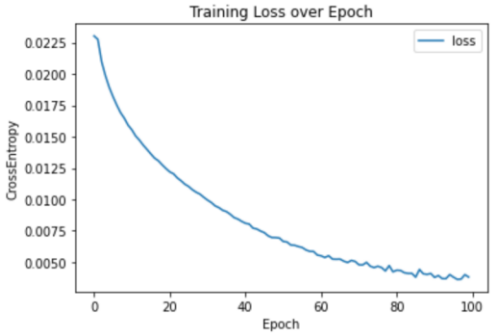
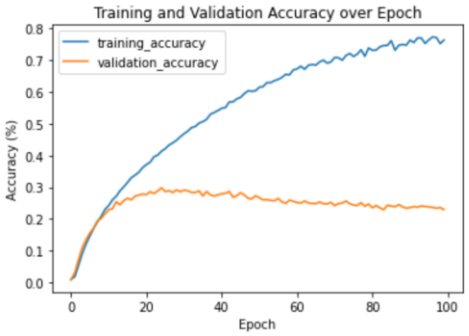
TEAM NAME: Luchao Qi

For the first model, it's a pretty simple convolutional neural network using just convolution, max pooling, and RELU. The big advantage is the speed, which means it's time-saving compared to other complicated models. Link to the model

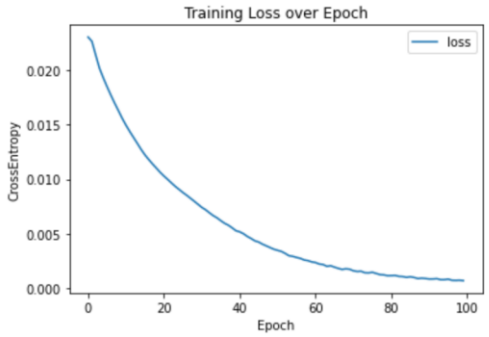
RESULT1:

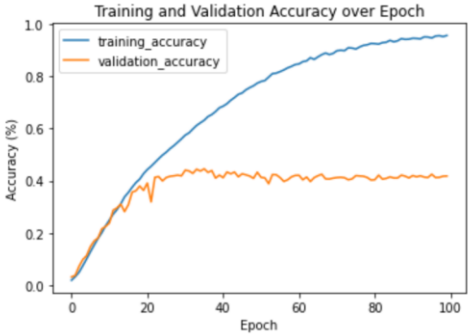




The second model is more complicated and improves the dev accuracy to ~40%, which is an improvement of previous model of ~20%. These changes including adding more convolutional layers, using dropout, batch normalization improve the accuracy a lot! Link to the model

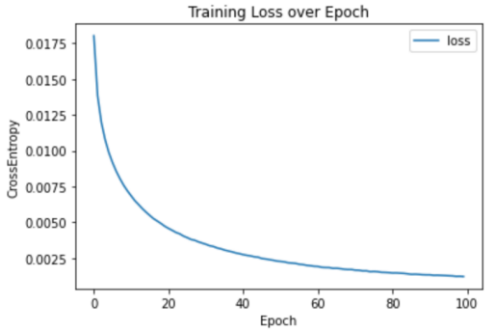
RESULT2:





The third model use different optimization method (Adam) which is an adaptive learning rate optimization algorithm that's been designed specifically for training deep neural networks. The dev accuracy converged quicker than the second model using SGD. Link to the model

RESULT3:



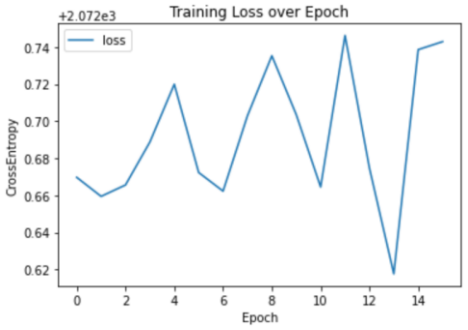
Also, some training images are rotated to help improve the dev accuracy to ~50%. Due to different initializations of weights in neural network, dev accuracy of this network shown in the picture is 46.96 %.

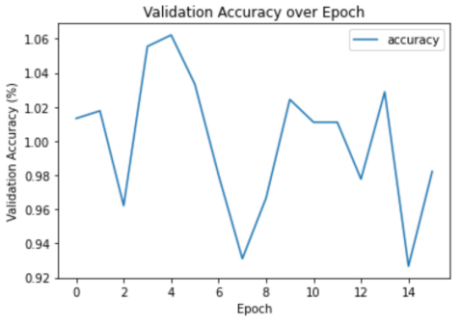
Ablation study

Epoch: From RESULT1, we can see the dev accuracy goes down slightly after reaching the highest value while the training accuracy keeps going up. In this case, we should try smaller size of epoch (early stopping) to avoid overfitting the model.

Optimizer Adam vs. SGD: From the results (RESULT2 and RESULT3) shown above, Adam help accuracy converge quicker when training the neural network

Batch Size: Smaller batch sizes make it easier to fit one batch worth of training data in memory (i.e. when using a GPU). But in practice, the loss value goes up and down and cannot converge and the following case may occur:





So the batch size should be lager during the training.

Further comparison between different batch_size settings in MNIST has demonstrated the trade-off

Batch_Size	5000	2000	1000	500	256	100	50	20	10	5	2	1
Total Epoches	200	200	200	200	200	200	200	200	200	200	200	200
Total Iterations	1999	4999	9999	19999	38999	99999	199999	499999	999999	1999999	cannot converge	
Time of 200 Epoches	1	1.068	1.16	1.38	1.75	3.016	5.027	8.513	13.773	24.055		
Achieve 0.99 Accuracy at Epoch	-	-	135	78	41	45	24	9	9	-		
Time of Achieve 0.99 Accuracy	-	-	2.12	1.48	1	1.874	1.7	1.082	1.729	-		
Best Validation Score	0.015	0.011	0.01	0.01	0.01	0.009	0.0098	0.0084	0.01	0.032		
Best Score Achieved at Epoch	182	170	198	100	93	111	38	49	51	17		
Best Test Score	0.014	0.01	0.01	0.01	0.01	0.008	0.0083	0.0088	0.008	0.0262		
Final Test Error (200 epoches)	0.0134	0.01	0.01	0.01	0.01	0.009	0.0082	0.0088	0.008	0.0662		

Further idea: Construct neural network based on ResNet to improve accuracy.