Traffic sign classifier Project.

Build a Traffic Sign Recognition Project

The goals / steps of this project are the following:

- Load the data set (see below for links to the project data set)
- Explore, summarize and visualize the data set
- Design, train and test a model architecture
- Use the model to make predictions on new images
- Analyze the softmax probabilities of the new images
- Summarize the results with a written report

You're reading it! and here is a link to my project code.

Data Set Summary & Exploration

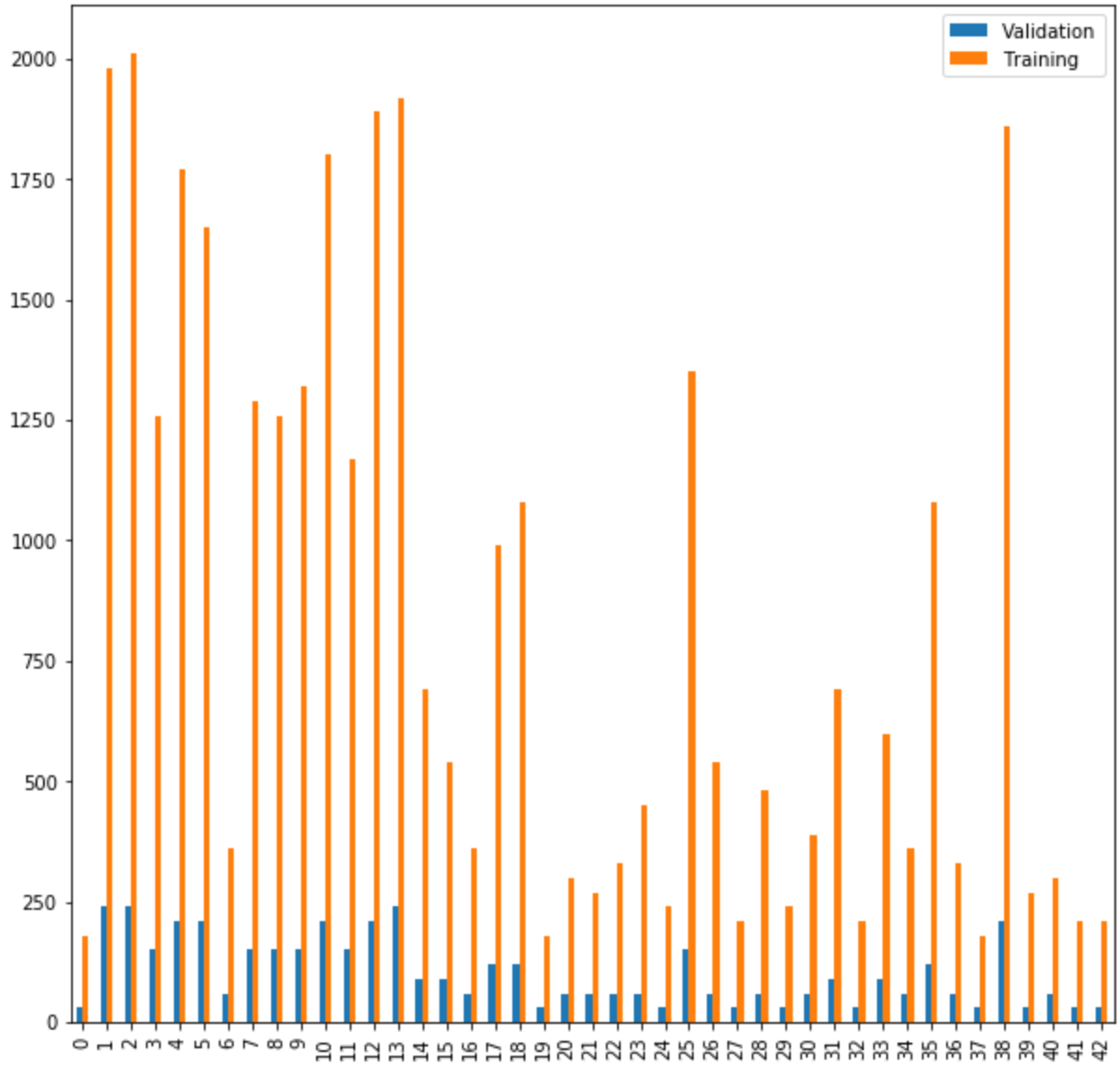
1. Basic summary of the data set

I used the pandas library to calculate summary statistics of the traffic signs data set:

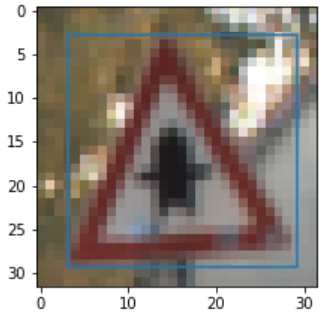
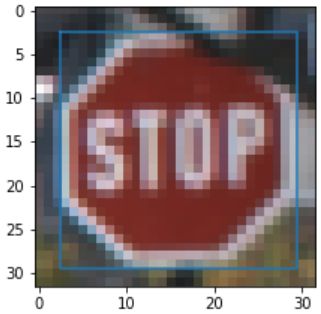
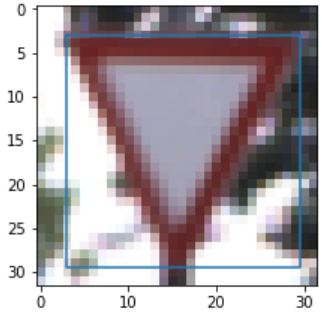
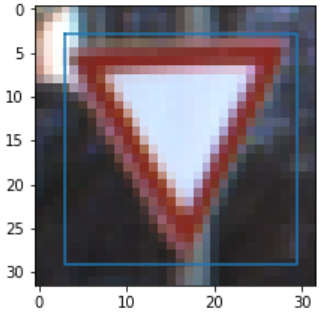
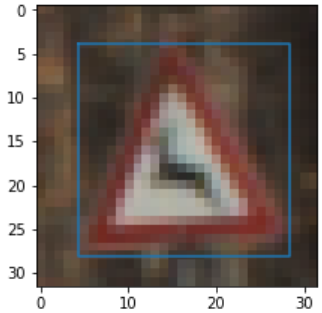
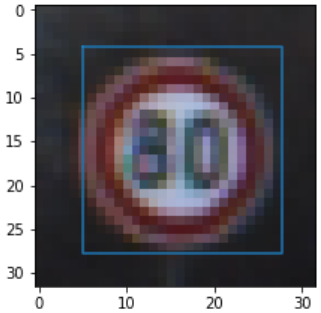
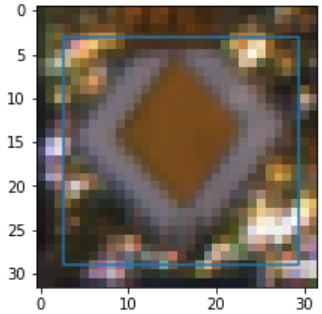
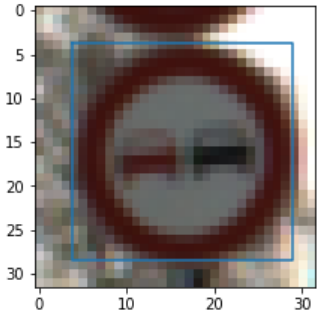
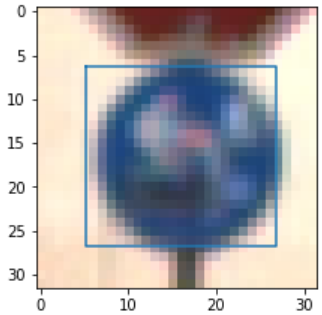
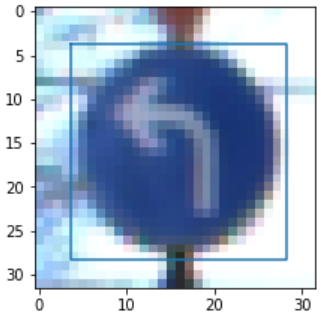
- The size of training set is ?
  - 34799
- The size of the validation set is ?
  - 4410
- The size of test set is ?
  - 12630
- The shape of a traffic sign image is ? Example image 200
  - Size = (53, 55]) , image location is (6,6) - (48,50)
- The number of unique classes/labels in the data set is ?
  - 43

2. Exploratory visualization of the dataset.

This figure show the number of images in the training and validation set per classification.



This Set of images is a random sampling of the training images. I overlayed a bounding box on the area of the image that contains the sign.



STOP




30





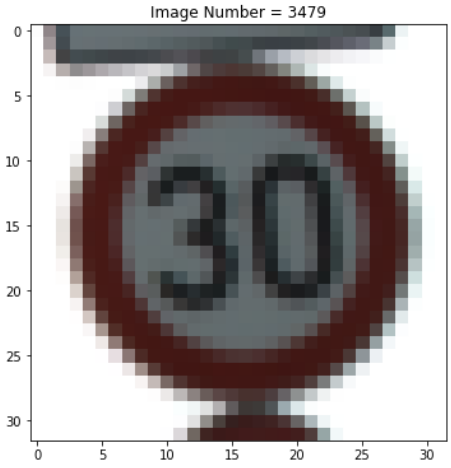
Design and Test a Model Architecture

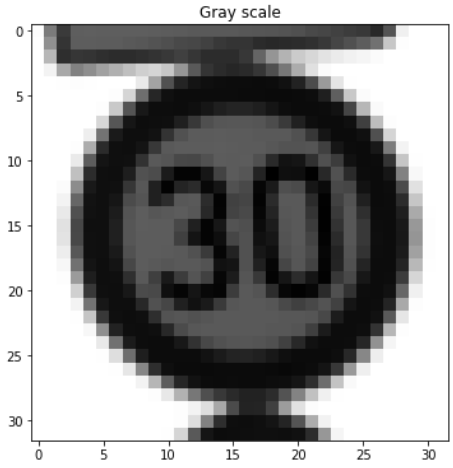
Preprocessing

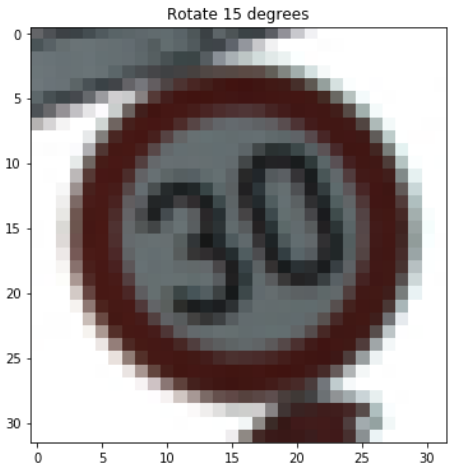
I used a simple normalization procedure as suggested in the notebook, I just converted from the [0-255,0-255,0-255] images to [-1.0-1.0,-1.0-1.0,-1.0-1.0] using the fast val - 128.0 / 128.0

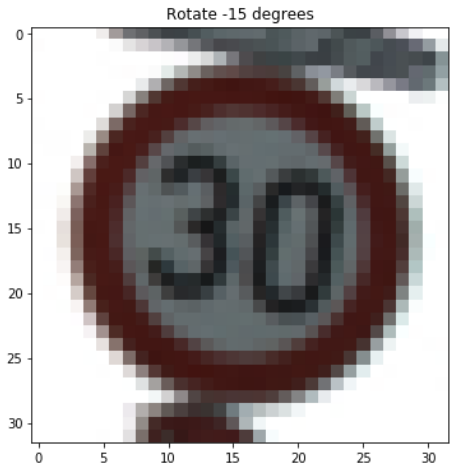
I experimented with gray scaling the images, and I also extended the training image dataset by adding rotations of the image +- 15 degrees. I read this in paper , they also used other techniques to extend the

Here is an example set of changes to a single image (3479) during the preprocessing.









Model architecture.

My final model consisted of the following layers:

Layer 1: Convolutional. Input = 32x32x3. Output = 28x28x9.
Relu Activation.
Pooling. Input = 28x28x9. Output = 14x14x6.

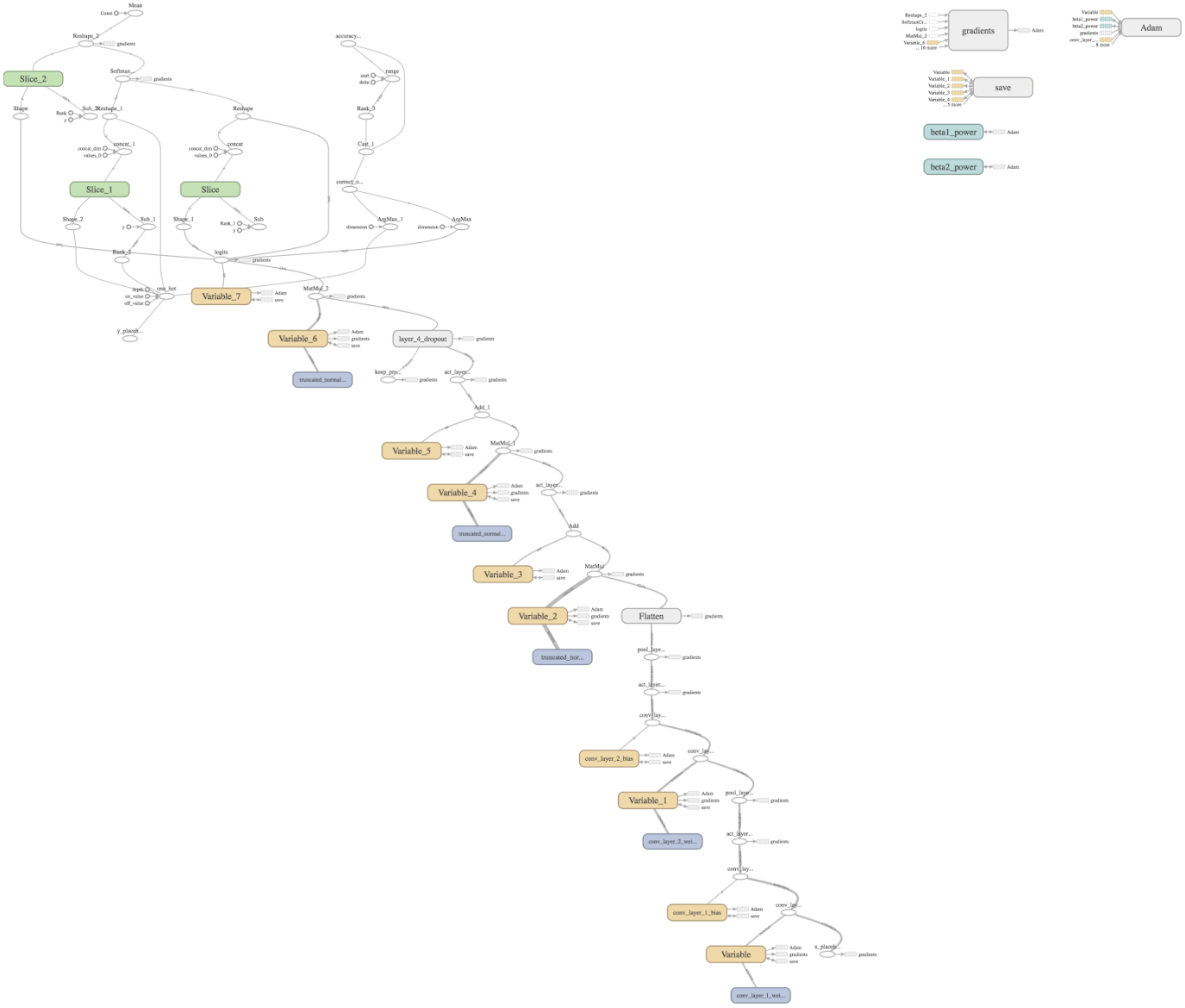
Layer 2: Convolutional. Output = 10x10x21.
Relu Activation.
Pooling. Input = 10x10x21. Output = 5x5x21.

Flatten. Input = 5x5x16. Output = 525.
Layer 3: Fully Connected. Input = 525. Output = 150.
Relu Activation.

Layer 4: Fully Connected. Input = 150. Output = 84.
Relu Activation.
Dropout of 50% connections between Layer 4 and Layer 5

Layer 5: Fully Connected. Input = 84. Output = 43.

Below is an diagram of the CNN model.
using TensorBoard (TensorFlow's built in graph viewer).



3 .Training the model

I started with the same set of hyper parameters as used in the LeNet sample.
Using the AdamOptimizer for the learning rate decay.

I didn't have time to experiment much beyond extending the number of Epochs.

Final hyper parameters chosen

Initial learning rate of 0.001
Drop out of .5
Batch size 128
Number of epochs 20
Initial Weights (mu = 0, sigma = 0.1)

4. Approach

Initially I tried a version of the LeNet model but expanded the first convolution layer to take 32x32x3 as the input. I left all other layers the same until the output layer that needed to be changed to having 43 outputs - one for each image classification type.

With this I achieved approx .90 accuracy.

Next I extended the layers to add more connections. Thinking that classifying into 43 buckets might require the network to have more fine grained features. I increased the depth of the convolution layers and then widened the fully connected layers.

Next I tried playing with coordinates of the actual sign with in the image. First I blanked out the borders with 0s as I didn't think that data was relevant. That actually reduced the accuracy significantly. Then I thought that perhaps if I randomized the borders it would help the model to focus on the more important parts of the image. However this didn't lead to an improvement either, and just slowed the training process. I then tried grayscale image pre processing. As a final step I extended the training dataset by adding a copy of images rotated 15degrees ,and a copy rotated -15 degrees. This help boost the accuracy a few more %points. I also experimented between grayscale and color with rotation. There wasn't much in the difference in terms of final performance ,so I ended up going with the color based model.
In hindsight I think grayscale might have been better , just for the preprocessing performance boost if nothing else.

Training set Accuracy = 0.999
Validation set Accuracy = 0.952
Test set Accuracy = 0.946

My final model results were:

- training set accuracy of ?
    - 99.9%
- validation set accuracy of ?
    - 95.2%
- test set accuracy of ?
    - 94.6%

Iterative Approach

- What was the first architecture that was tried and why was it chosen?
    - Chose the default LeNet architecture modified for 3 channels of color on in the input and 43 outputs in the final layer..

- What were some problems with the initial architecture?
    - Using the same hyper parameters from LeNet achieved approx 90% accuracy on the validation set.
- How was the architecture adjusted and why was it adjusted?
    - First I tried to expand the number of connections (making the mode wider). My intuition was that given we have 43 possible outcomes up from 10 in the original LeNet architecture more connections would allow the model to better differentiate the images.
    - Secondly I added a drop out layer.
    - The combination of these 2 improvements in the architecture boosted the validation performance to approx 95%
- Which parameters were tuned?
    - I trained the model locally on a CPU so time constraints limited the amount of tuning. I researched potentially using Google Cloud Engine's Hyptertuning to help do this, but I didn't have time to experiment , you have to package your tensor flow model in a specific way to do this.
    - I did some basic tuning with the number of Epochs but ended up just using 10.

- How does the final model's accuracy on the training, validation and test set provide evidence that the model is working well?

Comparison of validation results per epoch for 5 different approaches.

Epoch	Basic LeNet with Color input Accuracy	Wider LeNet with Color	Wider LeNet gray scale	Wider LeNet gray scale Rotation +- 15degrees	Wider LeNet color Rotation +- 15degrees
1	0.713	0.789	0.789	0.877	0.88
2	0.792	0.876	0.876	0.914	0.924
3	0.826	0.905	0.905	0.934	0.936
4	0.879	0.922	0.922	0.934	0.943
5	0.881	0.936	0.936	0.938	0.939
6	0.863	0.95	0.95	0.943	0.939
7	0.887	0.941	0.941	0.956	**0.95**
8	0.899	0.944	0.944	0.953	0.944
9	0.884	**0.95**	**0.95**	**0.96**	0.953
10	**0.908**	0.939	0.939	0.958	0.952

Test a Model on New Images

I choose a random sampling of images, which I cropped to be similar to the input images in terms of being mostly the sign, then I ran them through the model.

The images , their predictions and softmax probabilities are all displayed in the image below

[1|Speed limit (30km/h)=99.992%]
[0|Speed limit (20km/h)=0.008%]
[4|Speed limit (70km/h)=0.0%]
[16|Vehicles over 3.5 metric tons prohibited=0.0%]
[9|No passing=0.0%]

[28|Children crossing=100.0%]
[30|Beware of ice/snow=0.0%]
[20|Dangerous curve to the right=0.0%]
[23|Slippery road=0.0%]
[11|Right-of-way at the next intersection=0.0%]

[14|Stop=99.808%]
[17|No entry=0.192%]
[10|No passing for vehicles over 3.5 metric tons=0.0%]
[13|Yield=0.0%]
[29|Bicycles crossing=0.0%]

[33|Turn right ahead=99.999%]
[35|Ahead only=0.001%]
[34|Turn left ahead=0.0%]
[37|Go straight or left=0.0%]
[5|Speed limit (80km/h)=0.0%]

[23|Slippery road=100.0%]
[21|Double curve=0.0%]
[30|Beware of ice/snow=0.0%]
[28|Children crossing=0.0%]
[19|Dangerous curve to the left=0.0%]

[5|Speed limit (80km/h)=54.171%]
[25|Road work=24.316%]
[1|Speed limit (30km/h)=9.355%]
[18|General caution=5.496%]
[24|Road narrows on the right=4.258%]

[36|Go straight or right=100.0%]
[38|Keep right=0.0%]
[41|End of no passing=0.0%]
[40|Roundabout mandatory=0.0%]
[34|Turn left ahead=0.0%]

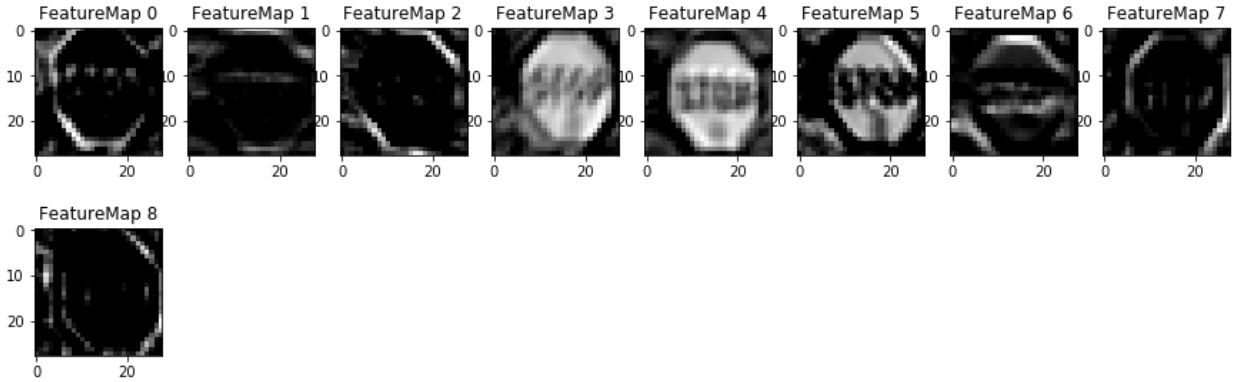
[14|Stop=100.0%]
[5|Speed limit (80km/h)=0.0%]
[17|No entry=0.0%]
[13|Yield=0.0%]
[29|Bicycles crossing=0.0%]

The model performed at 71% over the 7 images.
It's somewhat understandable that the 6th image - which is partially obscured with ice/snow fails.
The The model did not do a good job with image 3 - it is not obscured , but it gave it a 99% probability of being a stop sign, and a 0.2% probability of being the correct image.

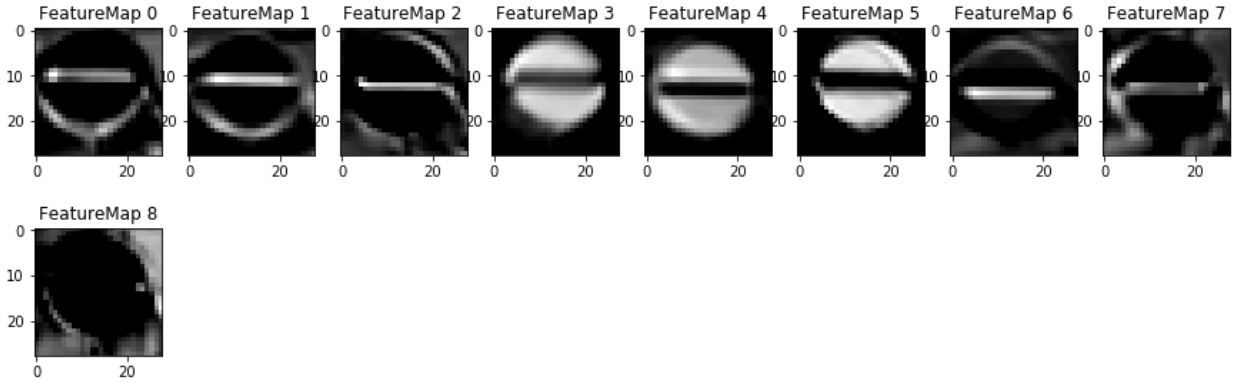
It's interesting to dig down into the feature maps produced by the convolution layers for the 2 images.

First layer

These are for the Stop sign which is correctly classified.

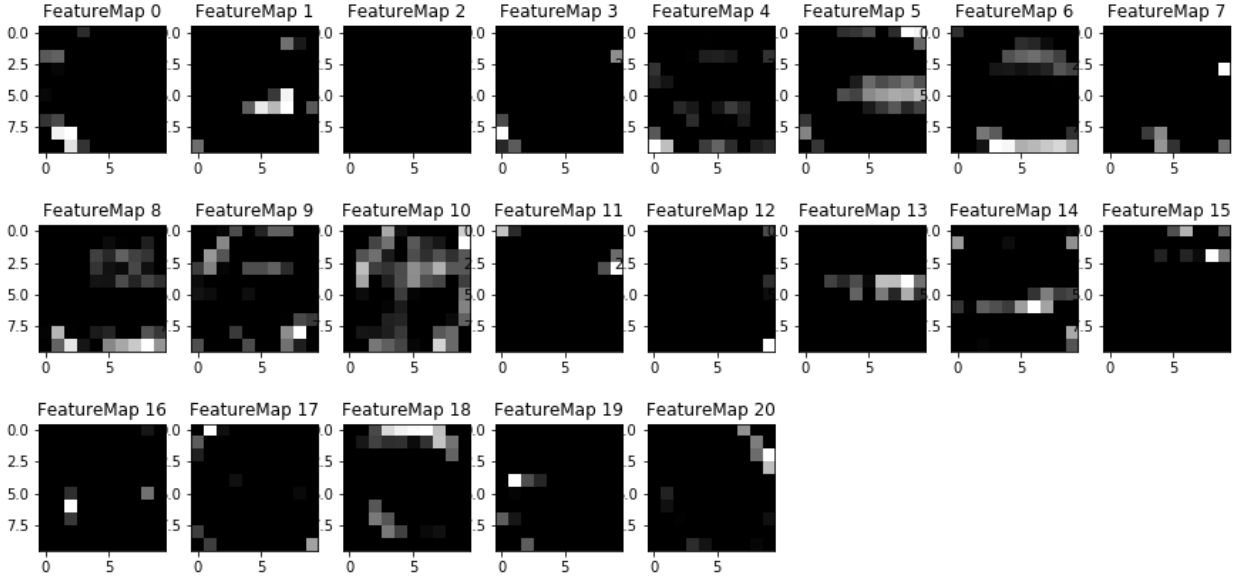


These are for the Do Not Enter sign that is incorrectly classified.

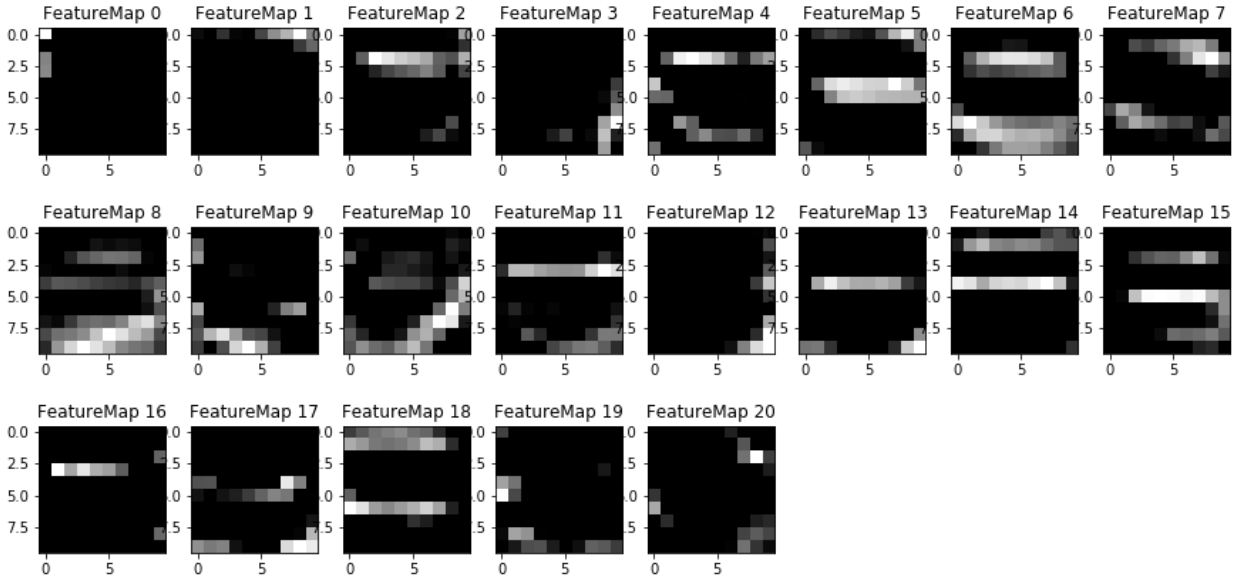


Second layer

Stop Sign



Do not enter sign.



My theory is that the convolution is minimizing the difference between the STOP and ----- in the 2 input images.

It’s possible that modifying the network further might be able to a better job, but I’m not sure exactly how. This is an area I would be interested in exploring further in the future.