Photo OCR

Sidharth Baskaran

June 2021

Problem Description and Pipeline

Pipeline

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

[Image] → [Text Detection] → [Character Segmentation] → [Character recognition]

Sliding Windows

- Text detection → difficult due to varying aspect ratios of detection zones
- Supervised learning approach
  - Decide on set pixel dimension for all images in training set
  - Positive ($y = 1$) and negative ($y = 0$) examples
- Slide a window through an unaltered image, can use a *stride parameter*, want to choose reasonable value for compromise of efficiency
- Image processing
  - Can take larger patches and scale down to set size
  - Colormap of text detection outlines where text is found, using the classifier and expanding/extrapolating positive regions
    - Can then createlarge bounding boxes

Artificial Data Synthesis

- Obtain more data artifically through synthetic methods
- Want more data for a low bias algorithm
- Ex: different fonts against random backgrounds, apply blurring, scaling, rotation, etc.
- Synthesis through distortions
  - Can introduce distortions and background noise for speech recognition
- Introduced distortions should be representative of the type of noise/distortions in **test set**
- Meaningless distortion does not help
- Getting more data
  - Want a low-bias classifier first before (plot learning curves)
    - Increase features of hidden units in an NN until low bias is achieved
  - Methods
    - Artificial data analysis
    - Collect/label it manually
    - Crowd source
    - Ask how much work it would be to get 10x data as currently had

Ceiling Analysis

- Help decide which components of timeline are best worth time
- Ceiling analysis → estimating errors due to each component

[Image] → [Text Detection] → [Character Segmentation] → [Character recognition]

- Single $\mathbb{R}$ evaluation metric, e.g. accuracy
  - Give each component the correct ground-truth output, i.e. simulating 100% accuracy and record accuracy for **entire system**
  - Can see how much overall performance increases along the pipeline, shows which areas need improvement