Learning a Neural Network for Text Recognition

Perrin Ruth Aparajithan Venkateswaran Derek Wright

James Meiss
APPM 3310, Fall 2017
University of Colorado Boulder

Description

Modelling the human brain using science and mathematics has always been a tough and fascinating problem. In recent times, due to advancements in computing power, neural networks are being used widely to model the human brain at a very abstract level - neurons receiving electrical inputs from various sources and sending signals across synapses when it surpasses a threshold. The applications of neural networks involve automating many activities that a human can perform effortlessly, including object detection, speech recognition, text analysis, etc.

Deciphering text in an image with a computer, in particular, is a very difficult one [1]. Our project will focus on constructing a neural network to detect the presence of, and recognize text from a given image. This can be viewed as two sub-problems. The first problem is detecting the text itself in an image. This involves identifying a bounding box for the text, and each character. The second problem is classifying each letter into one of the 26 letters, 10 digits, or any of the special characters. But recent works have unified it into a single problem [3].

Previous research done in this field includes the famous LeNet 5 architecture, which was developed in 1998 for online handwriting recognition [2]. Another import work that is more recent is the multi-digit number recognition model developed in 2015 by Google for recognizing multi-

character text in natural photographs [3]. We will be using these works, as well as other relevant research, as inspiration to develop our model.

Applications

This technology has many applications in the real world. Transferring old texts and manuscripts from the written world to the digital world; deciphering inscriptions from tablets, documents and other archaeological artifacts; creating searchable images (menus, directories, other large textual datasets) are a few of them.

Google Translate uses this technology to help travelers quickly translate text in their immediate environment for a more seamless tourist experience.

CAPTCHA is a service that protects websites against bots by producing an image that a human can easily read, that a computer cannot. Recent advances in this field have forced us to look at more sophisticated methods for preventing bots from accessing websites.

References

[1] Mohit Agarwal Baijnath Kaushik (2015) *"Text recognition from image using Artificial Neural Network and Genetic Algorithm"*

[2] Yann LeCun, Leon Bottou, Yoshua Bengio, Patrick Haffner (1998) *"Gradient-Based Learning Applied to Document Recognition"*

[3] Ian J. Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, Vinay Shet (2014) *"Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks"*