Large Vocabulary Handwritten Text Recognition Based on Deep Neural Networks

Kaixiang Huang, Yehan Wang, Jimmy Yu, Jianzhe Hu
{kh595, yw569, jy392, jh1207}@scarletmail.rutgers.edu
Advisor: Prof. Ivan Marsic  Group: S17-37

Motivations and Objectives

Motivations
In this era of digital information, computers and mobile devices play a pivotal role in how students learn. Countless time people find themselves wanting to convert their handwritten documents into digital, editable files with the ability to format it for further use. Existing solutions are not efficient or accurate enough to accomplish this task.

Objectives
Our goal is to take a picture of a handwritten document with a mobile device and send it to a computer to recognize the text and format of the contents of the photo accurately, in order to create a digital file that can be edited through standard processing applications such as Microsoft Word and Adobe Acrobat.

Background

Image Processing
OpenCV is a library of programming functions mainly aimed at real-time computer vision. We use it to process photos taken through our application.

Neural Network
Keras is an open source neural network library written in Python.
TensorFlow is an open source software library for machine learning across a range of tasks.
CUDA is a parallel computing platform and application programming interface (API) model created by Nvidia.
We built our network based on VGG16 using keras with TensorFlow backends and train it with NIST Special Database 19 using CUDA.

Research Challenges

Communication between cell phone application with computer recognition module.
Erase noises in the photo like shadows or pictures.
Identify the word if some letters can’t be recognized.

Acknowledgement
We would like to thank our adviser, Prof. Ivan Marsic and Dr. Xinyu Li for their guidance in this project.

System Design

Take a photo with application

Android Application
Application is responsible for creating a user-friendly interface that takes pictures of single-page documents and transfers it to the computer.
Application utilizes Android Camera API and Bluetooth 4.1.

Image Processing

OpenCV Image Processing

Neural Network

The CNN will have some misrecognition, that is why we apply a LSTM network after CNN. This particular network will take three letters to predict next letter. Here’s an example.

| Heat Map/Confusion Matrix of CNN |

If the letter recognized by CNN matches prediction, that is likely to be correct. Otherwise, we will give different options for user to choose which is the correct word. We will develop a better solution in the future.

References