Convolutional Networks Applied to House Numbers Digit Classification

Pierre Sermanet, Soumith Chintala and Yann LeCun

The Courant Institute of Mathematical Sciences - New York University

sermanet@cs.nyu.edu, sc3104@cs.nyu.edu, yann@cs.nyu.edu

Characters recognition in documents can be considered a solved task for computer vision whether hand-written or typed. It is however a harder problem in the context of complex scenes like photographs where best current methods lag behind human performance. [1] recently introduced a new digit classification dataset built from house numbers extracted from street level images. It is similar to the popular MNIST dataset (10 digits, 32x32 inputs), but an order of magnitude bigger (600,000 labeled digits) and contains color information.

[1] demonstrate the superiority of features learned all the way through the architecture as opposed to hand-designed features. Such superiority was also previously shown among others in a traffic sign classification challenge [2] where two independent teams obtained all top scores against various other approaches using convolutional neural networks (ConvNet) with learned features [3, 4]. [1] also argue about the superiority of unsupervised learning on this task. We obtain a 2.7 points accuracy improvement (93.3% accuracy) over the current state-of-the-art of 90.6% using a fully supervised ConvNet. This ConvNet departs from the traditional ConvNet architecture by using L2 pooling layers rather than average subsampling, and skipping connections between stages to the classifier [3]. It is implemented using the EBLearn C++ open-source framework [1] [5].

Algorithm	SVHN-Test Accuracy
Binary Features (WDCH)	63.3%
HOG	85.0%
Stacked Sparse Auto-Encoders	89.7 %
K-Means	90.6%
Supervised ConvNet	**93.3%**
Human Performance	98.0%

Figure 1: **Left:** 32x32 cropped samples for classification. **Right:** Performance reported by [1] with the addtional Supervised ConvNet state-of-the-art accuracy of 93.3%.

[1] http://eblearn.sf.net

References

[1] Netzer, Y, Wang, T, Coates, A, Bissacco, A, Wu, B, and Ng, A. Y. Reading digits in natural images with unsupervised feature learning. In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning*, 2011.

[2] Stallkamp, J, Schlipsing, M, Salmen, J, and Igel, C. The German Traffic Sign Recognition Benchmark: A multi-class classification competition. In *IEEE International Joint Conference on Neural Networks*, pages 1453–1460, 2011.

[3] Sermanet, P and LeCun, Y. Traffic sign recognition with multi-scale convolutional networks. In *Proceedings of International Joint Conference on Neural Networks (IJCNN)*, 2011.

[4] Ciresan, D. C, Meier, U, Masci, J, and Schmidhuber, J. A committee of neural networks for traffic sign classification. In *International Joint Conference on Neural Networks*, pages 1918–1921, 2011.

[5] Sermanet, P, Kavukcuoglu, K, and LeCun, Y. Eblearn: Open-source energy-based learning in c++. In *Proc. International Conference on Tools with Artificial Intelligence (ICTAI'09)*. IEEE, 2009.