NEW YORK UNIVERSITY

CV Assignment 2

A Report

on

The German Traffic Sign Benchmark

Report By:
Vikas Patidar
vp1274

Master of Science in Computer Science

Introduction

The German Traffic Sign Benchmark (GTSB) is a multi-class image classification challenge. The goal of GTSB is to get researchers in the field of Computer Vision to a common domain, where they could create models to recognize traffic signs and test it on a common dataset.

Initial Experiments:

- Used AlexNet architecture and trained the model to achieve an accuracy of 93.99 on the test dataset.
- Used VGG-11 architecture and trained the model to achieve an accuracy of 95.122 on the test dataset.

Main Approach:

The most prominent part of my final architecture is the Spatial Transformer Network (STN) [1], which was first introduced by Google DeepMind.

- Experiment 1
  Used two STN units with one batch normalization layer. Trained the model for 17 epochs and achieved an accuracy of 98.939.
- Experiment 2
  Increased the number of STN units to 3 and increased the batch normalization layers. Trained the model for 20 epochs and achieved an accuracy of **99.255(Public leaderboard - Cr7).**

Architecture

I implemented the architecture similar to the architecture implemented in “Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods”. [2]

This architecture has three conv and max-pool layers with Relu activations, three fully connected layers and three STN units. Each of the STN unit is placed before a conv layer. Instead of the linear contrast normalization layers used in the paper “Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods”, my architecture uses batch normalization layers.

STN Unit

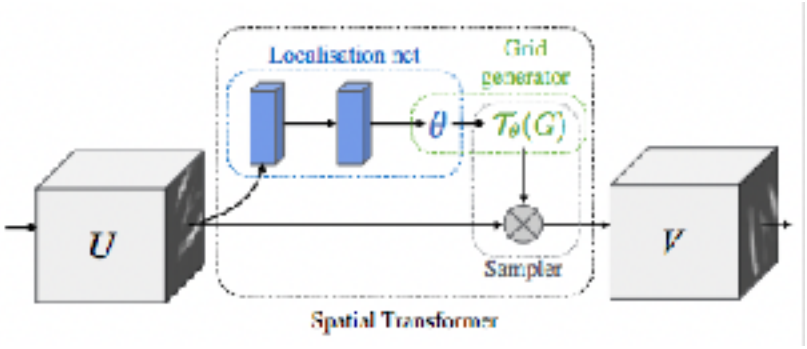

Fig 2: One STN Unit

Detailed Architecture

STN 1			
BatchNorm 1			
Conv 1	Kernels: 7X7	Padding: 2	Output Channels: 200
Max Pool 1	Stride: 2		
Relu			
STN 2			
BatchNorm 2			
Conv 2	Kernels: 4X4	Padding: 2	Output Channels: 250
Max Pool 2	Stride: 2		
Relu			
STN 3			
BatchNorm 3			
Conv 3	Kernels: 4X4	Padding: 2	Output Channels: 350
Max Pool 3	Stride: 2		
Relu			
Linear	400 Units		

Relu	
Linear - Softmax	43 Units

Hyperparameter Tuning & Results

- Tested different batch sizes like 8, 16 and 64. A batch size of 64 gave the best test accuracy.
- Learning rate of 0.0001 worked well out of 0.1, 0.001 and 0.0001.
- Momentum of 0.9 worked better than 0.5.
- Used Leaky ReLU instead of ReLU in between STN units and it improved accuracy by 0.2%.

Graphs of train loss vs epoch and validation loss vs epoch are shown below.

After 20th epoch,

Train Loss:	0.00002
Validation Loss:	0.0507
Validation Accuracy:	98%
Test Accuracy:	99.255%

Train Loss vs Epoch Graph

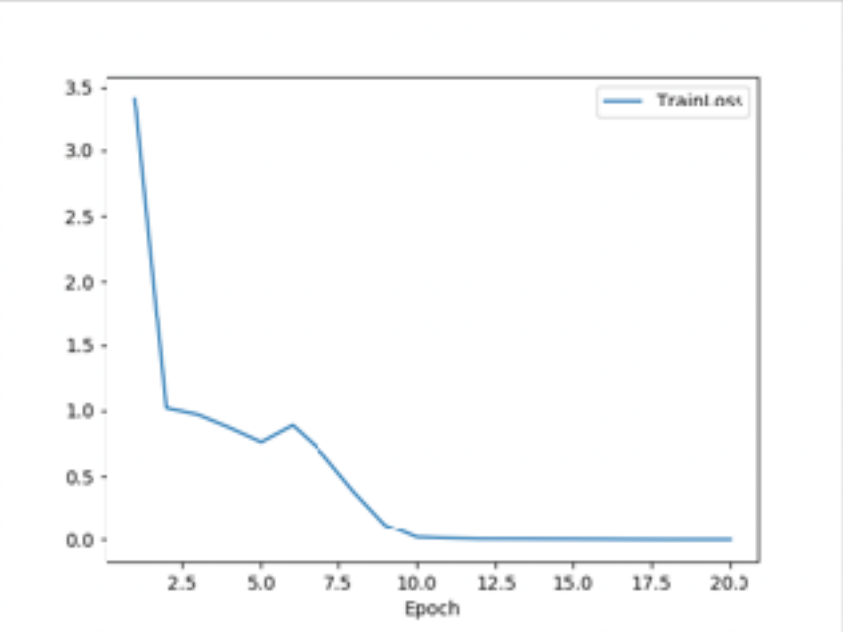

Fig 3: Train Loss vs Epoch

Validation Loss vs Epoch Graph

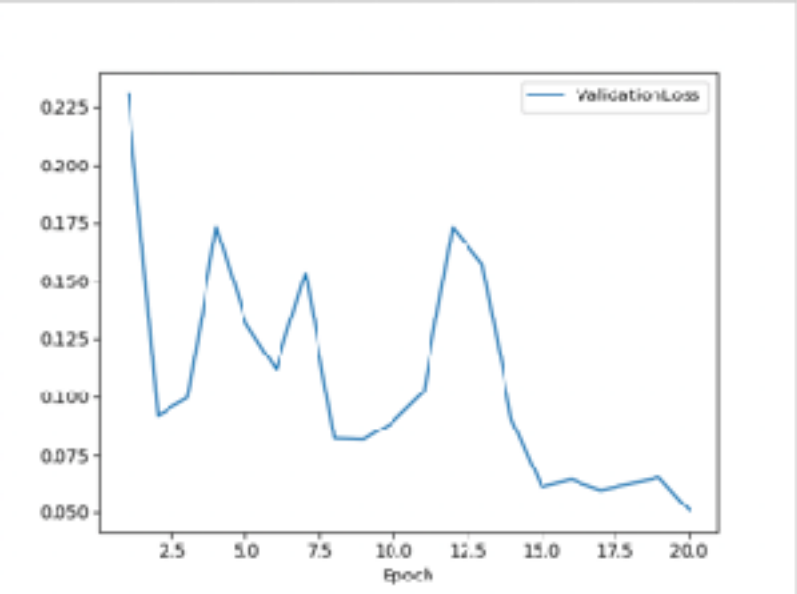

Fig 4: Validation Loss vs Epoch

References

[1] M. S. K. a. Z. A. Jaderberg, "Spatial transformer networks. In Advances in neural information processing systems," *Advances in neural information processing systems,* pp. 2017-2025, 2015.

[2] Á. Á.-G. J. a. S.-M. L. Arcos-García, "Deep neural network for traffic sign recognition systems: An analysis of spatial transformers and stochastic optimisation methods.," *Neural Networks,* vol. 99, pp. 158-165, 2018.