ML

Andrew Ng

Coursera

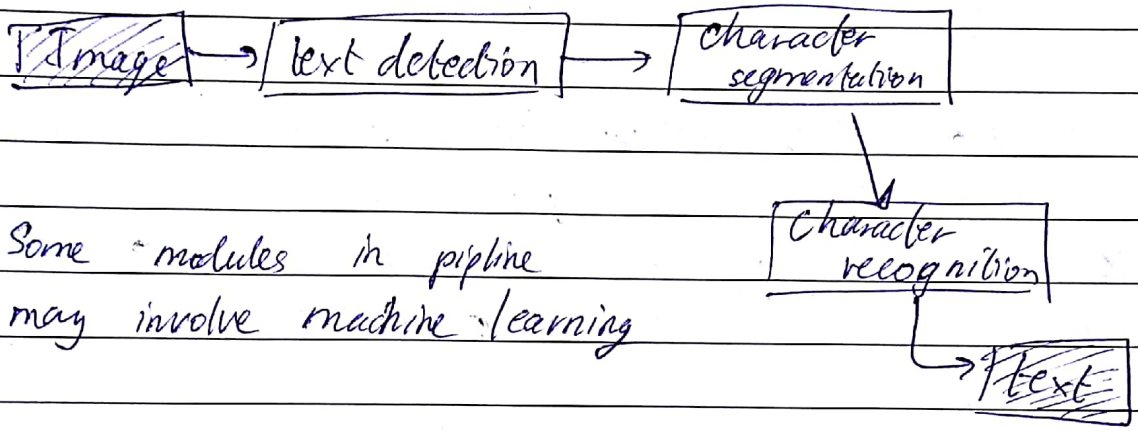
Week 11

Date:

Problem description & pipeline

Photo OCR pipeline

1) Text detection
2) Character segmentation ← break down text into characters
3) Character classification



Some modules in pipline may involve machine learning

Sliding Windows

For pedestrian detection in photos the aspect ratio of a person tends to be almost the same. (height to width ratio almost the same)

To ~~lear~~ teach an algo pedestrian detection get a dataset of +ve & -ve examples of pedestrians. Images can be of the same pixel size → supervised learning (82 x 36 patches suppose)

Date:

Suppose we have a big image & we wish to find pedestrians in it via the pedestrian classifier
(Note: there can be multiple pedestrians in test image in any location)

Sliding window detection - take 82 × 36 patches of the test image by sliding over the image a few pixels at a time & checking if pedestrian is present there

↳ ~~The sliding of the pixels~~ the # of pixels slided over is called step size / stride

↳ Then increase ~~decrease~~ image patch size & detect for ~~de~~ pedestrians

↳ you will have to convert the larger image into 82 × 36 for the classifier

In text detection the aspect ratio of the text is not the same. Once sliding window detection is done then apply "expansion" operator to results → expands regions where text was detected to form a rectangle around it

↳ this is how we combat the different aspect ratio problem

Date:

Character segmentation -

Train such a classifier where +ve examples denotes a portion of the overall text where there is a gap between 2 characters, -ve where this does not exist.

Once trained, then do 1D sliding window to figure out where different characters are by placing boundaries between letters

boundary right in the middle

Getting lots of data: Artificial data Synthesis

works in certain situations only

Artificial data synthesis for photo OCR

There are many word processing softwares with many different types of fonts. You those fonts & paste on random background to generate artificial data. → make data from ground up

You can also take real data & introduce distortions.

Date:

Distortion introduced should be representation of the type of noise/distortions in the test set.