Offline Mathematical Expression Recognizer

Shawn Peng: yisupeng@iu.edu
Yang Zhang: zhang505@Indiana.edu

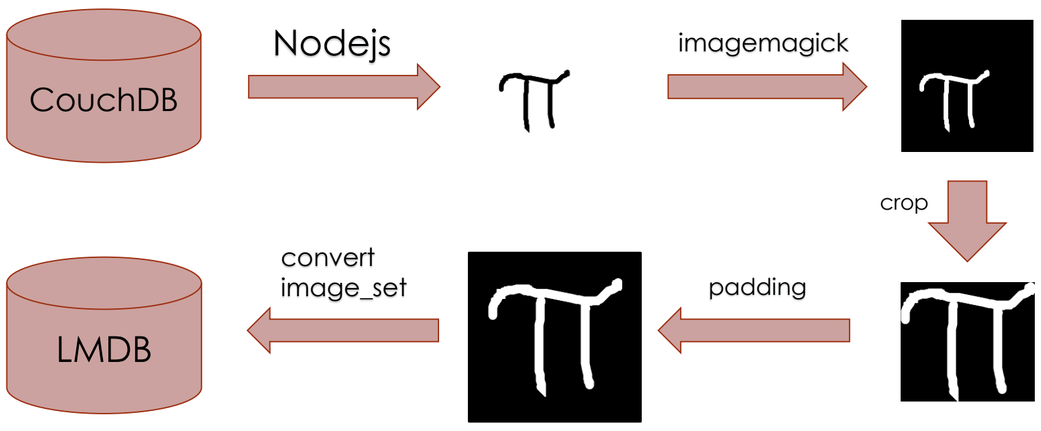
2016 Spring B657 Computer Vision
Prof. David Crandall

Data Set

7	Digits	MNIST
K	English	Chars74K, Write-Math
Ψ	Greek	Detexify
∫	Math Symbols	Write-Math

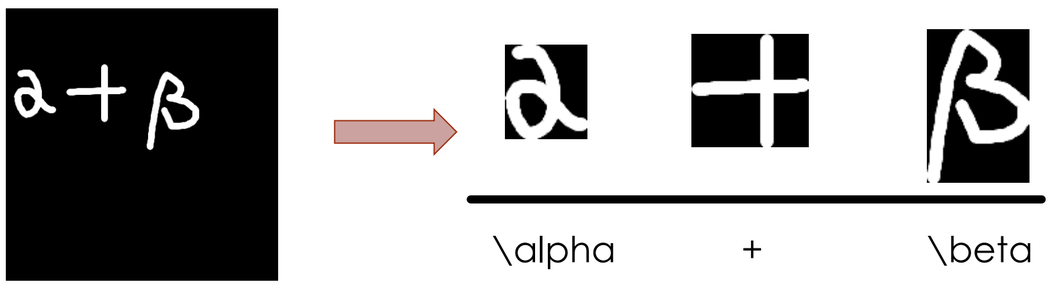
Since we use different data sources, and they use different image formats, we need to pre-process those data. Some of them are pre-processed, like Mnist and Chars74K dataset, but Chars74k, Detexify and write-math dataset uses white background.

Preprocessing



- We use Nodejs to convert CouchDB data to SVG and make the stroke border.
- We use imagemagick to negate, resize the image and convert it to png
- We use python to crop the image and add our image.
- We balance the training and testing data set by resampling
- The last step is to use Caffe tool: conver_image_set, to convert the image to lmdb

Result



Conclusion & Discussion

First, we found it is very useful to have the character to be properly bold before we send it to the convolutional neural network. For SVG image this can be done very easily, for images one thing we can use is the dilution.
Second, in the training of our convolutional neural network (we use LeNet), we find it is very useful to perform the trim operation on images.
On the other hand, the problem is that we do not have enough data to train a neural network. We have enough digit number and geek letter data, but for English letter we only get 50 samples. We believe that is why we only get 81.5% accuracy.
We have not begun to do grammar parse because of the poor performance of classifications.
We thank Prof. David Crandall for really helpful comments and discussions.
We also thank Dr. Daniel Kirsch and Martin Thoma to provide us the permission to the handwritten data.

Reference:

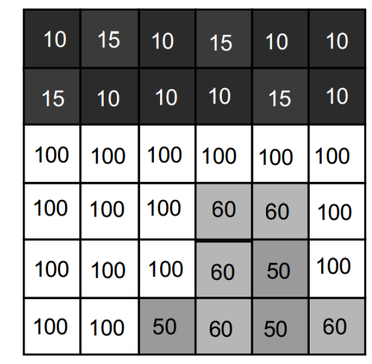
[1] Adam Coates, Blake Carpenter, Carl Case, Sanjeev Satheesh, Bipin Suresh, Tao Wang, David J. Wu, and Andrew Y. Ng. Text detection and character recognition in scene images with unsupervised feature learning. In Proceedings of the 2011 International Conference on Document Analysis and Recognition, ICDAR '11, pages 440{445, Washington, DC, USA, 2011. IEEE Computer Society.
[2] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Ecient graph-based image segmentation. Int. J. Comput. Vision , 59(2):167{181, September 2004.
[3] Lan Nguyen Nicolas D. Jimenez. Recognition of Handwritten Mathematical Symbols with PHOG Features.
[4] Michael Nielsen. Neural Networks and Deep LearningNeural Networks and Deep Learning. http://neuralnetworksanddeeplearning.com/chap6.html, 2016.

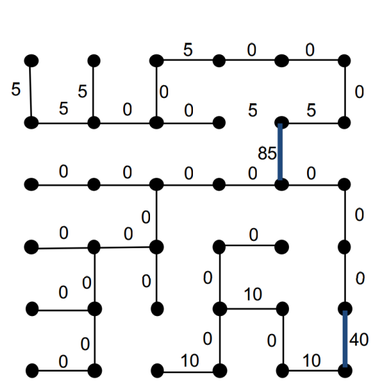
Segmentation

$$e^{xi} = \cos x + i \sin x$$

$$e^{xi} = \cos x + i \sin x$$

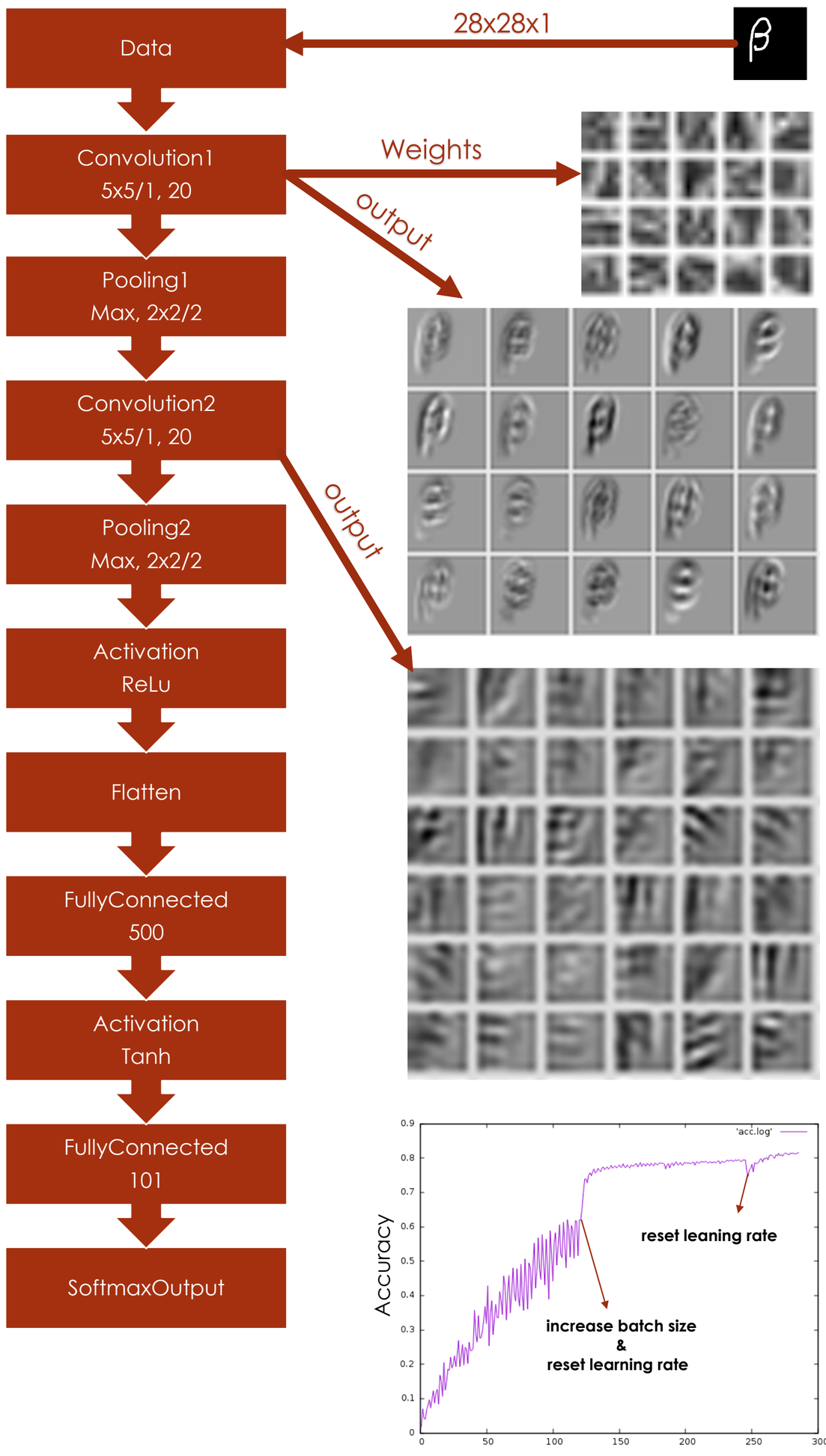
We use the spanning tree algorithm, which treat the image as a graph, and assign edge to every two adjacent pixels, according to their difference in color. After we construct the graph, we find the minimum spanning tree for this graph, and cut some edges with weights over a threshold to get a forest.





From: Prof. Crandall's B657 course ppt

Convolutional Neural Network



Tech we used

Caffe