Object classification for autonomous vehicle navigation of Stanford campus

Heather Blundell and Sarah Thornton {hrblun, smthorn}@stanford.edu

Problem

Many autonomous vehicle platforms are outfitted with low-cost cameras to capture data similarly to human eyes. We applied and compared the following techniques for image classification:

1. Softmax Regression
2. Support Vector Machines
3. Convolutional Neural Networks

Datasets

CIFAR-10 and CIFAR-100 32×32 pixels

ImageNet Cropped and scaled to 128×128 pixels

GoPro Footage Parsed with automatic object recognition algorithm [1], then cropped and scaled to 128×128 pixels

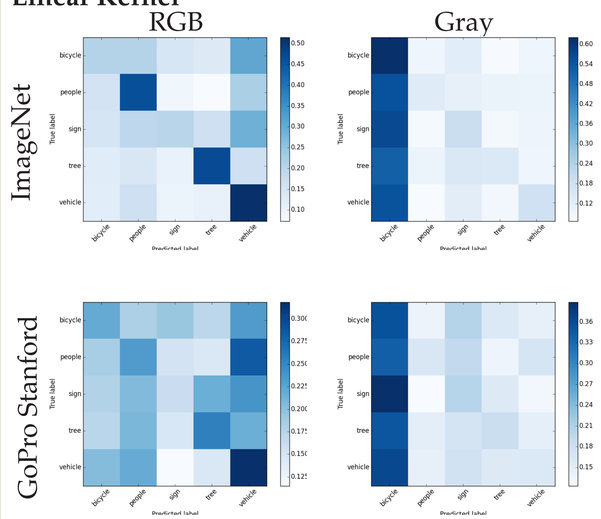
Dataset sizes*

	Train	Validate	Test
CIFAR	500	100	-
ImageNet	∼4,000	1,000	-
GoPro	-	-	600

* values are per class.

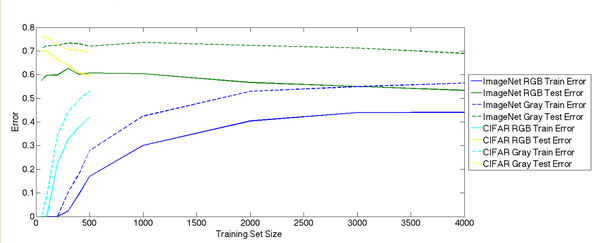
Support Vector Machines

Linear Kernel

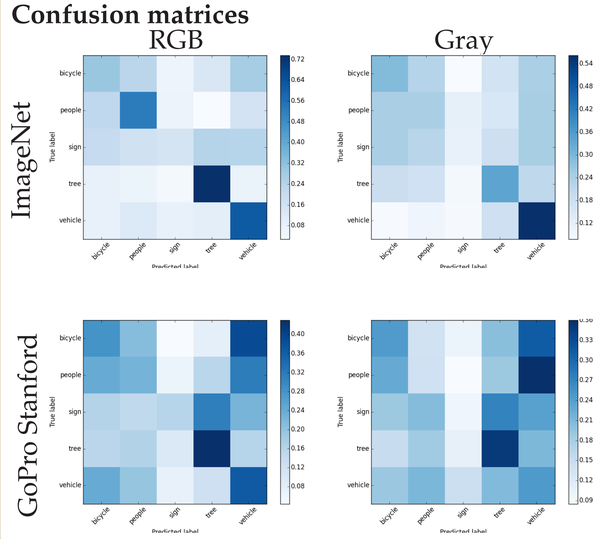


Softmax Regression

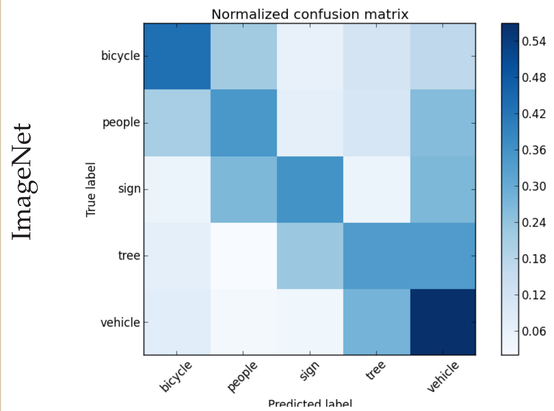
Learning Curve

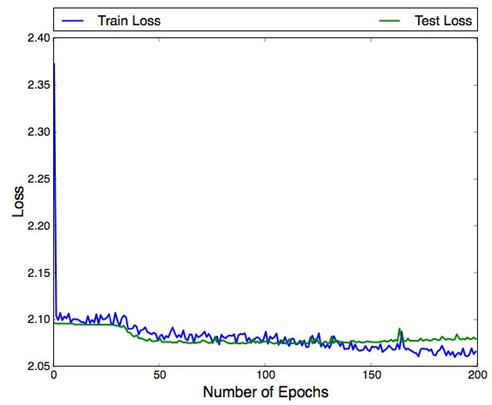


Confusion matrices



Convolutional Neural Networks





We have implemented a 4-layer CNN architecture (VGG-style [2]) using a categorical cross-entropy loss function to classify our ImageNet dataset using the Keras Deep Learning Library. We began with a 2D convolutional layer with 12 small (3×3) convolution filters. To control against overfitting, we applied Dropout twice [3], reduced the size of our two Dense layers, and added L2 regularization. However, the the CNN still began to overfit at approximately 100 epochs.

Discussion

Lessons learned from datasets:

	Pros	Cons
CIFAR	consistent size/resolution	small # examples; poor resolution
ImageNet	larger number of examples	varying quality
GoPro	images from car driving on Stanford campus	requires a lot of pre-processing and images inconsistent (i.e. exposure)

Overall, CNN worked best as indicated by visual inspection of the confusion matrices. Each dataset we investigated with the supervised learning algorithms had their own challenges. With CIFAR, it is a small dataset with poor resolution, but contains clean and consistent images. ImageNet provides a much larger dataset but they vary in size, quality, and resolution requiring more time to process before passing to our algorithms.

After more hyperparameter tuning to improve its accuracy, we plan to test our CNN against our Stanford footage dataset.

References

[1] K. E. A. van de Sande, J. R. R. Uijlings, T. Gevers, and A.W. Smeulders. Segmentation as selective search for object recognition. In *ICCV*, 2011. Matlab pcode available at `http://koen.me/research/selectivesearch/`

[2] K. Simonyan and A. Zisserman, âĂIJVery deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[3] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. "Dropout: A simple way to prevent neural networks from over-fitting." *The Journal of Machine Learning Research*, 2014.