CS7015-Deep Learning
Programming Assignment 3

Jeshuren Chelladurai (CS17M017)
Ajith Kumar M (CS17M009)

March 27, 2018

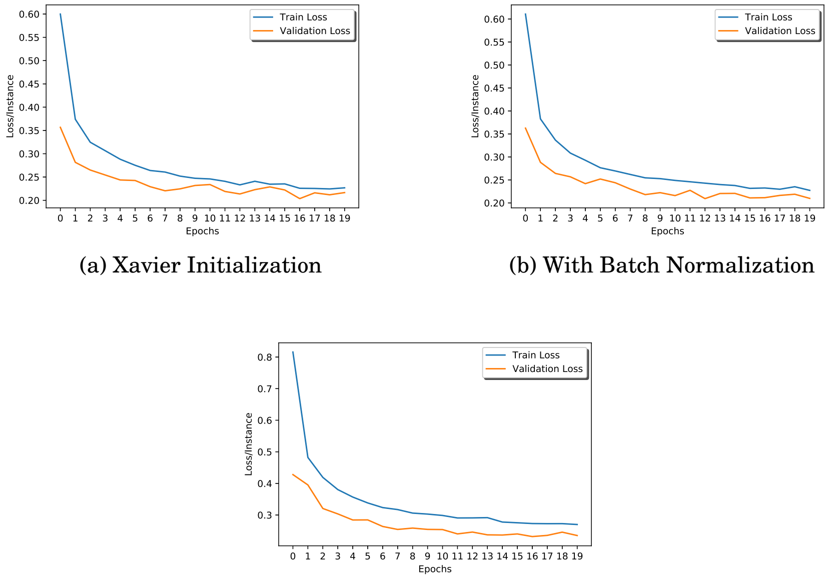
1. Architecture of our Best Model

Table 1: Best Model

Layer	**Parameters**
CONV1	32 (3x3) kernels, same padding, ReLU
CONV2	64 (3x3) kernels, same padding, ReLU
POOL1	(2x2) max pooling, (2x2) strides
DROPOUT1	0.4 dropout rate
CONV3	128 (3x3) kernels, same padding, ReLU
CONV4	256 (3x3) kernels, same padding, ReLU
POOL2	(2x2) max pooling, (2x2) strides
DROPOUT2	0.4 dropout rate
FC1	128 units ,ReLU
DROPOUT3	0.4 dropout rate
FC2+SOFTMAX	10 units , Linear

Validation Accuracy Achieved : 94.58%

2. Learning Curve



(a) Xavier Initialization

(b) With Batch Normalization

(c) He Initialization

Figure 1: Learning Curve for the model described in the assignment problem statement

3. Performance on the test data for the best model
   Kaggle Public Leaderboard F-score : 94.2
   Kaggle Private Leaderboard F-score : TBD

4. Training Parameter Settings for the best model :

   - Adam Optimizer with initial learning rate = 0.001 and retraining with learning rate = 0.0001.
   - Batch size : 100
   - Cross-entropy loss

5. Dimensions of Input and Output at Each Layer

Table 2: Dimensions of Input and Output at Each Layer

Layer	**Dimension of Input**	**Dimension of Output**
Conv1	28 × 28 × 1	28 × 28 × 64
Pool1	28 × 28 × 64	14 × 14 × 64
Conv2	14 × 14 × 64	14 × 14 × 128
Pool2	14 × 14 × 128	7 × 7 × 128
Conv3	7 × 7 × 128	7 × 7 × 256
Conv4	7 × 7 × 256	7 × 7 × 256
Pool3	7 × 7 × 256	3 × 3 × 256
FC1	1 × 2304	1024 × 1
FC2	1024 × 1	1024 × 1
Softmax	1024 × 1	10 × 1

6. Parameters in the network

Table 3: Parameters in the Conv and FC Layers

Layer	**Weights**	**Biases**	**Total**
Conv1	576(3×3×64)	64	640
Conv2	73,728(3×3×64×128)	128	73,856
Conv3	294,912(3×3×128×256)	256	295,168
Conv4	589,824(3×3×256×256)	256	590,080
Total Parameters in Conv Layers(I)			959,744
FC1	2,359,296(2304×1024)	1024	2,360,320
FC2	1,048,576(1024×1024)	1024	1,049,600
Softmax	10240(1024 × 10)	10	10,250
Total Parameters in FC Layers(II)			3,420,170
Total Parameters in the Network(I+II)			4,379,914

7. Neurons in the Network

Table 4: Neurons in the Conv and FC Layers

Layer	**Neurons**
Conv1	64
Conv2	128
Conv3	256
Conv4	256
FC1	1024
FC2	1024
Softmax	10
Total	2506

8. Effect of Batch Normalization

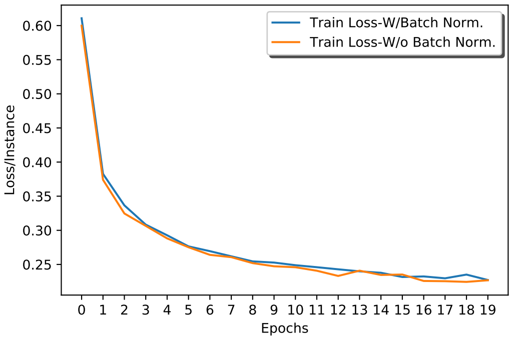


Figure 2: Plot of the learning curve between a model trained with and without batch normalization with the other settings remaining constant.

Effect of different initializers

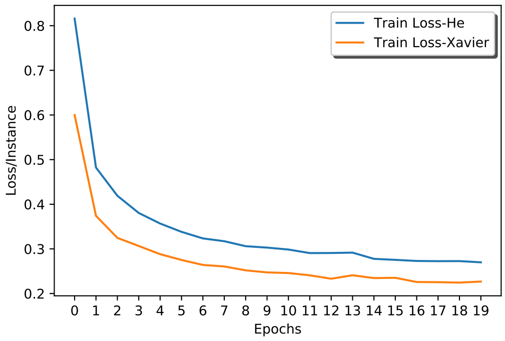


Figure 3: Plot of the learning curve between a model trained with Xavier and He initializers with the other settings remaining constant.

9. Plot of 64 layer 1 filters

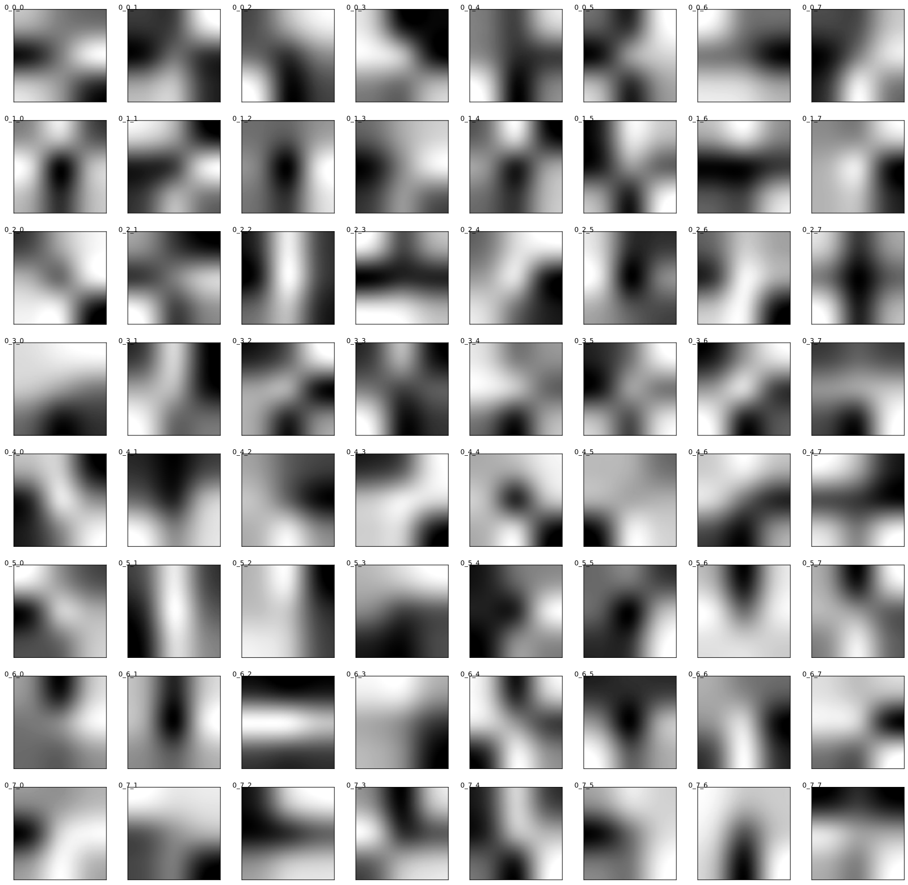
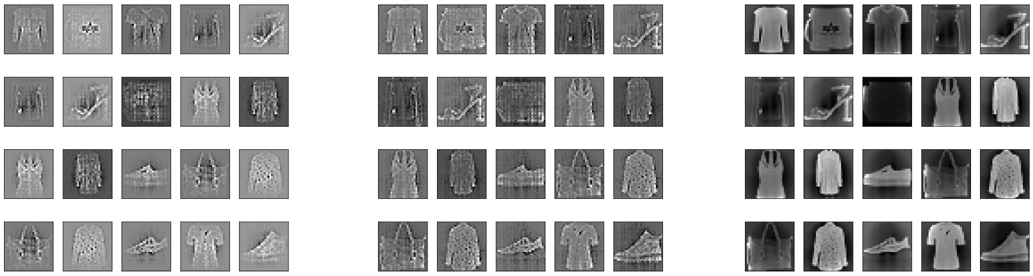


Figure 4: Plot of the 64 filters in the first layer. Rescaled to 300×300 for easier viewing.

10. Guided backpropogation



(a) One Neuron

(b) 10 Neurons

(c) Original Image

Figure 5: Guided backpropogation

11. Fooling the network

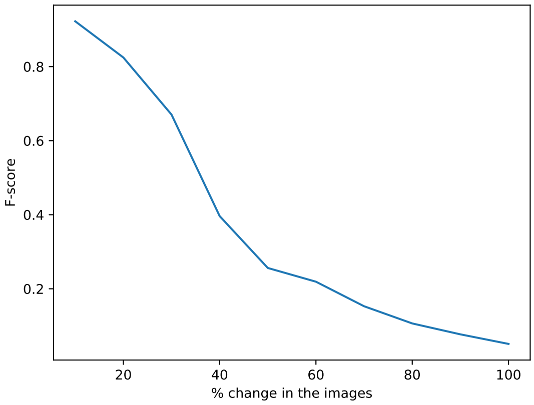


Figure 6: Decrease in F-score for random % of change in pixels in the image

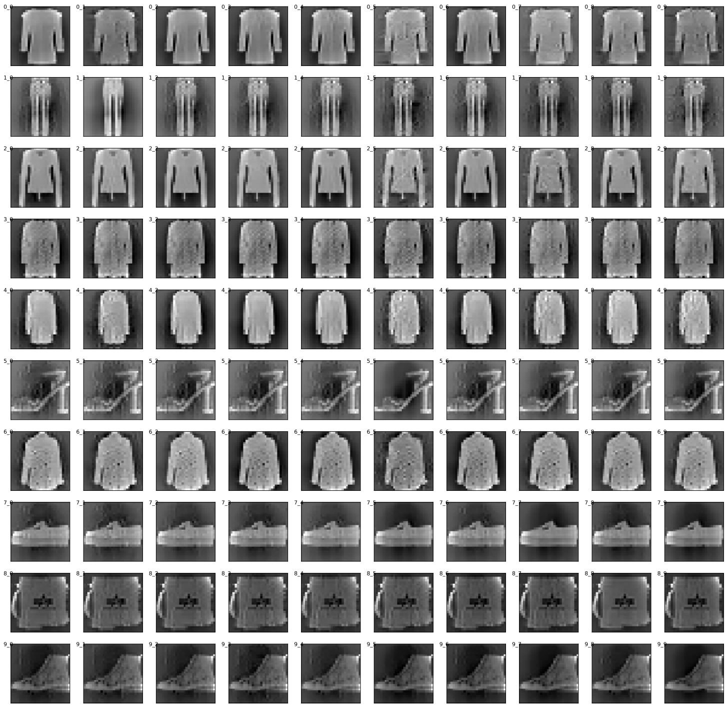


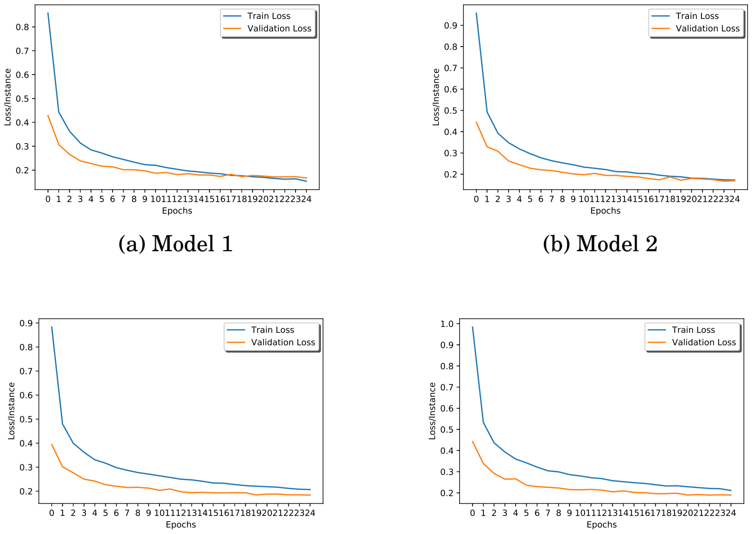
Figure 7: Plot of the images used to fool the network. The images along the diagonal are the original ones, whereas the off-diagonal images are the modified images used to make the network think that they are from the other class.

12. HyperParameter Tuning

We tried different hyper-parameter settings for the number of convolutional layers, filter size, activations etc., The models that we experimented are given in Table 5

Table 5: Different Model Configurations

Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
Conv1 - 32 3x3	Conv1 - 32 3x3	Same as Model1 with Data Augumentation	Same as Model2 with Data Augumentation	Same as Model1	Same as Model 1	Same as Model 2	Same as Model 2
Conv2 - 64 3x3	Conv2 - 64 3x3						
MaxPool - 2x2 w/ dropout(0.5)	MaxPool - 2x2 w/ dropout(0.5)						
Conv3 - 128 3x3	Conv3 - 128 3x3						
Conv4 - 256 3x3	Conv4 - 256 3x3						
MaxPool 3x3 w/ dropout(0.5)	MaxPool - 3x3 w/ dropout(0.5)						
FC1 - 256 (Leaky ReLU) w/dropout(0.5)	FC1 - 256 (ReLU) w/dropout(0.5)						
FC2 - 256 (Leaky ReLU) w/dropout(0.5)	FC2 - 256 (ReLU) w/dropout(0.5)						
Softmax - 10	Softmax - 10						
Learning Rate - 0.001	Learning Rate - 0.001			LR - 0.01	LR - 0.001	LR - 0.001	LR - 0.001
Batch Size - 500	Batch Size - 500			Batch Size - 100	Batch Size - 1000	Batch Size -100	Batch Size - 1000



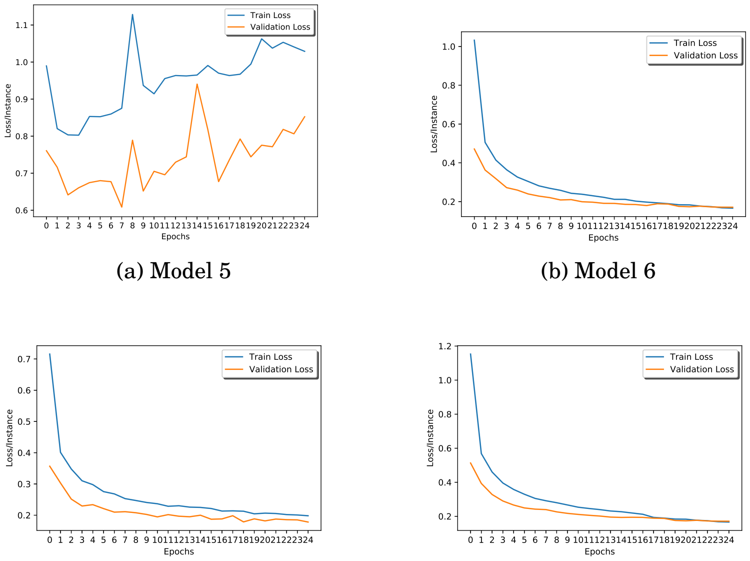
(a) Model 1

(b) Model 2

(c) Model 3

(d) Model 4

Figure 8: Learning Curves



(a) Model 5

(b) Model 6

(c) Model 7

(d) Model 8

Figure 9: Learning Curves

13. Data Augumentation

We tried the following data augumentation,

- Vertical Flip
- Horizontal Flip
- Rotation
- Adding Random Noise