**On the benefits of Convolutional Neural Network Combinations in Offline Handwriting Recognition**

(Suryani Dewi, Doetsch Patrick et al., 15th International Conference on Frontiers in Handwriting Recognition , 2016)

Summary:

CNN ist state of the art in the recognition of offline handwriting text. However, a combination of CNN and long short-term memory recurrent networks (LSTM) can improve the results significantly. There are two experiments (database for chinese characters: CASIA (online)) conducted in this paper:

* Single character per input image
* Sequence modeling for more characters in one image (= full text line images). LSTMs are used here with CNNs as preprocessing layers

CNN with 5 layers is used for feature recognition due to memory constraints. The fully connected layer is replaced by three BLSTM layers (512 cells for each direction).

Recurrent Neural Networks (RNNs) consider previous output in their hidden layers as well and thus have memory (information can be saved and reused) but have a problem called *vanishing gradient* problem, which led to the usage of LSTMs (3 additional gates that protect gradient information within the neural network).

Hidden Markov Model (HMM) is an algorithm and widely used for sequential or temporal data. It contains transition probabilities for segmentation ([Viterbi alignment](https://en.wikipedia.org/wiki/Viterbi_algorithm)).

Possible problems:

* Cursive style of handwriting -> CNN solves this in a good way.
* Different size and shape of each character
* Large vocabularies
* Segmentation of sentences into letters and words

Takeaways for our project:

* **Feature extraction with CNN**
* Sliding window (*we need further investigation on this*) for text
* Further **reading**:
  + [2] “Tandem HMM with CNN for handwritten word recognition,” in 2013 IEEE ([Link](https://drive.google.com/open?id=1HuBSSDW-jiy21YBeal1TPdC7UcyuO3-l))
  + [16] “Framewise phoneme classification with bidirectional LSTM networks,” in 2005 IEEE ([Link](https://drive.google.com/open?id=19uHahGfII23M3hLa830NW5vu5L4Tmt5S))
* Approach:
  + Segment the lines *using* [*ground truth*](https://en.wikipedia.org/wiki/Ground_truth)*, 8x32 window, sliding 3 pixels*
  + Examine CNN on single characters

Others:

1. Database

(<http://www.fki.inf.unibe.ch/databases/iam-handwriting-database>) is a **database** that contains **forms of unconstrained handwritten text**, which were **scanned at a resolution of 300 dpi** and saved as **PNG** images with **256 gray levels**. It **provides samples of a complete form, a text line and some extracted words**. This database was downloaded and might be helpful to train our CNN features in the beginning.

1. Mathwork projects

[1] Matlab ocr() function: [Link](https://www.mathworks.com/help/vision/ref/ocr.html#bt548t1-2_1)

[2] Optical Character Recognition (OCR): [Link](https://www.mathworks.com/matlabcentral/fileexchange/18169-optical-character-recognition--ocr-)

[3] OPTICAL CHARACTER RECOGNITION(lower case and space included): [Link](https://www.mathworks.com/matlabcentral/fileexchange/31322-optical-character-recognition-lower-case-and-space-included-)

Paper:

Suryani, D., Doetsch, P., & Ney, H. (2016, October). On the Benefits of Convolutional Neural Network Combinations in Offline Handwriting Recognition. In *Frontiers in Handwriting Recognition (ICFHR), 2016 15th International Conference on*(pp. 193-198). IEEE.

**Optical Character Recognition Implementation Using Pattern Matching**

(Faisal Mohammad, Jyoti Anarase, Milan Shingote, Pratik Ghanwat ISB&M School of Technology, Nande, Pune- 412115)

Summary:

The author performs the Pattern Matching algorithm based method for Optical Character Recognition (OCR) to convert images of handwritten or typewritten characters to the editable binary matrix, which effectively reduces the image processing time while maintaining the accuracy. The binary matrix is divided into different sections. A corresponding track-sector matrix is generated by calculating the number of ones in each intersection of sector and track. The pattern based recognition requires matching of the binary matrix with the existing template. The generated track-sector matrix is compared with the template. If all the parameters match, the character is identified.

Applicability:

This paper presents a basic OCR system and an effective algorithm, which helps us to build main structure of our own OCR system. It provides a pattern, which we can follow and make extension. Therefore, it is applicable. When we have complicated situation, we can try to change some parameter in the system to satisfy our achievement.

Issue:

The performance of the systems have been constrained by the dependence on font, size, and orientation. Our project involves recognition of different type of typewritten or handwritten, especially messy handwriting. We need to improve OCR system to process more complicated image. When the input image character is stick together, pre-processing module should be applied several times to get the correct bounding box of every character. When the input image characters are scrawled, pattern generation should be applied several times and compare every track-sector matrix which has a better recognition result.

Paper:

Mohammad, F., Anarase, J., Shingote, M., & Ghanwat, P. (2014). Optical character recognition implementation using pattern matching. International Journal of Computer Science and Information Technologies, 5(2), 2088-90.

**Tandem HMM with Convolutional Neural Network for Handwritten Word Recognition**

(Theodore Bluche, Hermann Ney, and Christopher Brown, *Spoken Language Processing Group.* 2013)

Summary:

The authors combine hidden Markov models (HMM) with convolutional neural networks (CNN’s) to recognize handwritten words. Both hybrid combinations and tandem combinations are used. In a hybrid combination, CNN’s are used in lieu of Gaussian mixtures to compute emission probabilities. In tandem combination, the CNN’s serve as the feature extractor for HMM’s. The combined approaches resulted in up to 60% improvement compared to Gaussian HMM’s. The CNN architecture used in this experiment is most similar to LeNet5. Preprocessing of the images fixes the slant in each image, crops the images, and enhances the contrast.

Applicability:

Our project involves recognizing handwritten characters using some type classifier, such as SVM’s or CNN’s. This paper uses convolutional neural networks in combination with other models to recognize handwritten words, so it is applicable. The paper states that there are both geometric and directional features to be extracted, which is useful information in the case that we use handcrafted features. The reference to LeNet5 offers some insight into what kind of CNN to use for transfer learning.

Issues:

The authors use classification models not covered in CSSE463, so those parts of their classifiers cannot be effectively replicated or referenced. We will only be using a single classifier, whether it is a CNN or SVM. The original paper reports significant improvement over such classifiers, so it is likely that our classifier will not perform as well as a combined classifier.

Paper:

Theodore Bluche, Hermann Ney, and Christopher Brown. Tandem HMM with Convolutional Neural Network for Handwritten Word Recognition. *LIMSI CNRS, Spoken Language Processing Group.* 2013.