision to serve data points efficiently in the training process.

Model architecture:

The first experiment with two conv2d layers each followers by a Maxpool at the end 2 Dense fully connected layers and an output layer was trained and the results were logged and can be observed in the previous pages, in experiment 2 another conv2d layer was added after the original 2 and without a maxpool were connected straight to the FC layer with ADAM as an optimizer instead of Momentum Stochastic Gradient Descent; the results were affected negatively and can be observed in the graphs provided.

Training process and Evaluation;

The dataset was splitted into train and test sets repectively. The train set was used to train the model and CrossEntropy was used as a loss function in both experiments.
The test set was used to calculate models with multiple accuracy metrics which are given above against all runs.