

# **THE STREET VIEW HOUSE NUMBERS (SVHN) DATASET PROJECT**

- [Dataset](#)
- [Code](#)
- [Papers](#)
- [Contact](#)
- [About](#)
- [FAQ](#)
- [Citation](#)
- [Attribution](#)
- [License](#)
- [Feedback](#)



# GOOGLE STREET VIEW IMAGES DATASET

SVHN is obtained from house numbers in Google Street View images

It has over 600,000 digit images and the numbers are in the natural scene images.

- The dataset has ten classes, one for each digit. Digit '0' has label 0, '9' has label 9.
- In the dataset, they separate data into two groups: 73,257 digits for training and 26,032 digits for testing. There also have 531,131 additional data to use as extra training data.
- In this project, we are going to use the 32-by-32 color images that all images are resized to the fixed resolutions.



# OBJECTIVE

We aimed at applying various deep learning techniques in order to do the image classification.

The models:

- Multilayer Perceptron Networks (MLP)
- Convolutional Neural Network (CNN) by Pytorch
- CNN by Caffe.

Though these different models, we can also find out how these algorithms compare to each other and how can we improve our model.



# MULTILAYER PERCEPTRON NETWORKS (MLP)

MLP network has two linear layers and one ReLU layer.

- has input size as  $3*32*32$  and output size of this layer as `hidden_size`(=500).
- has input size as `hidden_size` and output size as the number of classes(=10).
- Mini batch = 100, 200, 500, 1000.
  - we also find out the best result comes out when mini batch = 500
- Learning rate = 0.001

	<b>MLP</b>
<b>Accuracy</b>	<b>0.80</b>
<b>Time (s)</b>	<b>130.97</b>
<b>F1 score</b>	<b>0.79</b>



# RESULT FOR MLP

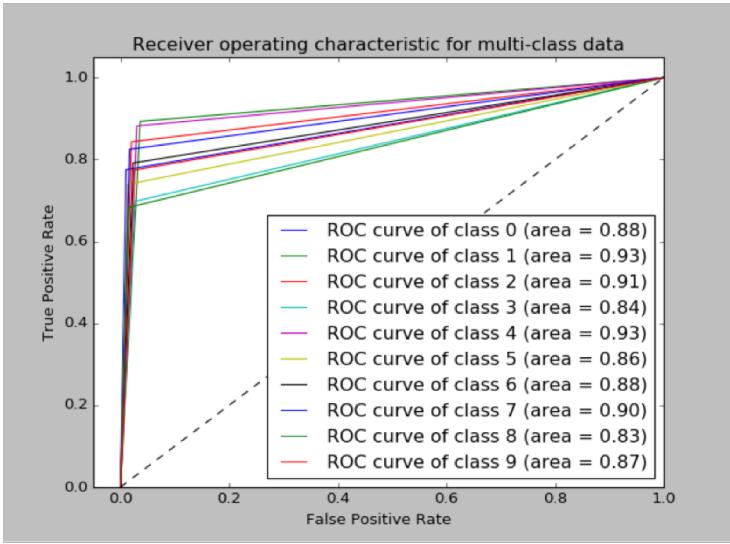


Figure :The ROC of MLP

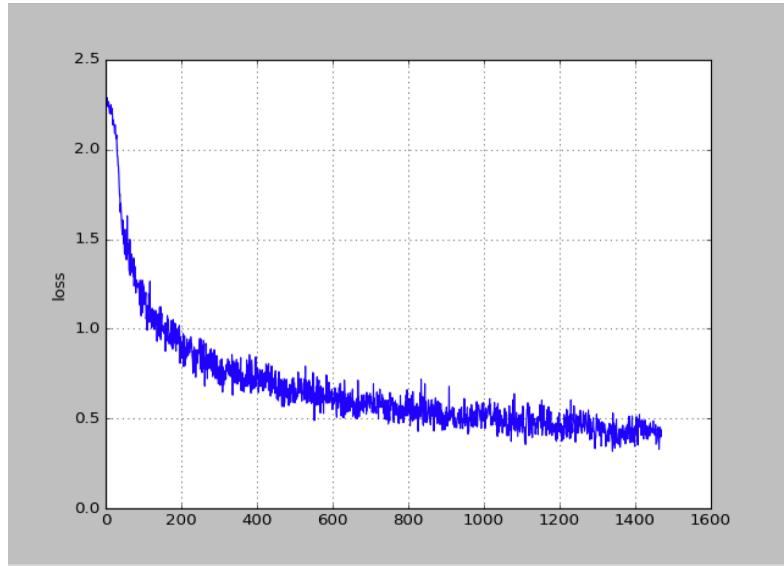


Figure: The loss of MLP

The loss of the MLP model maintains between 0.4 to 0.6. The highest single class accuracy can reach 89% (class 2), and the lowest single accuracy is around 68% (class 3). This simple MLP model can reach the 80% overall accuracy with 10 epochs training process, and the result doesn't change a lot when we adding more training iterations, try different batch sizes or the number of hidden sizes.



# ACCURACY & F1-SCORE

Accuracy of	0 : 77 %
Accuracy of	1 : 89 %
Accuracy of	2 : 84 %
Accuracy of	3 : 69 %
Accuracy of	4 : 88 %
Accuracy of	5 : 73 %
Accuracy of	6 : 79 %
Accuracy of	7 : 82 %
Accuracy of	8 : 68 %
Accuracy of	9 : 77 %

Classification report:					
	precision	recall	f1-score	support	
0	0.84	0.78	0.81	1744	
1	0.86	0.89	0.87	5099	
2	0.89	0.84	0.87	4149	
3	0.77	0.70	0.73	2882	
4	0.76	0.88	0.82	2523	
5	0.85	0.74	0.79	2384	
6	0.74	0.79	0.76	1977	
7	0.81	0.83	0.82	2019	
8	0.76	0.68	0.72	1660	
9	0.67	0.77	0.72	1595	
avg / total	0.81	0.81	0.81	26032	
Precision: 0.793					
Recall: 0.791					
F1: 0.790					



# CONVOLUTIONAL NEURAL NETWORK(CNN)

## CONVOLUTIONAL NEURAL NETWORK

MODEL 1:  
CONV1 + MAXPOOL  
CONV2 +MAXPOOL  
F1 SCORE : 0.844  
ACUARRY : 85%

MODEL 2:  
CONV1 + MAXPOOL+RELU  
CONV2 +MAXPOOL+RELU  
F1 SCORE : 0.87  
ACUARRY : 88%

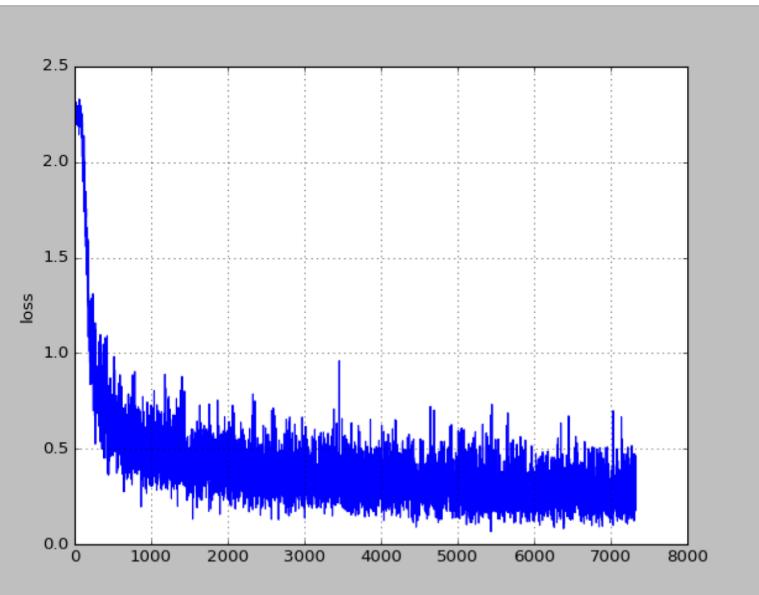
MODEL 3:  
CONV1 + MAXPOOL+RELU+DROPOUT(0.2)  
CONV2 +MAXPOOL+RELU+DROPOUT(0.2)  
F1 SCORE : 0.872  
ACUARRY : 88%

MODEL 4:  
CONV1 + MAXPOOL+RELU+DROPOUT(0.2)  
CONV2 +MAXPOOL+RELU+DROPOUT(0.2)  
FC1(RELU)  
FC2(RELU)  
FC3  
F1 SCORE : 0.9  
ACUARRY : 91%

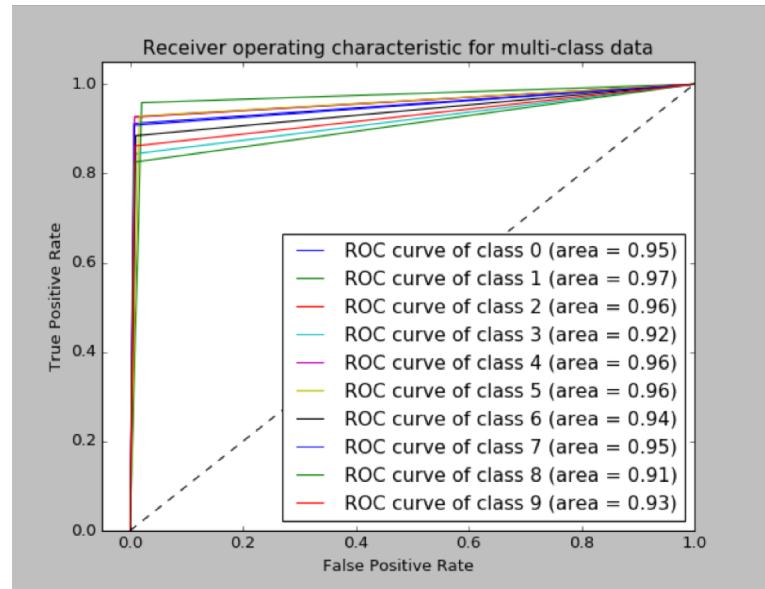
MODEL 5:  
CONV1 + MAXPOOL+RELU+DROPOUT(0.2)  
CONV2 +MAXPOOL+RELU+DROPOUT(0.2)  
FC1(RELU)  
FC2(TANH)  
FC3  
F1 SCORE : 0.875  
ACUARRY : 88%

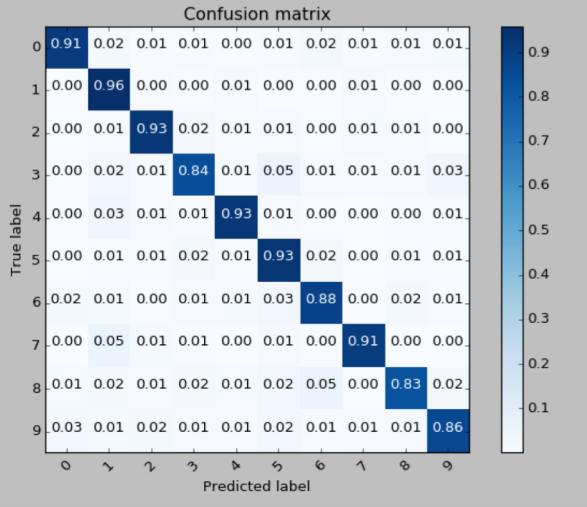


# RESULT FOR CNN



	CNN (ver1)	CNN (ver2)	CNN (ver3)	CNN (ver4)	CNN (ver5)
Accuracy	0.85	0.88	0.88	0.91	0.88
Time (s)	142.60	141.79	144.45	141.80	136.55
F1 score	0.844	0.87	0.872	0.90	0.875





The Normalized Confusion Matrix Plot of CNN Model 4

Confusion matrix, without normalization

```
[ [1590  31   11   10    3   14   34   13   16   22]
  [ 23 4887  24   20   47   16   14   48   15   5]
  [  9  47 3842  67   30   39   13   61   29   12]
  [ 10   71   30 2433  16  148   26   23   33   92]
  [  7   80   26   18 2339  13    6    9   12   13]
  [  7   16   13   40   15 2214   47    2   14   16]
  [ 39   28    4   17   27   55 1749    9   37   12]
  [  3  107   25   15    4   18    5 1833    3    6]
  [ 21   28   15   33   24   41   89    4 1370   35]
  [ 47   23   35   18   14   38   11   17   18 1374] ]
```

Confusion Matrix of CNN Model 4

Accuracy of 0	:	91 %
Accuracy of 1	:	95 %
Accuracy of 2	:	92 %
Accuracy of 3	:	84 %
Accuracy of 4	:	92 %
Accuracy of 5	:	92 %
Accuracy of 6	:	88 %
Accuracy of 7	:	90 %
Accuracy of 8	:	82 %
Accuracy of 9	:	86 %

	Classification report:			
	precision	recall	f1-score	support
0	0.91	0.91	0.91	1744
1	0.92	0.96	0.94	5099
2	0.95	0.93	0.94	4149
3	0.91	0.84	0.88	2882
4	0.93	0.93	0.93	2523
5	0.85	0.93	0.89	2384
6	0.88	0.88	0.88	1977
7	0.91	0.91	0.91	2019
8	0.89	0.83	0.85	1660
9	0.87	0.86	0.86	1595
avg / total	0.91	0.91	0.91	26032

Precision: 0.901

Recall: 0.898

F1: 0.899



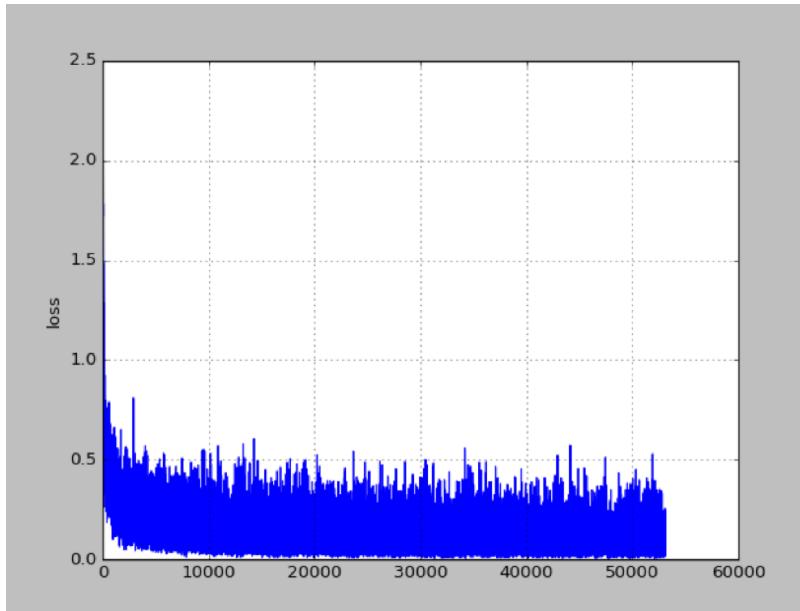
# MORE IMPLEMENTS

- Used 531,131 additional training data in order to see there have improvements for our models.
  - Extra training dataset has more clear background.

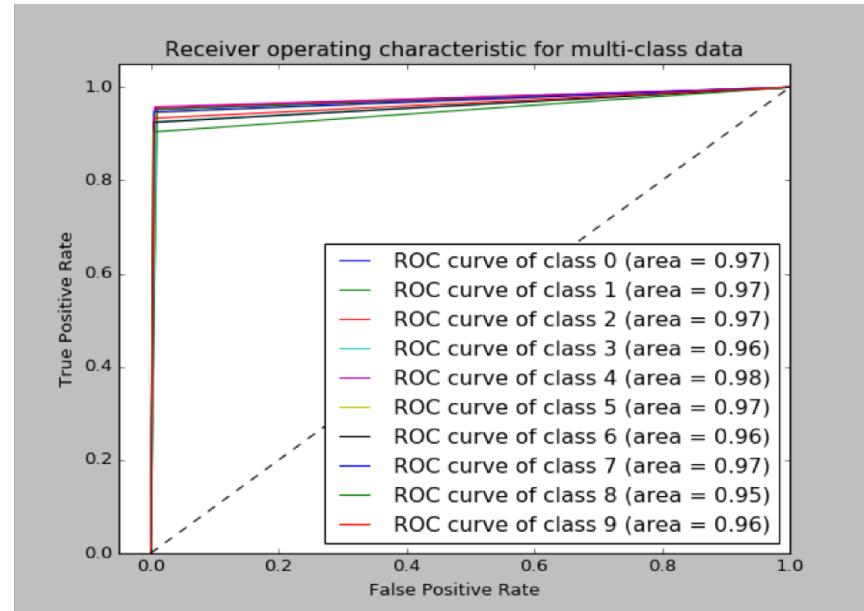
	<b>Training dataset</b>	<b>Extra Training dataset</b>
<b>Accuracy</b>	<b>0.91</b>	<b>0.94</b>
<b>Time (s)</b>	<b>141.80</b>	<b>1048.43</b>
<b>F1 score</b>	<b>0.90</b>	<b>0.939</b>



# CNN MODEL 4 BY USING EXTRA TRAINING DATA



The Loss of CNN Model 4



The ROC of CNN Model 4



# CONCLUSION & FUTURE RESEARCH

- **We try CNN model 2 by using Caffe, the time is much shorter than CNN model. (80 vs 140 seconds)**
  - With training dataset,
    - MLP has the lowest accuracy of 80% and f1 score of 0.79
    - CNN model 4 has the highest accuracy of 91% and f1 score of 0.9
  - With the same CNN model 4:
    - the performance on the test dataset with training dataset(~ 70k) has the accuracy of 91%
    - the performance on the test dataset with the extra training dataset(~ 531k) has the accuracy of 94%.
- **The results above show that the CNN has higher accuracy and lower loss than the MLP model.**

## Future Research

- Add more layer:
  - Someone add CNN model to 20 layers and get about 93% accuracy:
  - The link: <https://github.com/hangyao/street view house numbers>
- Changing the different Pooling layers:
  - Another research shows CNN with mix-pooling(50 average-pooling /50 max-pooling) can increase the accuracy.



**THANK YOU**

