照片光学字符识别（photo optical character recognition）

Photo OCR

机器学习流水线（machine learning pipeline）

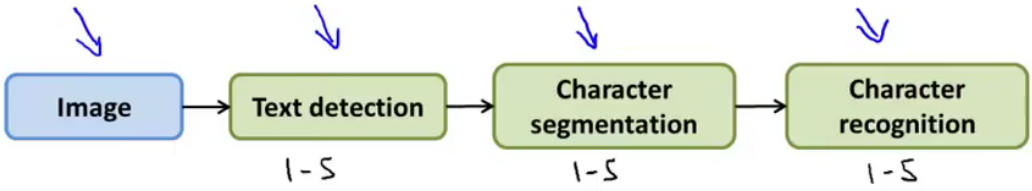
Photo OCR pipeline

→ 1. Text detection

→ 2. Character segmentation

→ 3. Character classification

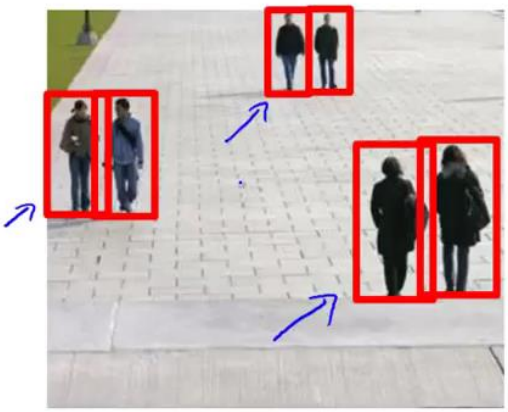
Photo OCR pipeline



滑动窗(sliding windows)的分类器

Text detection

Pedestrian detection



Supervised learning for pedestrian detection

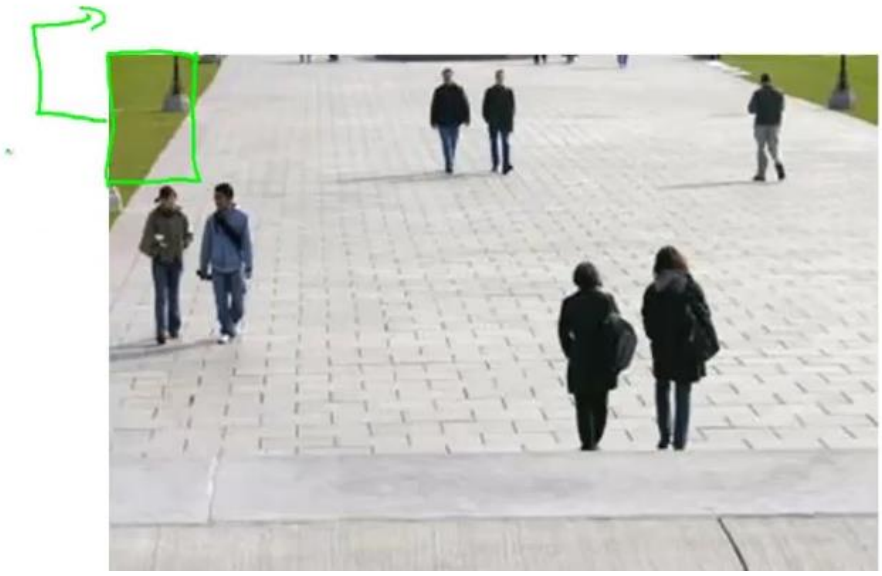
$x =$ pixels in 82x36 image patches

1,000
10,000
⋮

Positive examples $(y = 1)$

Negative examples $(y = 0)$

Sliding window detection

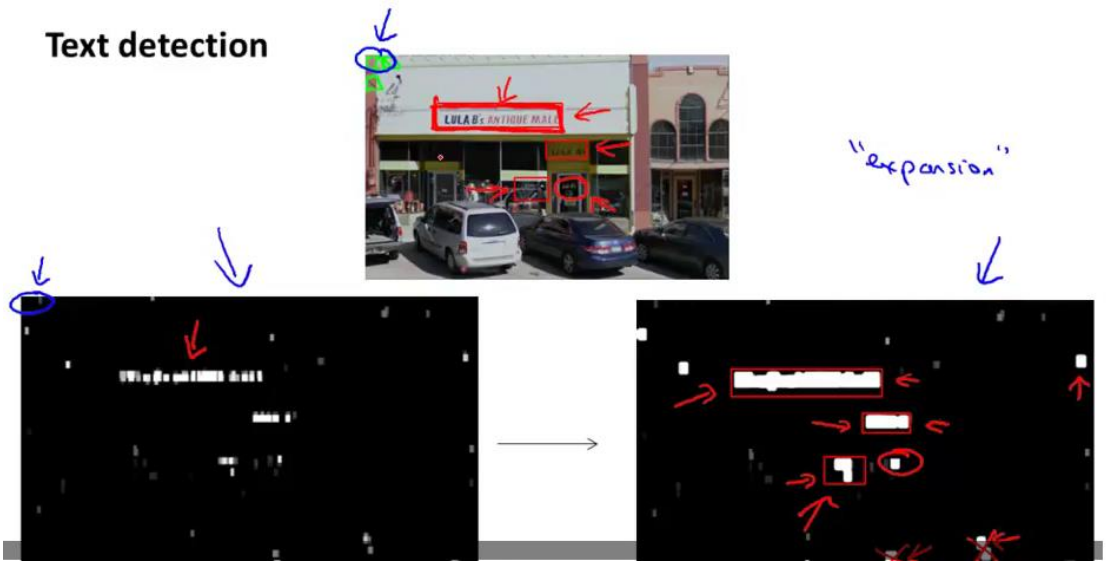


Text detection

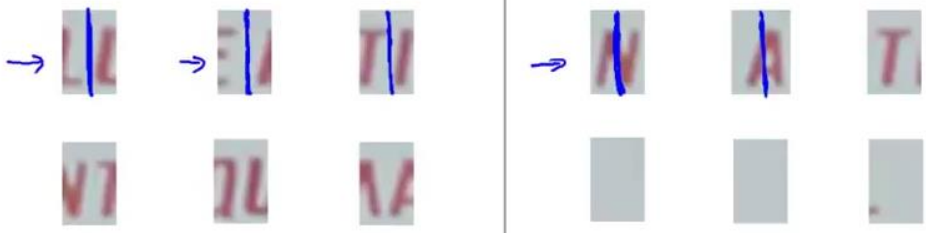
Positive examples $(y = 1)$

Negative examples $(y = 0)$

Text detection



1D Sliding window for character segmentation



Positive examples $(y = 1)$ | Negative examples $(y = 0)$

Photo OCR pipeline

1. Text detection

2. Character segmentation

3. Character classification

"人工数据合成"（artificial data synthesis）

Artificial data synthesis for photo OCR

Real data

Synthetic data

声音合成

Synthesizing data by introducing distortions: Speech recognition

Original audio:

Audio on bad cellphone connection

Noisy background: Crowd

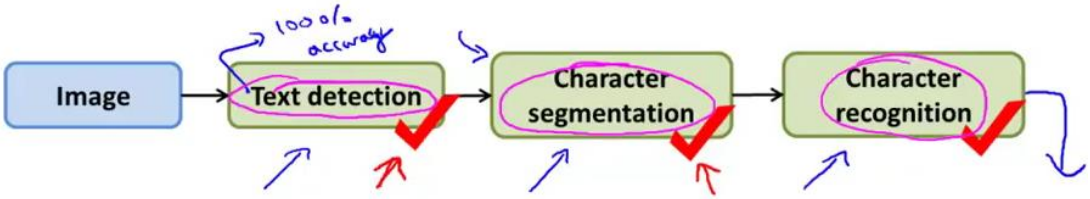
Noisy background: Machinery

Discussion on getting more data

1. Make sure you have a low bias classifier before expending the effort. (Plot learning curves). E.g. keep increasing the number of features/number of hidden units in neural network until you have a low bias classifier.
2. "How much work would it be to get 10x as much data as we currently have?"
   - Artificial data synthesis
   - Collect/label it yourself
   - "Crowd source" (E.g. Amazon Mechanical Turk)

上限分析(ceiling analysis)的内容

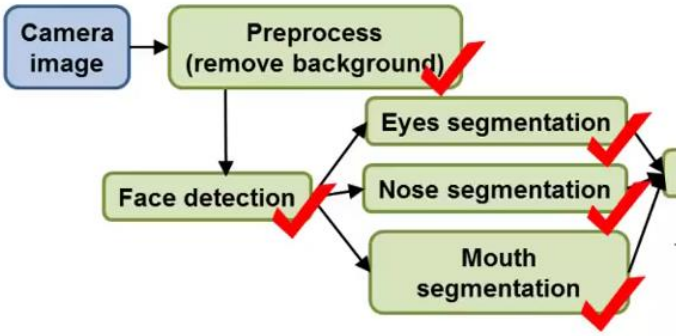
Estimating the errors due to each component (ceiling analysis)



What part of the pipeline should you spend the most time trying to improve?

Component	Accuracy
Overall system	72%
Text detection	89%
Character segmentation	90%
Character recognition	100%

Another ceiling analysis example



Component	Accuracy
Overall system	85%
Preprocess (remove background)	85.1%
Face detection	91%
Eyes segmentation	95%
Nose segmentation	96%
Mouth segmentation	97%