ECS 189G-001

Deep Learning

Winter 2024

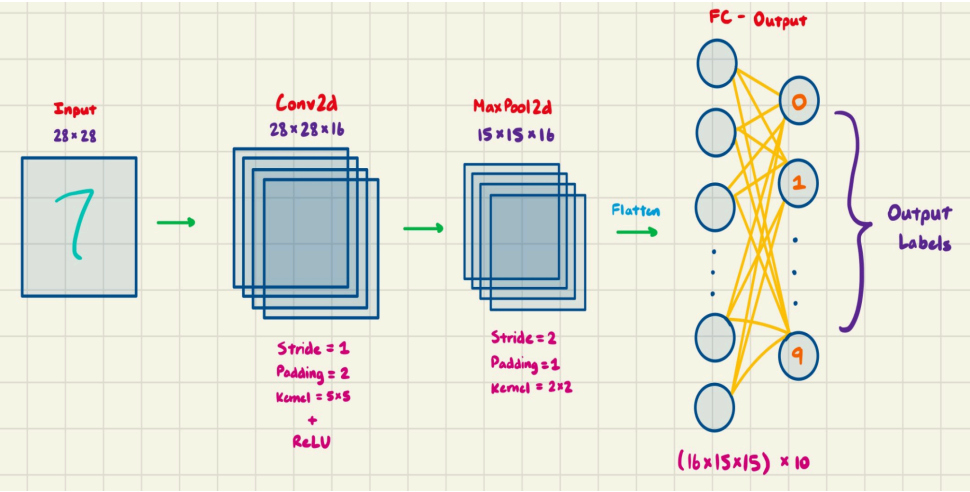
Stage 3 Report

Team Name: TheTransformers		
Aditya Seth	xxxxxxxxxx	aseth@ucdavis.edu
Parth Pawar	xxxxxxxxxx	pupawar@ucdavis.edu
Brian Li	xxxxxxxxxx	xqnli@ucdavis.edu
Ryan Yu	xxxxxxxxxx	ryuyu@ucdavis.edu

Section 1: Task Description

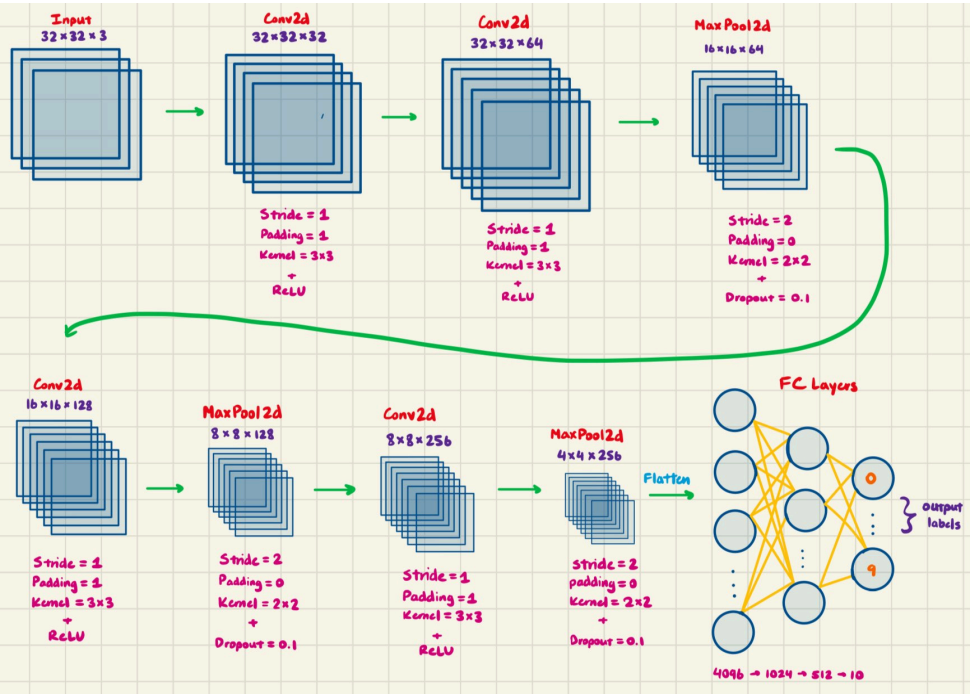
Build and train 3 Convolutional Neural Networks (CNNs) on 3 different datasets.

Section 2: Model Description



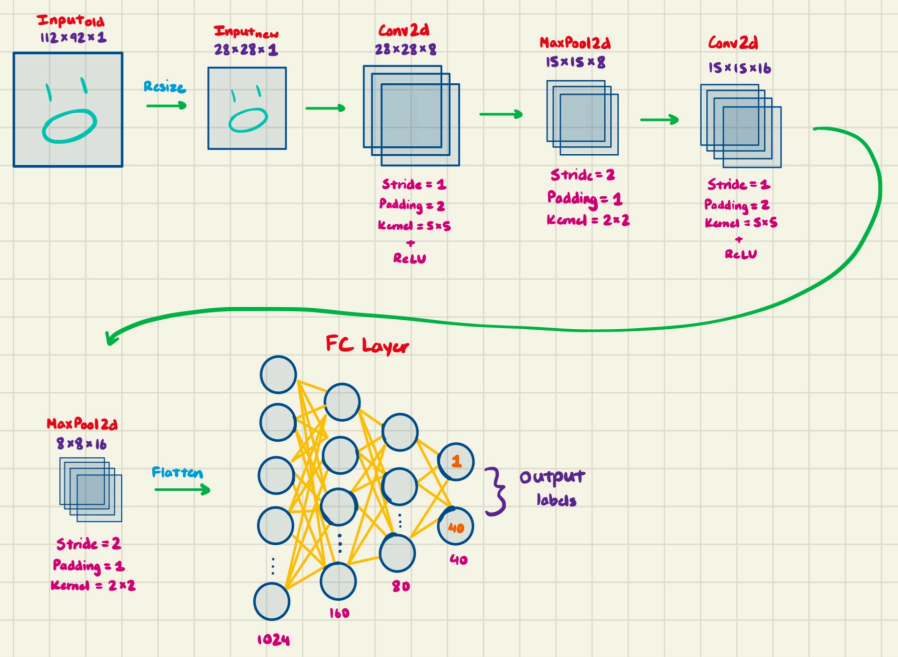
MNIST

The model has 1 convolutional layer with 1 input channel (because grayscale) and 16 output channels. A 5x5 kernel, stride of 1, & padding of 2 is applied in the convolution, followed by the ReLU activation. The result is fed into a max pooling layer, which is then flattened to a 15x15x16 fully connected layer to 10 output neurons.



CIFAR-10

The model contains four convolutional layers each with stride 1, padding 1, 3x3 kernels, and ReLU activation functions. Additionally, it has three max pooling layers consisting of stride 2, padding 0, 2x2 kernels, and a dropout rate of 0.1. Lastly, it has 3 fully connected neural layers that reduce our features into 10 potential outputs. The neuron in the output layer with the highest value is selected as the predicted label. Dropout is utilized in our model here as a way to reduce the likelihood of overfitting.



ORL

Before we process our image, we resize it from 112x92x1 to 28x28x1 to reduce computational intensity. Afterwards, we pass it into our model which consists of two convolutional layers of stride 1, padding 2, 5x5 kernels, and ReLU. There also consists of two max pooling layers of stride 2, padding 1, and 2x2 kernels used to condense our features and remove noise. FInally, we flatten our processed features into 1024 neurons, which are reduced to 160, 80, and lastly 40 for the output labels.

Section 3: Experiment Settings

3.1 Dataset Description

3 datasets were provided, each pre-partitioned into training and testing sets. The MNIST dataset has 60,000 grayscale train images of handwritten digits 0 to 9, with 10,000 test images. The CIFAR-10 dataset contains 50,000 RBG images across 10 classes, with 10,000 test images. Lastly, the ORL dataset is made up of 360 training and 40 testing images of human faces.

3.2 Detailed Experimental Setups

For all 3 models, we used mini-batch gradient descent so that less computational power and memory is needed when training. It also may help improve generalizability. Batch size of 64 was used for MNIST, 256 for CIFAR, and 41 for ORL respectively. Furthermore, we used Adam for optimization, Cross Entropy for loss, and $1e^{-3}$ for the learning rate.

MNIST: We trained the model for 1 epoch because we noticed that was enough for the model to converge.

CIFAR: We normalized the images to ensure consistent pixel values, which reduces learning time and improves accuracy. We opted to apply max pooling after two consecutive convolutional layers to allow the network to extract more detailed features from the images before extraneous information is removed in pooling. The max number of epochs was set to be 50 to ensure the loss converged. We also included support for CUDA for faster training time.

ORL: During our dataset preparation, we resized the input images from 112x92 to 28x28 pixels as a way to speed up computation. For this dataset, we used the minimum batch size, which is 41, since the classes are labeled from 0 to 40. We set the maximum number of epochs as 50. We determined 50 epochs as the max because that was enough for the model to converge.

Please refer to section 2 for details regarding activation functions and layer dimensions for all 3 models.

3.3 Evaluation Metrics

In the experiment, we will use Accuracy, Precision, Recall, F1 score as our evaluation metrics.

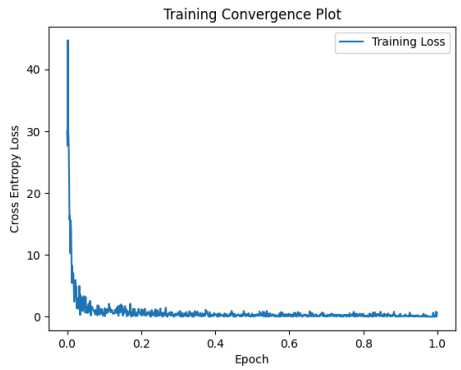
Using the classification_report function from the sklearn.metrics library, we obtained the accuracy, precision, recall, and F1 score for all the epochs during training and testing. Out of these four evaluation metrics, we aimed for models with a high accuracy and F1 score when processing our test set. Accuracy is our main focus mainly because we aimed for the models to have high capabilities in correctly labeling the images. We also put slight emphasis on F1 score to ensure that our model is balanced when it comes to labeling correctly and incorrectly.

3.4 Source Code

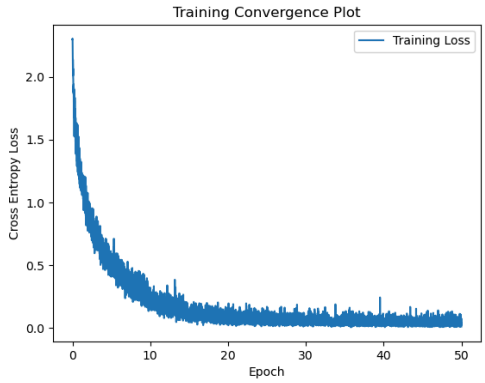
https://github.com/adityaseth8/ecs189g . The repository does not contain the datasets due to their size.

3.5 Training Convergence Plot

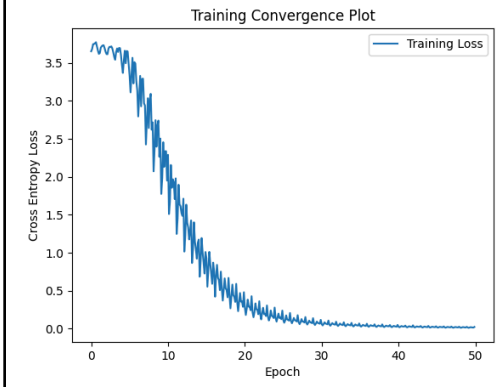
MNIST



CIFAR-10



ORL



3.6 Model Performance

MNIST

	precision	recall	f1-score	support
0	0.99	0.97	0.98	980
1	0.99	0.99	0.99	1135
2	0.97	0.97	0.97	1032
3	0.99	0.92	0.95	1010
4	0.95	0.98	0.97	982
5	0.87	0.99	0.93	892
6	0.99	0.92	0.95	958
7	0.97	0.96	0.96	1028
8	0.91	0.97	0.94	974
9	0.98	0.91	0.94	1009
accuracy			0.96	10000
macro avg	0.96	0.96	0.96	10000
weighted avg	0.96	0.96	0.96	10000

```
************ Overall Performance ************
CNN MNIST Accuracy: 0.9589 +/- None
************ Finish ************
```

CIFAR-10

	precision	recall	f1-score	support
0	0.83	0.81	0.82	1000
1	0.88	0.91	0.89	1000
2	0.74	0.62	0.68	1000
3	0.57	0.62	0.59	1000
4	0.83	0.65	0.73	1000
5	0.64	0.70	0.67	1000
6	0.70	0.89	0.79	1000
7	0.87	0.79	0.83	1000
8	0.86	0.89	0.87	1000
9	0.87	0.84	0.85	1000
accuracy			0.77	10000
macro avg	0.78	0.77	0.77	10000
weighted avg	0.78	0.77	0.77	10000

```
************ Overall Performance ************
CNN CIFAR Accuracy: 0.7727 +/- None
```

ORL

	precision	recall	f1-score	support
1	1.00	1.00	1.00	1
2	1.00	1.00	1.00	1
3	1.00	1.00	1.00	1
4	1.00	1.00	1.00	1
5	0.50	1.00	0.67	1
6	1.00	1.00	1.00	1
7	1.00	1.00	1.00	1
8	1.00	1.00	1.00	1
9	1.00	1.00	1.00	1
10	1.00	1.00	1.00	1
11	1.00	1.00	1.00	1
12	1.00	1.00	1.00	1
13	1.00	1.00	1.00	1
14	1.00	1.00	1.00	1
15	1.00	1.00	1.00	1
16	1.00	1.00	1.00	1
17	1.00	1.00	1.00	1
18	1.00	1.00	1.00	1
19	1.00	1.00	1.00	1
20	1.00	1.00	1.00	1
21	1.00	1.00	1.00	1
22	1.00	1.00	1.00	1
23	0.50	1.00	0.67	1
24	1.00	1.00	1.00	1
25	1.00	1.00	1.00	1
26	1.00	1.00	1.00	1
27	1.00	1.00	1.00	1
28	1.00	1.00	1.00	1
29	0.50	1.00	0.67	1
30	1.00	1.00	1.00	1
31	1.00	1.00	1.00	1
32	1.00	1.00	1.00	1
33	1.00	1.00	1.00	1
34	1.00	1.00	1.00	1
35	1.00	1.00	1.00	1
36	1.00	1.00	1.00	1
37	1.00	1.00	1.00	1
38	0.00	0.00	0.00	1
39	0.00	0.00	0.00	1
40	0.00	0.00	0.00	1
accuracy			0.93	40
macro avg	0.89	0.93	0.90	40
weighted avg	0.89	0.93	0.90	40

```
************ Overall Performance ************
CNN ORL Accuracy: 0.925 +/- None
************ Finish ************
```

3.7 Ablation Studies

MNIST

Trial A: Instead of a kernel size of 5 in the convolutional layer, the kernel size is 3. All other hyperparameters are the same.

```
evaluating performance...
              precision    recall  f1-score   support

           0       0.98      0.98      0.98       980
           1       0.98      0.99      0.99      1135
           2       0.97      0.97      0.97      1032
           3       0.99      0.91      0.95      1010
           4       0.97      0.97      0.97       982
           5       0.91      0.98      0.95       892
           6       0.97      0.98      0.98       958
           7       0.94      0.97      0.95      1028
           8       0.97      0.97      0.97       974
           9       0.98      0.93      0.95      1009

    accuracy                           0.97     10000
   macro avg       0.97      0.97      0.97     10000
weighted avg       0.97      0.97      0.97     10000

************ Overall Performance ************
CNN MNIST Accuracy: 0.9667 +/- None
```

Trial B: Instead of a kernel size of 2 in the convolutional layer, the kernel size is 3. All other hyperparameters are the same.

```
evaluating performance...
              precision    recall  f1-score   support

           0       0.98      0.96      0.97       980
           1       0.98      0.98      0.98      1135
           2       0.97      0.94      0.95      1032
           3       1.00      0.84      0.91      1010
           4       0.95      0.97      0.96       982
           5       0.83      0.99      0.90       892
           6       0.99      0.92      0.95       958
           7       0.97      0.95      0.96      1028
           8       0.83      0.98      0.90       974
           9       0.97      0.91      0.94      1009

    accuracy                           0.94     10000
   macro avg       0.95      0.94      0.94     10000
weighted avg       0.95      0.94      0.94     10000

************ Overall Performance ************
CNN MNIST Accuracy: 0.9442 +/- None
```

CIFAR-10

Trial A: Images not normalized prior to training. 2 fully connected layers used. Batch normalization used after every convolutional layer. Higher dropout rates. 10 epochs.

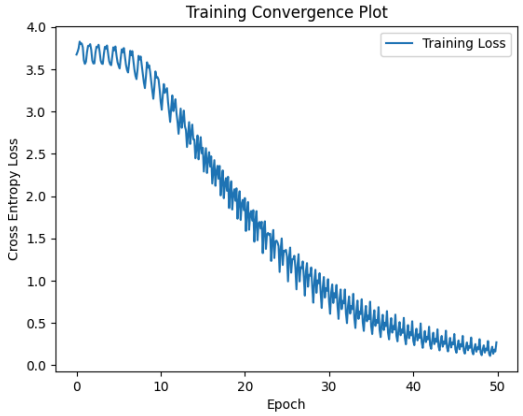
```
              precision    recall  f1-score   support

           0       0.55      0.61      0.58      1000
           1       0.60      0.63      0.61      1000
           2       0.42      0.34      0.38      1000
           3       0.30      0.38      0.33      1000
           4       0.49      0.41      0.45      1000
           5       0.42      0.43      0.42      1000
           6       0.60      0.56      0.58      1000
           7       0.60      0.57      0.59      1000
           8       0.68      0.65      0.66      1000
           9       0.55      0.59      0.57      1000

    accuracy                           0.52     10000
   macro avg       0.52      0.52      0.52     10000
weighted avg       0.52      0.52      0.52     10000

************ Overall Performance ************
CNN CIFAR Accuracy: 0.5167 +/- None
************ Finish ************
```

Trial B: Images not normalized prior to training. 5 convolutional layers used to condense the image down to 1x1x512. 2 fully connected layers and softmax used. Batch normalization used. No dropout. 10 epochs.

```
              precision    recall  f1-score   support

           0       0.52      0.47      0.49      1000
           1       0.54      0.50      0.52      1000
           2       0.39      0.33      0.36      1000
           3       0.31      0.27      0.29      1000
           4       0.39      0.32      0.35      1000
           5       0.41      0.39      0.40      1000
           6       0.48      0.53      0.50      1000
           7       0.46      0.58      0.51      1000
           8       0.50      0.64      0.56      1000
           9       0.48      0.52      0.50      1000

    accuracy                           0.45     10000
   macro avg       0.45      0.45      0.45     10000
weighted avg       0.45      0.45      0.45     10000

************ Overall Performance ************
CNN CIFAR Accuracy: 0.4529 +/- None
************ Finish ************
```

ORL

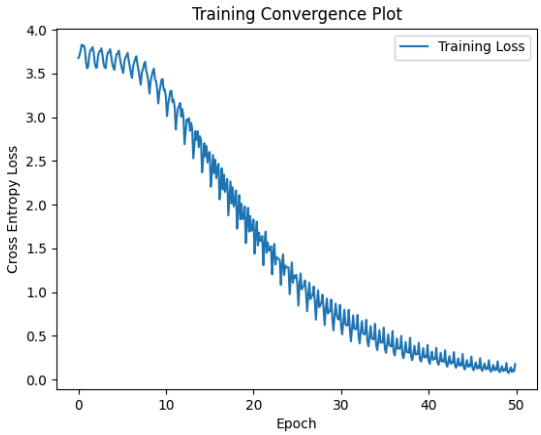
Trial A: The first linear layer replaces Tanh with Softmax. The second linear layer replaced LeakyReLU with ReLU. All other hyperparameters are the same.



	precision	recall	f1-score	support
1	0.00	0.00	0.00	1
2	1.00	1.00	1.00	1
3	1.00	1.00	1.00	1
4	1.00	1.00	1.00	1
5	0.50	1.00	0.67	1
6	1.00	1.00	1.00	1
7	1.00	1.00	1.00	1
8	1.00	1.00	1.00	1
9	0.50	1.00	0.67	1
10	1.00	1.00	1.00	1
11	1.00	1.00	1.00	1
12	1.00	1.00	1.00	1
13	1.00	1.00	1.00	1
14	1.00	1.00	1.00	1
15	1.00	1.00	1.00	1
16	0.50	1.00	0.67	1
17	1.00	1.00	1.00	1
18	1.00	1.00	1.00	1
19	1.00	1.00	1.00	1
20	1.00	1.00	1.00	1
21	1.00	1.00	1.00	1
22	0.50	1.00	0.67	1
23	1.00	1.00	1.00	1
24	1.00	1.00	1.00	1
25	1.00	1.00	1.00	1
26	0.50	1.00	0.67	1
27	1.00	1.00	1.00	1
28	1.00	1.00	1.00	1
29	1.00	1.00	1.00	1
30	1.00	1.00	1.00	1
31	1.00	1.00	1.00	1
32	1.00	1.00	1.00	1
33	1.00	1.00	1.00	1
34	1.00	1.00	1.00	1
35	1.00	1.00	1.00	1
36	1.00	1.00	1.00	1
37	0.00	0.00	0.00	1
38	0.00	0.00	0.00	1
39	0.00	0.00	0.00	1
40	0.00	0.00	0.00	1
accuracy			0.88	40
macro avg	0.81	0.88	0.83	40
weighted avg	0.81	0.88	0.83	40

********** Overall Performance **********
CNN ORL Accuracy: 0.875 +/- None
********** Finish **********

Trial B: Instead of Tanh being applied to the first linear layer, Sigmoid applied. All other hyperparameters are the same.



	precision	recall	f1-score	support
1	1.00	1.00	1.00	1
2	1.00	1.00	1.00	1
3	1.00	1.00	1.00	1
4	1.00	1.00	1.00	1
5	0.50	1.00	0.67	1
6	1.00	1.00	1.00	1
7	1.00	1.00	1.00	1
8	1.00	1.00	1.00	1
9	1.00	1.00	1.00	1
10	1.00	1.00	1.00	1
11	1.00	1.00	1.00	1
12	1.00	1.00	1.00	1
13	1.00	1.00	1.00	1
14	1.00	1.00	1.00	1
15	1.00	1.00	1.00	1
16	1.00	1.00	1.00	1
17	1.00	1.00	1.00	1
18	1.00	1.00	1.00	1
19	1.00	1.00	1.00	1
20	1.00	1.00	1.00	1
21	1.00	1.00	1.00	1
22	1.00	1.00	1.00	1
23	0.50	1.00	0.67	1
24	1.00	1.00	1.00	1
25	1.00	1.00	1.00	1
26	1.00	1.00	1.00	1
27	1.00	1.00	1.00	1
28	1.00	1.00	1.00	1
29	0.50	1.00	0.67	1
30	1.00	1.00	1.00	1
31	1.00	1.00	1.00	1
32	1.00	1.00	1.00	1
33	1.00	1.00	1.00	1
34	1.00	1.00	1.00	1
35	1.00	1.00	1.00	1
36	1.00	1.00	1.00	1
37	1.00	1.00	1.00	1
38	0.00	0.00	0.00	1
39	0.00	0.00	0.00	1
40	0.00	0.00	0.00	1
accuracy			0.93	40
macro avg	0.89	0.93	0.90	40
weighted avg	0.89	0.93	0.90	40

********** Overall Performance **********
CNN ORL Accuracy: 0.925 +/- None
********** Finish **********