Two-dimensional Word Embedding and Transfer Learning from Vision to Text with System Demo on 300mW CNN Accelerator Chip

Baohua Sun, Lin Yang, Wenhan Zhang, Michael Lin
Gyrfalcon Technology Inc.
1900 McCarthy Blvd, Milpitas, CA, US
baohua.sun@gyrfalcontech.com

Abstract

Power-efficient CNN Domain Specific Accelerator (CNN-DSA) chips are currently available for wide use in mobile devices. These chips are mainly used in computer vision applications. However, the recent work of Super Characters method for text classification and sentiment analysis tasks using two-dimensional CNN models has also achieved state-of-the-art results through the method of transfer learning from vision to text. In this paper, we implemented the text classification and sentiment analysis applications on mobile devices using CNN-DSA chips. The two-dimensional embedding is language-independent, because the text is converted into image. At the workshop, we have two interactive system demonstrations for the NLP tasks. The first demo classifies the input English Wikipedia sentence into one of the 14 classes. The second demo classifies the Chinese online-shopping review into positive or negative.

Figure 1: Efficient On-Device Natural Language Processing system demonstration. The CNN-DSA chip is connected to Raspberry Pi through the USB interface. Keyboard sends the typing text input to Raspberry Pi through USB. A monitor is connected to Raspberry Pi through HDMI for display. On the monitor, it shows the introduction for the demo (zoom in to see details). There are two demos. The first demo classifies the input English Wikipedia sentence into one of the 14 ontologies. The second demo classifies the Chinese online-shopping review into positive or negative.

1 Introduction

Power-efficient CNN Domain Specific Accelerator (CNN-DSA) chips are currently available for wide use. Sun et al. designed a two-dimensional CNN-DSA accelerator which achieved a power consumption of less than 300mW and an ultra power-efficiency of 9.3TOPS/Watt. Demos on mobile and embedded systems show its applications in real-world implementations. The 28nm CNN-DSA accelerator attains a 140fps for 224x224 RGB image inputs at an accuracy comparable to that of the VGG.

For Natural Language Processing tasks, RNN and LSTM models are widely used, which are different network architectures from the two-dimensional CNN. However, the recent work of Super Characters method using two-dimensional word embedding achieved state-of-the-art result in text classification and sentiment analysis tasks, showcasing the promise of this new approach. The Super Characters method is a two-step method. In the first step, the characters of the input text are drawn onto a blank image, so that an image of the text is generated with each of its characters embedded by the pixel values in the two-dimensional space. The resulting image is called the Super Characters image. In the second step, the generated Super Characters image is fed into a two-dimensional CNN models for classification. The two-dimensional CNN models are trained for the text classification task through the method of Transfer Learning, which finetunes the pretrained models on large image dataset, e.g. ImageNet, with the labeled Super Characters images for the text classsification task.

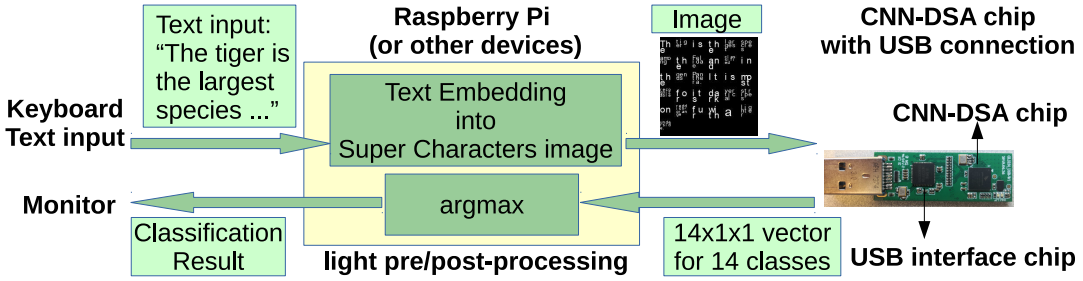

Figure 2: Data flow for the system demonstration of efficient on-device NLP using Super Characters method and CNN-DSA chip.

In this paper, we implemented NLP applications on mobile devices using the Super Characters method on a CNN-DSA chip as shown in Figure 1. It takes arbitrary text input from keyboard connecting to a mobile device (e.g. Raspberry Pi). And then the text is pre-processed into a Super Characters image and sent to the CNN-DSA chip to classify. After post-processing at the mobile device, the final result will be displayed on the monitor.

2 System Design and Data Flow

As shown in Figure 2, the keyboard text input is pre-processed by the Raspberry Pi (or other mobile/embedded devices) to convert into a Super Characters image. This pre-processing is only a memory-write operation, which requires negligible computation and memory resources.

The Super Characters method works well for Asian languages which has characters in squared shapes, such as Chinese, Japanese, and Korean. These glyphs are easier for CNN models to recognize than Latin languages such as English, which is alphabets-based in a rectangular shape and may have to break the words at line-changing. To improve the performance for English, a method of Squared English Word (SEW) is proposed to cast English word in a squared shape as a glyph. Figure 3 shows an example of this method. Basically, each word takes the same size of a square space lxl. Words with longer alphabets will have smaller space for each alphabet. Within the lxl space, the word with N alphabets will have each of its alpha in the square area of $\{l/ceil[sqrt(N)]\}^2$, where $sqrt(.)$ stands for square root, and $ceil[.]$ is rounding to the top.

The CNN-DSA chip receives the Super Characters image through the USB connection to the

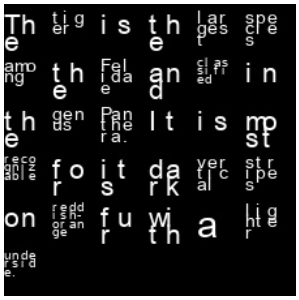

Figure 3: An example for Squared English Word (SEW) method. The two-dimensional embedding in this image corresponds to the text of ”The tiger is the largest species among the Felidae and classified in the genus Panthera. It is most recognizable for its dark vertical stripes on reddish-orange fur with a lighter underside.”

mobile device. It outputs the classification scores for the 14 classes in the Wikipedia text classification demo. The classification scores mean the probabilities for classification but before softmax. The mobile device only calculate the argmax to display final classification result on the monitor, which is also negligible computations. The CNN-DSA chip completes the complex CNN computations with low power less than 300mw.

3 Conclusion

We implemented efficient on-device NLP applications on a 300mw CNN-DSA chip by employing the two-dimensional embedding used in the Super Characters method. The two-dimensional embedding converts text into images, which is ready to be fed into CNN-DSA chip for two-dimensional CNN computation. The potential use cases for this demo system could be the intension recognition in a local-processing smart speaker or Chatbot.