Machine Learning Engineer Nanodegree

Capstone Proposal

Alex Fong Jie Wen
December 11th, 2017

Proposal

Domain Background

Japanese consists of three types of scripts, kanji, hiragana and katakana. Roman numerals and symbols are also commonly used. Offline Japanese character recognition has been applied to a number of applications, such as transcription of documents and even as a tool for learning the language. Convolutional neural networks (CNNs) have been used to obtain state-of-the-art performance in character recognition for various languages [1-3] but few attempts were made for the Japanese language. There is also remarkable progress in research for using CNNs in portable devices. However, the focus has been on commonly used languages such as Chinese [2], or CNNs in general [5].

Problem Statement

This capstone project aims to explore the feasibility of conducting deep learning based optical character recognition (OCR) for printed Japanese text, on mobile devices. More specifically, the goal is to obtain a model which requires a small amount of memory and runs with acceptable speed on the average mobile device. To focus the scope of the project, handwritten text will not be considered in testing. This is fair, given that OCR is used more frequently for printed material rather than written material with the exception of mail sorting and other niche usages.

Dataset and Inputs

The ETL-2 dataset will be used for this project. ETL-2 consists of Japanese characters printed for newspapers and characters printed for patent applications. Each image is a character, with dimensions 60 by 60. There are 52769 images in total. The characters come in 2 different fonts, Mincho and Gothic.

The ETL-2 dataset is balanced, with machines printing on a datasheet to create the dataset. Majority of JIS level one Kanji, hiragana, katakana, the roman alphabet and symbols are found in this dataset. The specific characters can be found here.
http://etlcdb.db.aist.go.jp/etlcdb/etln/etl2/e2code.jpg

The ETL-9B dataset will also be used for this project. ETL-9B consists of 3036 handwritten Japanese characters from 4000 different writers. JIS level one Kanji, hiragana and katakana, are present in this dataset. The specific characters can be found here.
http://etlcdb.db.aist.go.jp/etlcdb/etln/etl9/e9sht.htm

The ETL-9B dataset is balanced, with each person writing a number of characters once by filling up a datasheet.

Both datasets have been binalized with Otsu's method.
The two datasets can be found here:
http://etlcdb.db.aist.go.jp/?page_id=1721
http://etlcdb.db.aist.go.jp/?page_id=1711

The datasets in the ETL character database have been collected by Electrotechnical Laboratory, universities and other research organizations for character recognition researches from 1973 to 1984.

Solution Statement

A solution would be a model which runs without requiring large amounts of memory, and makes predictions at an acceptable speed on mobile phone CPUs. An acceptable amount of memory would be around 100mb or less considering the capacities of smartphones today. An acceptable speed would be about 5 characters in 1 second.

Datasets will be decoded to extract black & white PIL images. The model used will either be a custom deep learning network or a common one like ResNet, Inception or VGG. Feature map pruning will be conducted to reduce the memory consumption of the deep learning models.

The loss function used will be classification cross entropy loss, and we will be using stochastic gradient descent for training the models. Labels for the images will be left untouched since deep learning models do not benefit from one hot encoding. The images will be resized and normalized when fine tuning common networks for transfer learning.

Benchmark Model

A good benchmark model for memory and computation time is the results obtained by Xiao et al [2]. Their network for handwritten Chinese character recognition required 2.3MB of storage and took 9.7 ms per character image on a single threaded CPU. Chinese characters are extremely similar to Japanese characters, partially sharing the same character set. The network had a top-1 error of 2.91%. A similar performance with respect to memory and computation time is desired.

A good benchmark model for accuracy is the results obtained by Tsai [4]. His network for handwritten Japanese character recognition had an estimated recognition rate of above 96.1%. A slightly lower recognition rate is desired, given the accuracy and speed/memory tradeoff.

Evaluation Metrics

The models produced will be evaluated based on precision, amount of storage required, and the average time taken for each character on a single threaded CPU or possibly a mobile CPU.

Precision = (true positives)/(true positives + false positives)
Only precision will be used as a measure of accuracy, given that no mistakes are tolerated in OCR. A confusion matrix

Runtimes and memory usage will be recorded in python.

Project Design

The following networks will be examined for their viability in terms of memory, speed and accuracy.

1) transfer learning with ETL-2
2) transfer learning with ETL-2 & ETL-9
3) custom network with ETL-2
3) custom network with ETL-2 & ETL-9

All of the above networks will be pruned of their feature maps to reduce memory consumption and improve computational speed with the approach suggested by Molchanov et al. [5]

The candidates for transfer learning are ResNet and Inception. Further research will be conducted to select one of those networks. Since character images is quite different from natural images, performance from transfer learning might be poor. A custom network adopted by Xiao et al [2] for Chinese recognition will be attempted as well.
A part of the ETL-2 dataset will be partitioned out for testing. Training will be conducted on ETL-2 only, and also on ETL-2 + ETL-9 to see if additional but noisy data improves accuracy.

The networks will be trained on a Nvidia GTX 1060-6GB card, and will be tested on one core of an Intel I3-4130 machine. Testing may also be conducted on middle end mobile phones.

References

[1] Zhang, X., Bengio, Y., & Liu, C. (2016, June 18). Online and Offline Handwritten Chinese Character Recognition: A Comprehensive Study and New Benchmark. Retrieved December 10, 2017, from https://arxiv.org/abs/1606.05763

[2] Xiao, X., Jin, L., Yang, Y., Yang, W., Sun, J., & Chang, T. (2017, February 26). Building Fast and Compact Convolutional Neural Networks for Offline Handwritten Chinese Character Recognition. Retrieved December 10, 2017, from https://arxiv.org/abs/1702.07975

[3] Wojna, Z., Gorban, A., Lee, D., Murphy, K., Yu, Q., Li, Y., & Ibarz, J. (2017, August 20). Attention-based Extraction of Structured Information from Street View Imagery. Retrieved December 10, 2017, from https://arxiv.org/abs/1704.03549

[4] Tsai, C. Recognizing Handwritten Japanese Characters Using Deep Convolutional Neural Networks Retrieved December 10, 2017, from https://*cs231n.stanford.edu/reports/2016/pdfs/262_Report.pdf*

[5] Molchanov, P., Tyree, S., Karras, T., Aila, T., & Kautz, J. (2017, June 08). Pruning Convolutional Neural Networks for Resource Efficient Inference. Retrieved December 10, 2017, from https://arxiv.org/abs/1611.06440