A State-of-the-art Deep Learning Approach for Handwriting-based Recognition on Gender and Handedness

Yanbang Wang, Jiangshan Li, Tiancheng Cai
{ywangdr, jiangsli, caitch}@stanford.edu, Stanford University

Problem & Task

Motivation

- Handwriting-based identification of gender and handedness has a variety of applications in forensic biometrics and archaeology, while most existing methods rely heavily on handcrafted features combined with classic statistical methods.

Results

- We report a deep-learning based method that achieved state-of-the-art performance on both English and Chinese handwriting dataset.
  - Our model outperforms humans by a large margin and achieves an error rate at least 30% lower than the existing best models in two different test settings.
- We collect and publish a new handwriting dataset for gender classification consisting of 350 handwriting samples from 32 writers.
- We build a website that allows viewers to predict gender from their own handwriting images.

Data

- **IAM-OnDB (Public dataset)**
  - 1,700 English handwriting samples (13,049 text lines) from 221 writers with gender and handedness labels.
- **HIT-MW (Public dataset)**
  - 853 Chinese handwriting samples (8,644 text lines) from 780 writers with gender labels.
- **Collected Dataset**
  - We collect 350 handwriting writing samples from 32 writers in Chinese and English with gender and handedness labels.

We perform background whitening and foreground darkening using OSTU Threshold, detect white margin, zoom, resize, crop line images into images of size (80, 320) pixels and normalize.

FIGURE 1: IAM-OnDB image after preprocessing

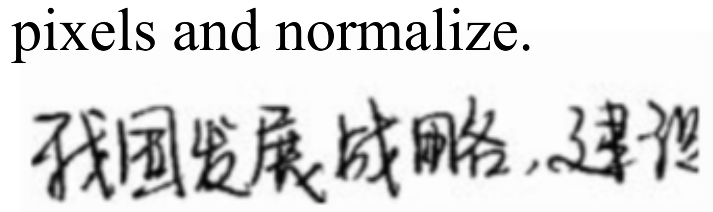

FIGURE 2: HITMW image after preprocessing

Features

We use preprocessed handwriting images as input to our CNNs. We consider data augmentation methods including rotation, shifting, rescaling and morphological operations (binary erosion/dilation) to apply to training data.

Models

Binary Classification loss:

$$\mathcal{J} = -\frac{1}{m}\sum_{i=1}^{m}(\alpha y_i \log \hat{y}_i + \beta(1-y_i)\log(1-\hat{y}_i))$$

α, β: weights for (potentially skewed categories)

IAM-OnDB

- CNN as shown in FIGURE 3, with BatchNorm, dropout = 0.5, Adam Optimizer

HIT-MW

- CNN as shown in FIGURE 4, with BatchNorm, dropout = 0.3, Adam Optimizer

Cross-Languages

- **Transfer Learning**
  - HIT-MW dataset is much smaller compared to IAM-OnDB
  - Freeze the previous layers from IAM-OnDB model and fine tunes the last two / four layers for Chinese dataset
- **SafetyNet for English/ Chinese classification**
  - CNN: 4 CONV layers (16, 32,64,128, 4x4 kernel); 3 FC layers (512, 128, 10), with BatchNorm and dropout = 0.1
  - 99.88% test accuracy

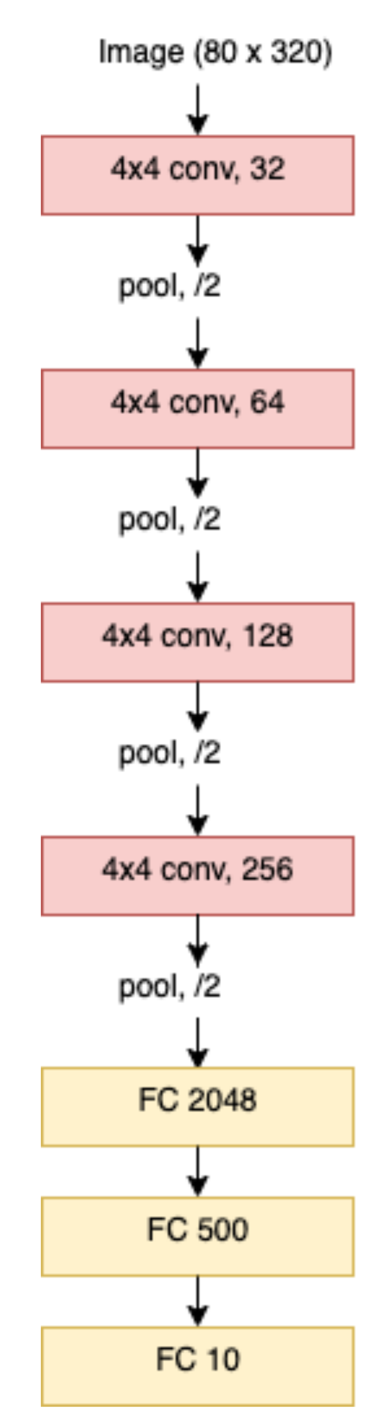

FIGURE 3: IAM-OnDB model

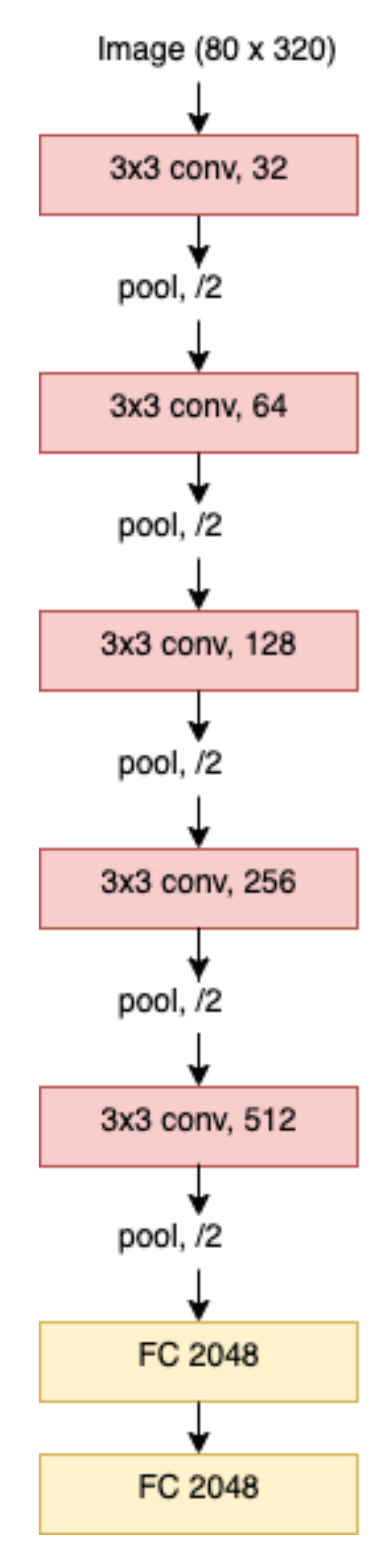

FIGURE 4: HITMW model

Results

IAM-OnDB

Labels	Training Accuracy	Test Accuracy
Our (original)	0.968	0.859
Morera et al. [1]	/	0.807
Bouadjenek et al. [2]	/	0.755
Our (gender balanced)	0.709	0.714
Liwicki et al. [3]	/	0.676

Table 3: Accuracy results on IAM-OnDB

Labels	Num Train Samples	Num Test Samples	Precision	Recall	F1 Score
Male	27853	6911	0.880	0.916	0.898
Female	13019	3307	0.809	0.739	0.772
Overall	40872	10218	0.844	0.828	0.830

Labels	Num Train Samples	Num Test Samples	Precision	Recall	F1 Score
Male	31499	3265	0.743	0.655	0.696
Female	13061	3265	0.691	0.773	0.730
Overall	44560	6530	0.717	0.694	0.705

Table 2: Performance results on gender-balanced IAM-OnDB

HIT-MW

Labels	Training Accuracy	Test Accuracy
HITMW	0.782	0.755

Table 5: Accuracy results on HITMW

Labels	Num Train Samples	Num Test Samples	Precision	Recall	F1 Score
Male	5854	632	0.712	0.698	0.705
Female	7413	844	0.854	0.891	0.872
Overall	13267	1476	0.783	0.795	0.780

Table 4: Performance results on HIT-MW Dataset

Discussion

- We achieve state-of-the-art accuracy on the task of gender classification for both Chinese and English Datasets.
- We didn't follow up on the classification of handedness because of the skewness of dataset (roughly 9:1)
- The task of gender classification from handwriting is inherently difficult, and we are amazed that we are able to beat human performance (around 64.88% on English) by a large margin.

Future

- We would study the transferability of our model further by integrating a third languages (for example, having train, dev and test sets coming from three different languages).
- We would work on the visualizing what the network has learned.
- We would like to study the fairness issue involved in such model.

Reference

[1] Ángel Morera, Ángel Sánchez, José Francisco Vélez, and Ana Belén Moreno, "Gender and Handedness Prediction from Offline Handwriting Using Convolutional Neural Networks," Complexity, vol. 2018, Article ID 3891624, 14 pages, 2018. https://doi.org/10.1155/2018/3891624.

[2] N. Bouadjenek, H. Nemmour, and Y. Chibani, "Histogram of Oriented Gradients for writer's gender, handedness and age prediction," in *Proceedings of the International Symposium on Innovations in Intelligent Systems and Applications, INISTA 2015*, IEEE, Madrid, Spain, August 2015.

[3] Liwicki, M., A. Schlapbach, H. Bunke. Automatic Gender Detection Using On-Line and Off-Line Information. – Pattern Analysis and Applications, Vol. 14, 2011, No 1, pp. 87-92.