Notes taken from Coursera Course – Andrew Ng's ML

by Chris Zeng, April 26, 2015

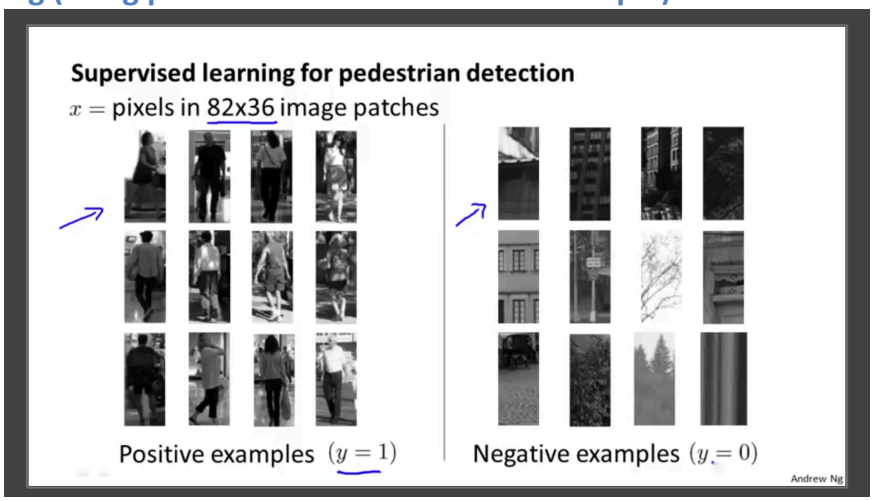
Photo OCR pipeline

1. Text detection
2. Character segmentation
3. Character recognition

1. Text detection

- Sliding window detection algorithm

1a.Training (using pedestrian detection as an example)



Using pedestrians' images as y=1 and other random images as y=0.

1b. Classifying (1st stage classification)

- Moving a rectangular window with a Step size/stride of 4 pixel (to the right, downwards) 1 pixel is the best, but will take a long time.
- Change the size of the window (but keeping the aspect ratio and rescale it to the size the classifier uses)

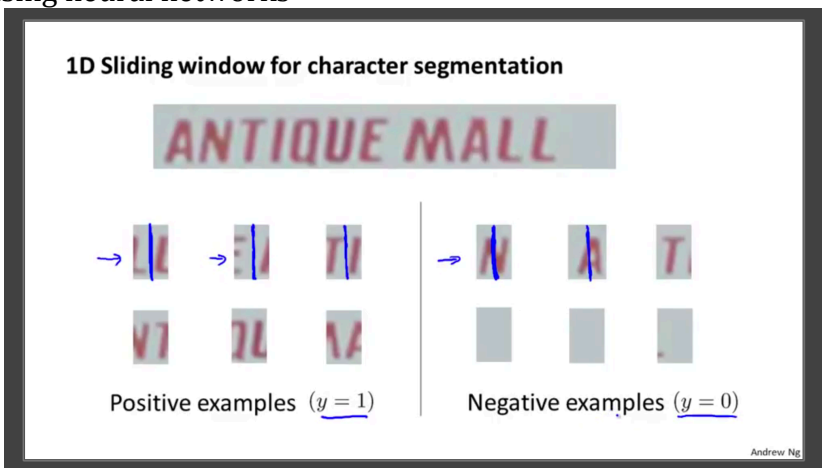
1c. Expansion (expanding the white regions so they become continuous region rather than disjointed)

- Is it within several distance from a white pixel, if so, classify it as 1.
- Looking at the aspect ratio of the white regions, if there are tall and thin rectangles, discard them as they are not very likely to be text.

2. Character segmentation

- Split characters

Training using neural networks



3. Character Classification

Using neural network or other similar techniques

Photo OCR pipeline

1. Text detection

2. Character segmentation

3. Character classification

Andrew Ng