

Contents

Application example: Photo OCR Optical Character Recognition

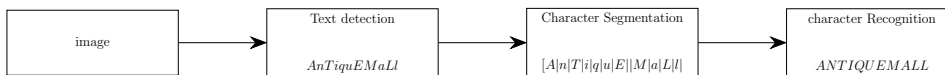
March 26, 2018

Part I

Week 11

0.1 OCR Problem Description

Photo OCR pipeline:



1. text detection using sliding window:

- use a labeled training set with positive examples ($y = 1$) when image is a text and $y = 0$ (negative example), the image has not text.
- the resulting image is a black and white image where the positive example are white
- Once text objects are detected, apply "expansion operation": this expands the white region for every pixel if the neighbouring white pixel is close

2. Character segmentation:

- Use supervised learning algorithm to look for gap in between letters : positive example $y = 1$ is an image with a gap between 2 letters.

3. Character recognition is performed using 1D sliding window

0.1.1 Sliding window for Pedestrian Detection

- This is a supervised learning, where the NN is trained on labeled images : image of pedestrian $y = 1$, image without pedestrian $y = 0$).
- on a test set image, we take a patch (green box below) and determine if $y = 1$ or $y = 0$, and move the box by pre-determined step-size: this is called **sliding window**
- we then increase the patch size



0.2 Artificial Data Synthesis

There are 2 techniques to new new data:

- create new data from scratch
- synthesizing data by introducing artificial distortions

This works for image, but also speech recognition (by introducing noise, background). Note that the distortion introduced should be representation of the type of noise/distorsions in the test data. However, usually it does not help to add purely random/meaningless noise to the data ($x \leftarrow x + \text{random noise}$)

0.3 Ceiling Analysis

With Ceiling Analysis, we estimate the errors due to each component of the pipeline, which helps determining what component is more likely to improve the overall model.

We illustrate the concept using OCR pipeline



component	accuracy test set
overall system	72%
Text detection plugging in the ground truth labels of the test set i.e making text detection 100% accurate	89%
(Manually) Character Segmentation	90%
(Manually) character recognition	100%

The largest improvement is obtained with improving "Text Detection".