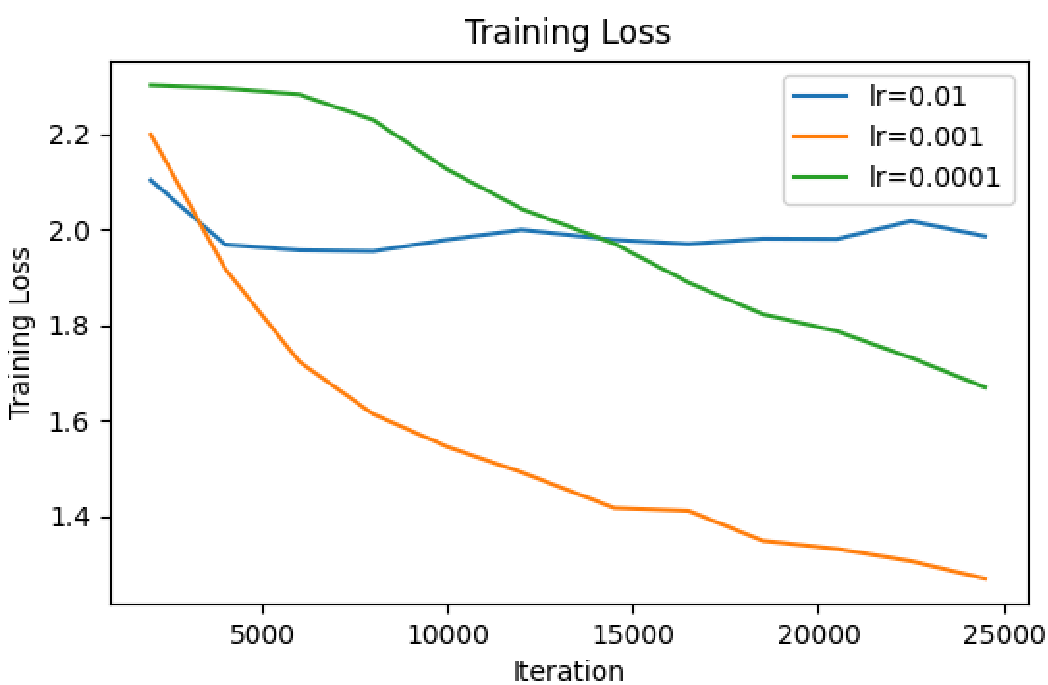
1
num_train_batches = 5
num_test_batches = 1
num_img_per_batch = 10000
num_train_img = 50000
num_test_img = 10000
size_batch_bytes = 30000
size_image_bytes = 3
size_batchimage_bytes = 30000
1a

4




Training Error
Training Error (%)
Iteration
90
80
70
60
50
40
5000
10000
15000
20000
25000
lr=0.01
lr=0.001
lr=0.0001


Test Error
Test Error (%)
Iteration
90
80
70
60
50
5000
10000
15000
20000
25000
lr=0.01
lr=0.001
lr=0.0001

4d.
At the beginning of training, the model with lr = 0.01 achieved lowest training loss, training error, and test error. However, as training progressed, its performance plateaued and even showed slight fluctuations, indicating unstable learning caused by the large learning rate.
The model with lr = 0.001 showed the most stable and consistent improvement, with both loss and error steadily decreasing throughout training.
The lr = 0.0001 model also improved gradually but at a much slower rate, suggesting that the learning rate was too small for effective convergence within only two epochs.
A large learning rate(0.01) tends to cause unstable updates and potential overfitting, while a small one(0.0001) slows down the learning process significantly. Given the current dataset and two-epoch training setup, lr = 0.001 provided the best overall performance. With longer training time and more iterations, lr = 0.0001 might eventually reach a similar performance to lr = 0.001, but not within the limited training duration of this experiment.