# Computer Vision: Assignment #2

Pramit Mallick

pm2758

October 25, 2018

Taking a closer look at the dataset, we realized that any normal network would require a lot of data augmentation to help train instances where an image or the region of interest in the image is rotated and scaled and cropped. Augmenting data and adding more CNN layers and fully connected layers into the given base model could've solved the problem by brute force if trained for enough epochs, but it would be prone to overfitting. After some research, we found the Spatial Transformer Network (STN) by Jaderberg et al. [2015] to be an interesting solution to the problem. STNs allow a neural network to learn how to perform spatial transformations on the input image in order to enhance the geometric invariance of the model. We trained a simple STN and achieved the following results -
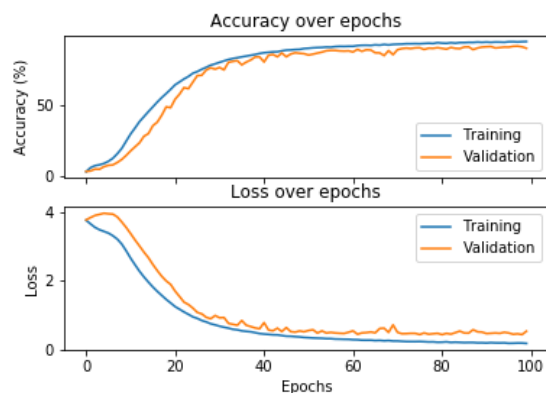
But it only achieved 91.44% accuracy on the validation set and 0.94726 score on the test set.

Closer inspection showed us that the model was oscillating in its error rate after a certain epoch, suggesting that the step size must be reduced. Thus to ensure convergence, I reduced the learning rate by half when the validation accuracy reached 80%, then again at 85% and finally at 90%. I also added another convolution layer (increased the number of filters at each stage as well), introduced batch normalization and added another fully connected layer at the end. This lead to the model performing 95.47% on the validation set and 0.97149 score on the test set. The convergence results are -
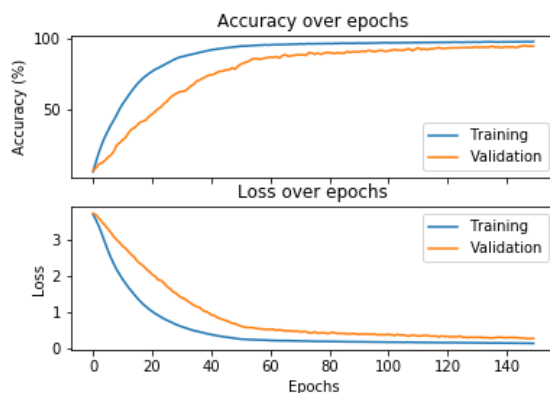
The relevant model is included in the zipped attachment with the name `model_stn_lr2_conv.pth`.

References

Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. In *Advances in neural information processing systems*, pages 2017–2025, 2015.