Assignment 2: Classifying STL-10 with a deep feed-forward convolutional neural net

Priyank Bhatia
New York University
Center for Urban Science + Progress
1 MetroTech Center, 19th Floor
Brooklyn, NY 11201
pb1672@nyu.edu

Emil Christensen
New York University
Center for Urban Science + Progress
1 MetroTech Center, 19th Floor
Brooklyn, NY 11201
erc399@nyu.edu

Peter Varshavsky
New York University
Center for Urban Science + Progress
1 MetroTech Center, 19th Floor
Brooklyn, NY 11201
pv629@nyu.edu

1 Data

Designed to study the use of unlabeled data for image classification, the STL-10 dataset [1] consists of three sets of 96x96 pixel RGB color images: 5000 labeled training images, 8000 labeled test images, and 100000 unlabeled images. Each image belongs to one of 10 categories. In this submission we only utilize the labeled portion of STL-10 and attempt to improve classification quality by image augmentation and depth of network.

2 Architecture

Two versions of a feed-forward convolutional neural net architectures were compared. The first (CP) architecture was an implementation of the baseline model by Christian Puhrsch with one convolutional layer of 23 7x7 pixel learned filters with a step size of 2, ReLU nonlinearity, 2 sq. pixel max pooling with step size 2, 50% dropout, a 50-node fully connected layer, ReLU, LogSoftMax and, and a negative log likelihood criterion. The second model (A1) used the same basic architecture but consisted of two convolutional layers of 200 and 400 5x5 pixel filters each, and a 800-neuron fully connected linear layer. Spatial pooling, rectified linear unit nonlinearities and dropout were applied as in the first network.

3 Preprocessing and augmentation

The original 5000 training images were split into training and validation sets of sizes 4500 and 500. To ameliorate the small training size and improve feature invariance the 4500 training images were cloned twice. The first cloned set was flipped horizontally, and the second was rotated counter-clockwise by 0.35 radians. This yielded an augmented training set of 13500 images. We further attempted to augment the training set using contrast HSV color space adjustments similar to contrast2 in [2], small random translations and rotations, but ran into training convergence issues, possibly due to coding errors. Augmented data were converted to YUV color space. Training images were globally normalized. Validation and test images were globally normalized using training mean and standard deviation. All images were further locally normalized and given a 2-pixel zero padding.

4 Training Procedure

Both models were trained with mini batch stochastic gradient descent with batch sizes 1, 8, 32, and 128. Additionally a shorter version of model A1 (two 200-filter convolutional layers, 400-node fully connected layer) was evaluated. Non fully stochastic mini batches improved both runtime and accuracy as suggested in [3]. The full A1 architecture yielded the best performance on the validation set with mini batch size 8. The best-performing model was then retrained on the full 5000-image training set and evaluated on the test set. A learning rate of 0.1 with annealing factor 0.001 were used.

	A1 model accuracy		
	Train	Validation	Test
1	75.76%	52.20%	-
8	93.27%	62.0%	64.04%
32	89.81%	61.6%	-
128	63.22%	54.2%	-

5 Results

The following table presents the results of both models in training on validation split, and on full data. The models trained on full training set were tested against the test set of 8000 images.

Model CP			
	train accuracy	validation accuracy	test accuracy
train with validation	82.82%	48.0%	-
train without validation	84.07%	-	51.75%

Model A1			
	train accuracy	validation accuracy	test accuracy
train with validation	93.27%	62.0%	-
train without validation	92.90%	-	64.04

6 A note on augmentation

The table shows A1 model results trained on several augmentations of the 5000-image training set. Because these evaluations were attempted close to the submission deadline, they were not done with the validation split. The entire training set was augmented, and accuracy was assessed on the test set. These results suggest that augmenting the training set can reduce the number of epochs needed for training convergence and improve classification accuracy.

Augmentations	Augmented size	Training accuracy	Test Accuracy	Number of Epochs
Original + horizontal reflection	10,000	99.68%	63.45%	37
Original + horizontal reflection + rotations (20° and 40°)	15,000	99.23%	67.48%	31
Original + horizontal reflection + rotation + rotation and reflection	20,000	99.43%	67.75%	29

obtained for different augmentations, for all the augmentations mentioned, same model that has been discussed was used and it was observed that increasing the train dataset size through different pre-processing techniques leads to faster convergence and number of epochs at which the most optimal test accuracy observed keeps on decreasing. Also the optimal test accuracy is observed when, rotation (20 degrees,), horizontal flip and rotation (40 degrees) are produced and added to the original training dataset. This increases the original dataset size from 5000 to 20000, and this configuration shows highest test accuracy of 67.75

References

[1] A. Coates, H. Lee, and A.Y. Ng. An analysis of single-layer networks in unsupervised feature learning. In Geoffrey Gordon, David Dunson, and Miroslav Dudk, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *JMLR Workshop and Conference Proceedings*, pages 215–223. JMLR W&CP, 2011.

[2] Alexey Dosovitskiy, Jost Tobias Springenberg, Martin Riedmiller, and Thomas Brox. Discriminative unsupervised feature learning with convolutional neural networks. *CoRR*, abs/1406.6909, 2014.

[3] Y. LeCun, L. Bottou, G. Orr, and K. Muller. Efficient backprop. In G. Orr and Muller K., editors, *Neural Networks: Tricks of the trade*. Springer, 1998.