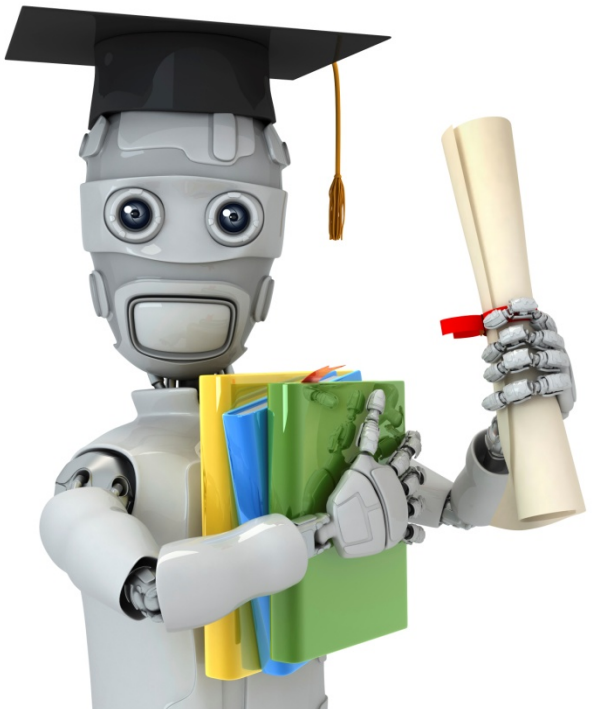


Machine Learning

Application example: Photo OCR 图片 OCR

Problem description and pipeline

OCR pipeline

The Photo OCR problem

Photo OCR pipeline

1. Text detection

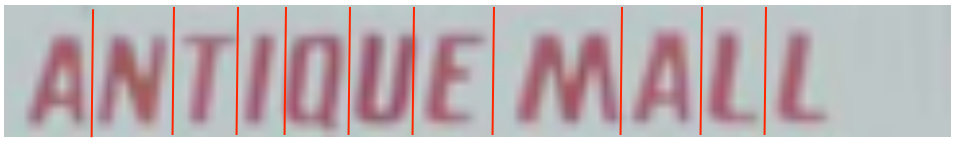
对输入图像，找文字信息区域

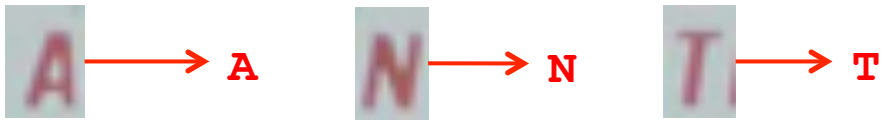
2. Character segmentation

识别单个字符区域

3. Character classification

分类器识别字符



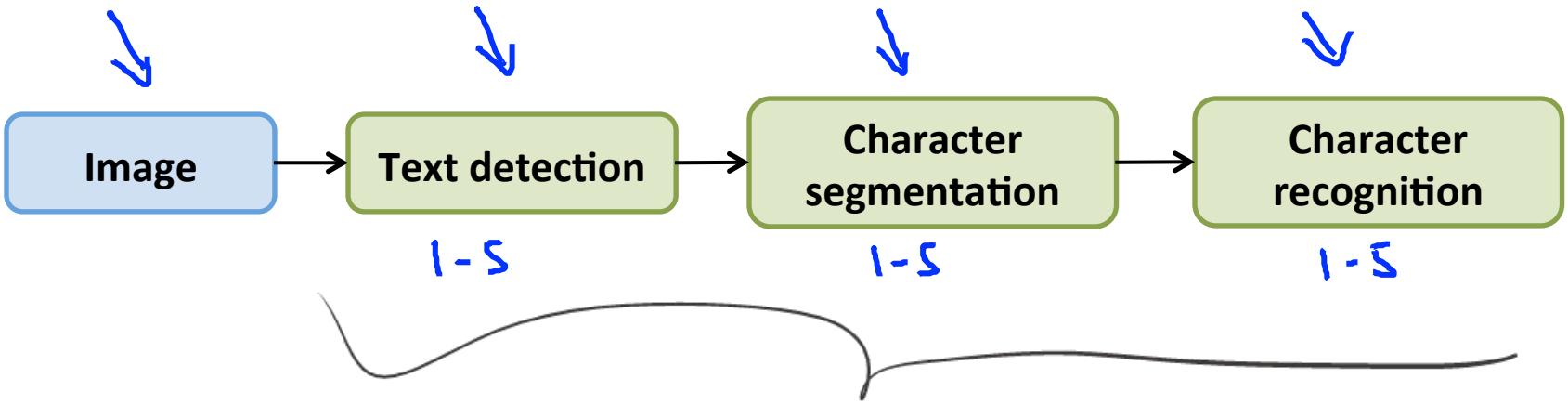


Cleaning → Cleaning

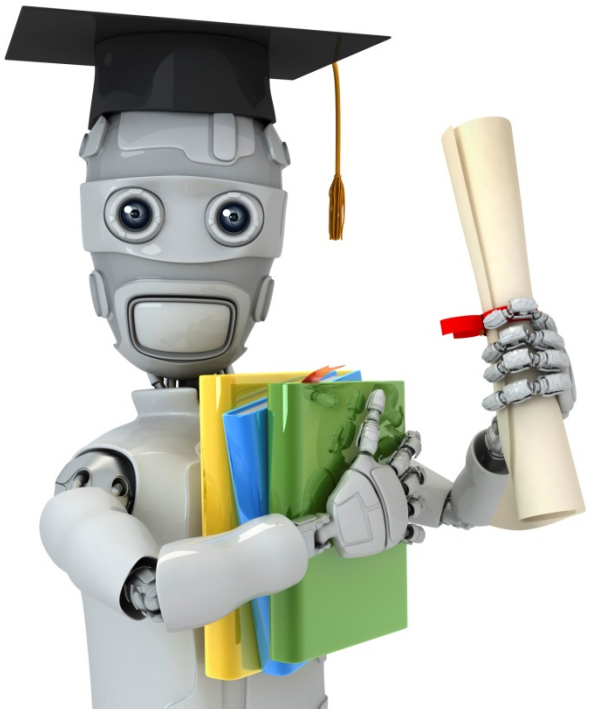
Andrew Ng

拼接整个系统，对字符分类算法识别出的字母进行修正

Photo OCR pipeline 照片OCR流水线



分别搭建组成流水线，便于分工



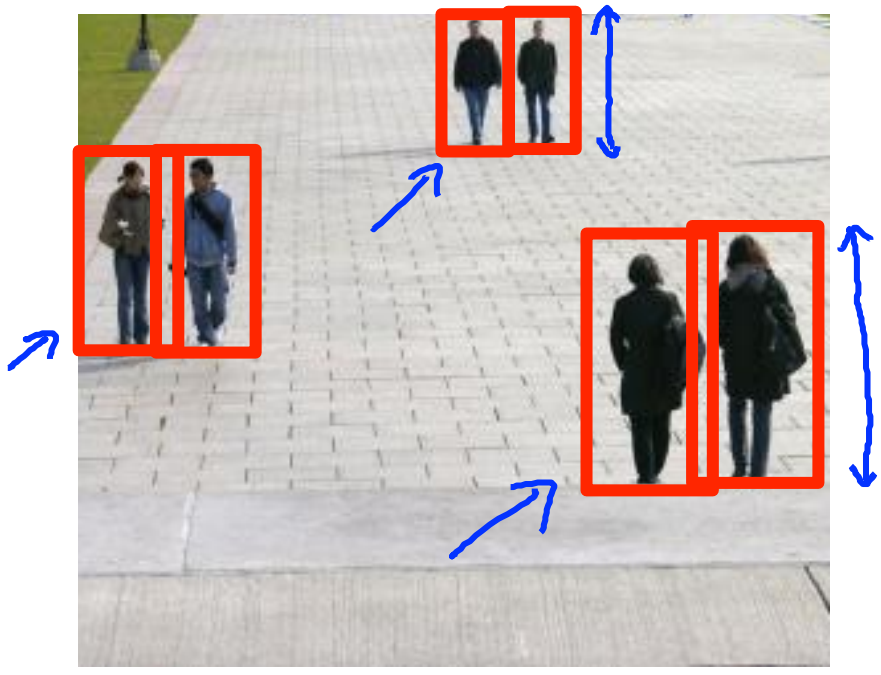
Machine Learning

Application example: Photo OCR

Sliding windows

Text detection

Pedestrian detection



Supervised learning for pedestrian detection

$x =$ pixels in 82x36 image patches

Positive examples $(y = 1)$

Negative examples $(y = 0)$

Andrew Ng

Sliding window detection

step-size / stride

对某大小图像，移动步长遍历整个图像

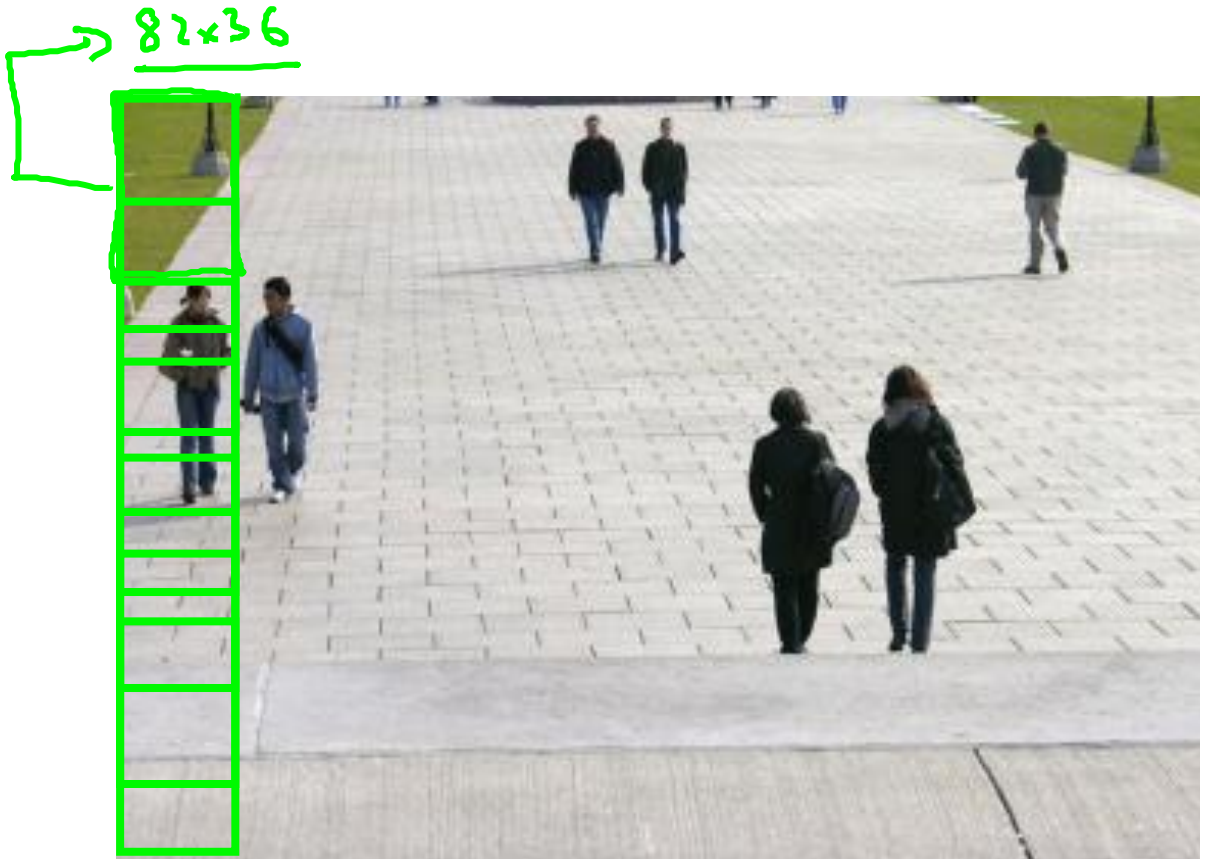
修改窗口大小，移动步长遍历整个图像

（每个图像放入分类器分类）

Andrew Ng

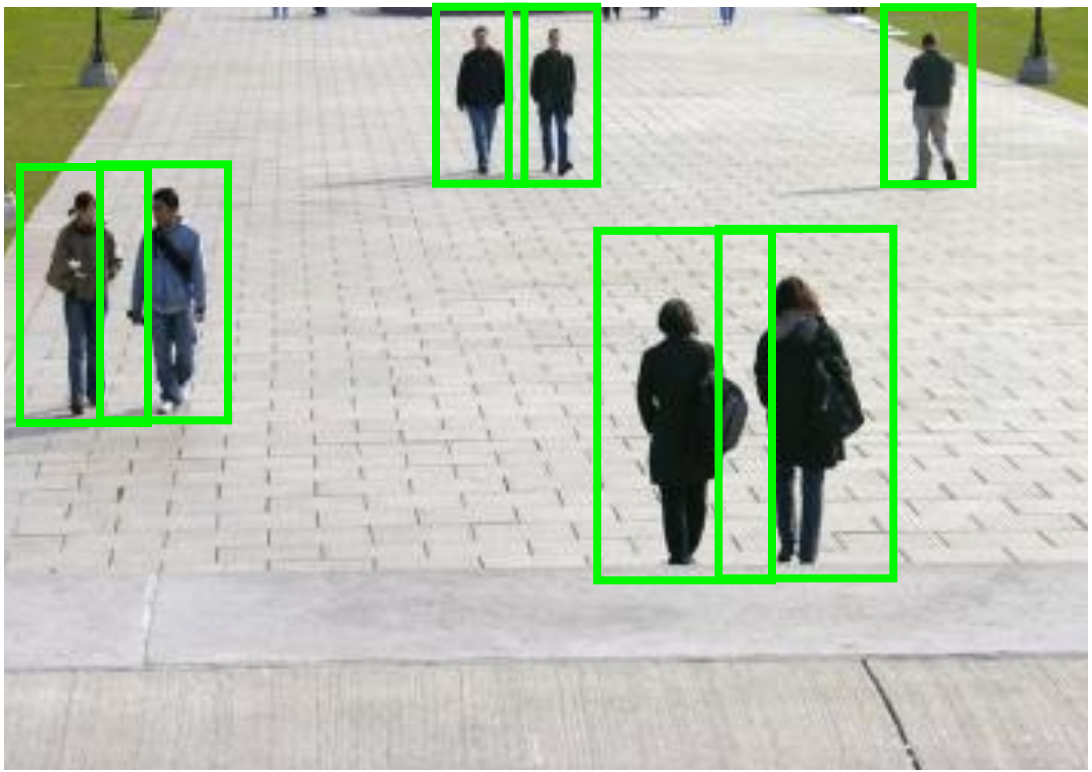
优化计算

Sliding window detection



Sliding window detection

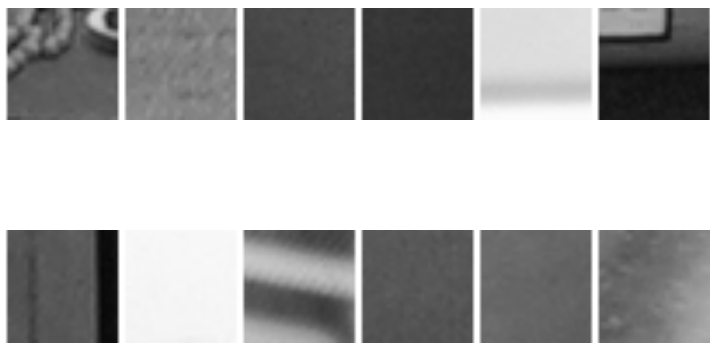
Sliding window detection



Text detection

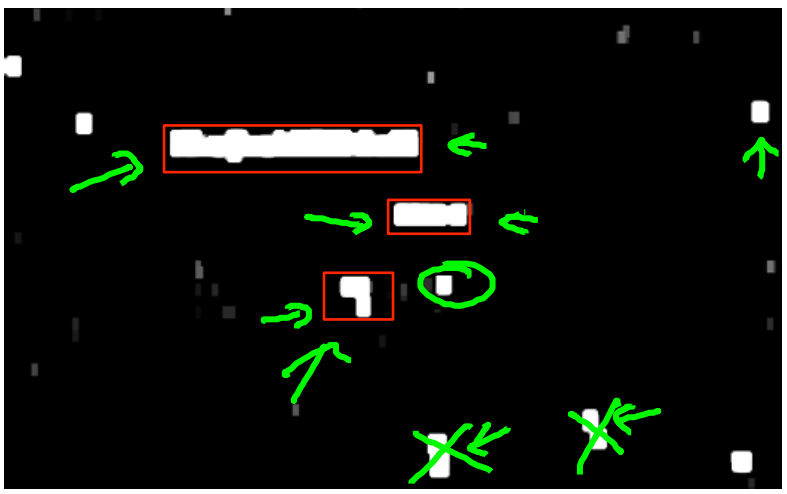
Text detection

Positive examples $(y = 1)$



Negative examples $(y = 0)$

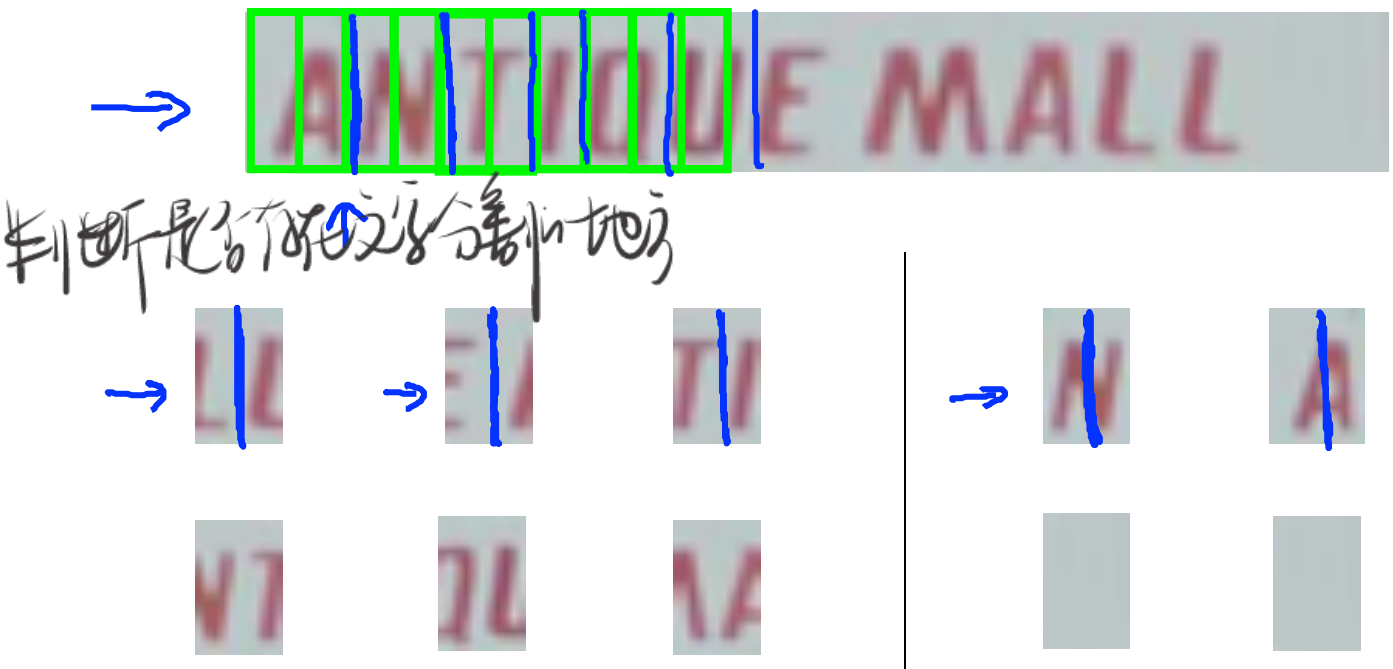
Text detection



[David Wu]

Andrew Ng

1D Sliding window for character segmentation



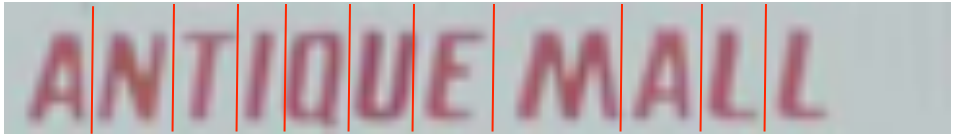
Positive examples $(y = 1)$

Negative examples $(y = 0)$

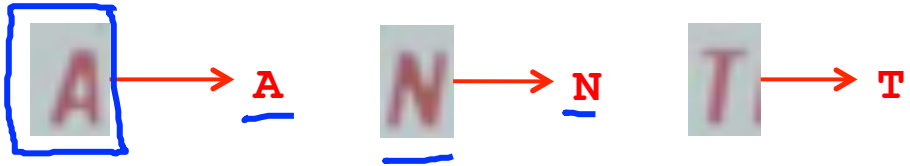
Photo OCR pipeline

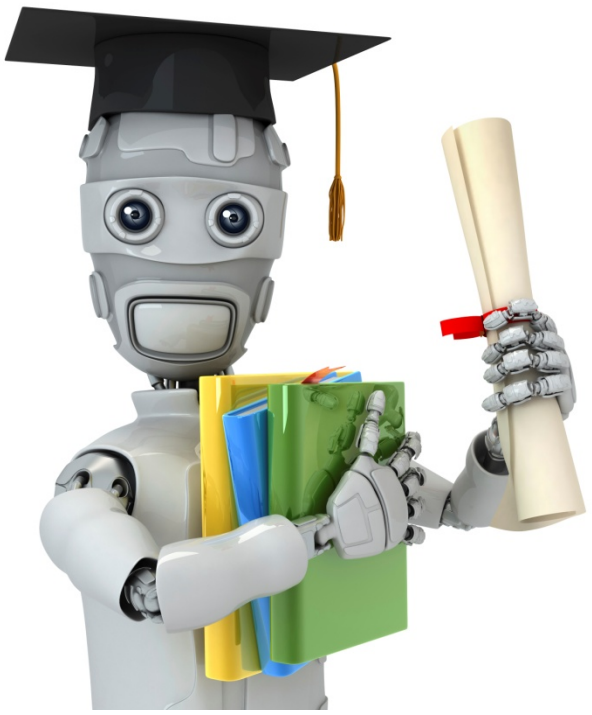
1. Text detection

2. Character segmentation



3. Character classification





Machine Learning

Application example: Photo OCR

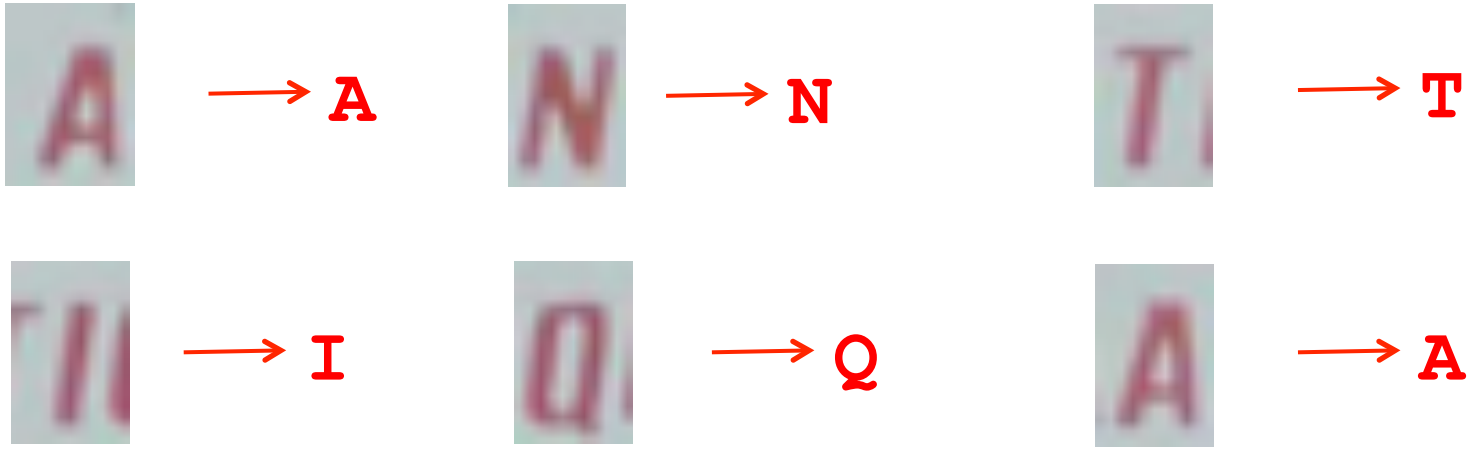
Getting lots of data: Artificial data synthesis

人工数据合成(得到更多数据)

从0开始创造新数据 (1)

确小好到训练集，再以某种方式扩充 (2)

Character recognition



Artificial data synthesis for photo OCR

Real data

[Adam Coates and Tao Wang]

Abcdefg
Abcdefg
Abcdefg
Abcdefg
Abcdefg

找大字体库，生成不同字体的字符

↓

再粘贴到任意不同的背景 →

Andrew Ng

or

模糊算子

仿射变换：等分 缩放 旋转

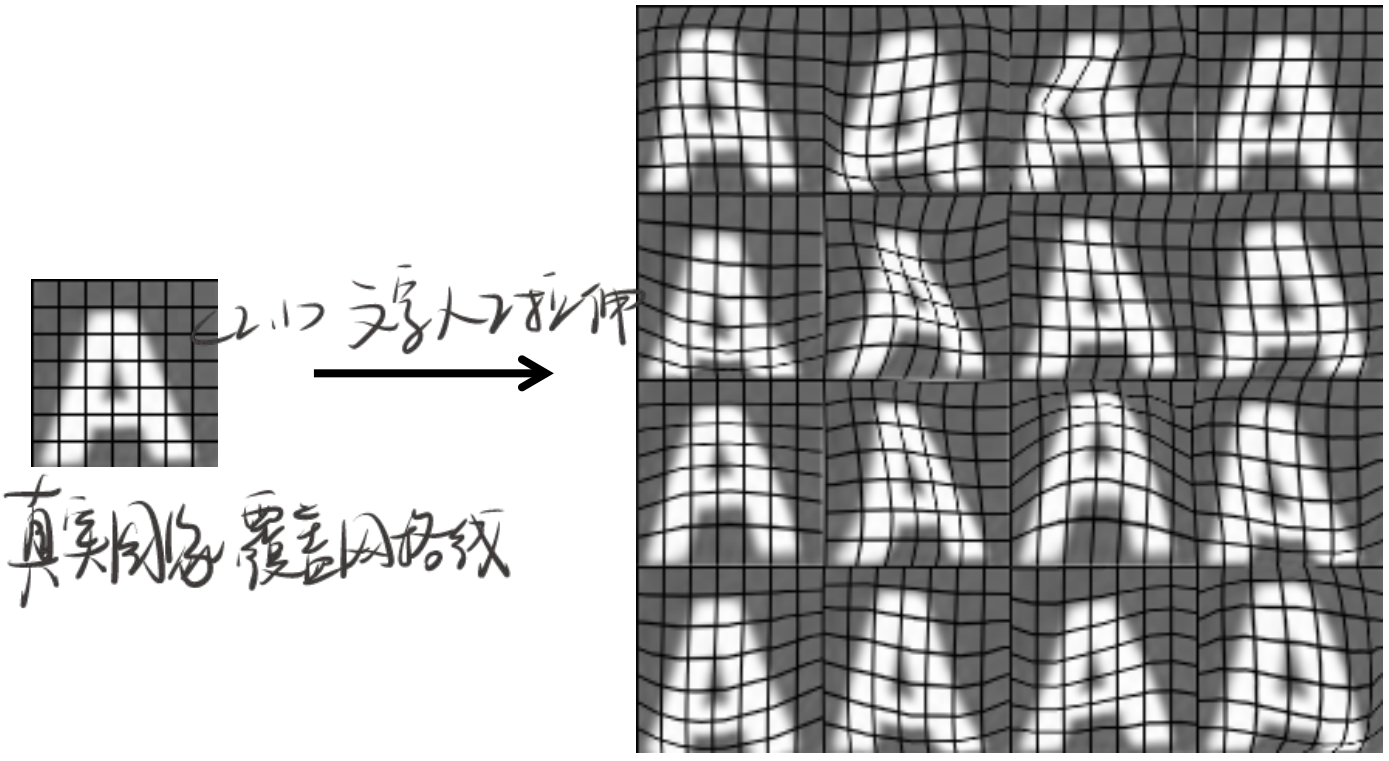
Artificial data synthesis for photo OCR

Real data

Synthetic data

[Adam Coates and Tao Wang]

Synthesizing data by introducing distortions



[Adam Coates and Tao Wang]

Synthesizing data by introducing distortions: Speech recognition

Original audio:

Audio on bad cellphone connection

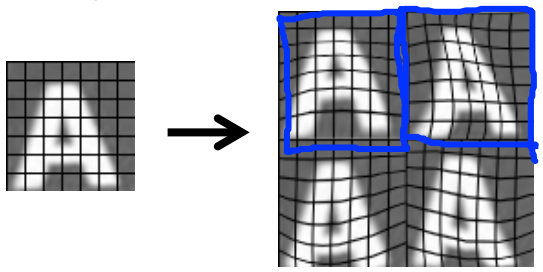
Noisy background: Crowd

Noisy background: Machinery

[www.pdsounds.org]

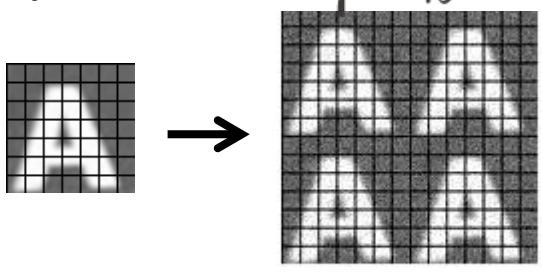
Synthesizing data by introducing distortions

- Distortion introduced should be representation of the type of noise/distortions in the test set. 引入的失真应有代表性



- Audio:
  Background noise,
  bad cellphone connection

- Usually does not help to add purely random/meaningless noise to your data. 引入无意义的失真(非像素级噪音)

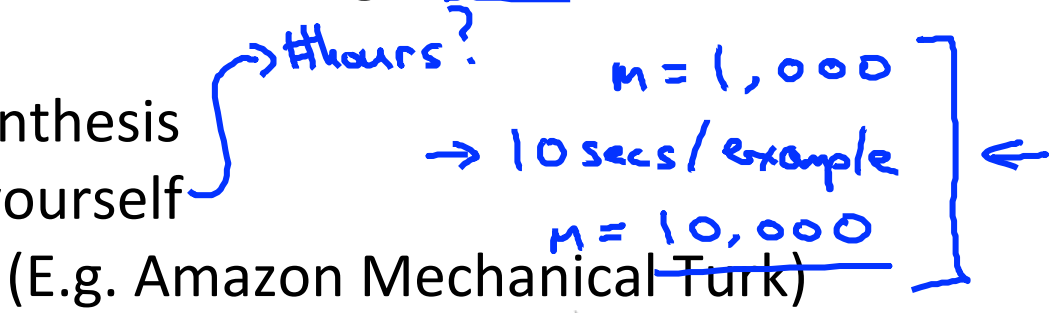


- x_i = intensity (brightness) of pixel i
- $x_i \leftarrow x_i +$ random noise

[Adam Coates and Tao Wang]

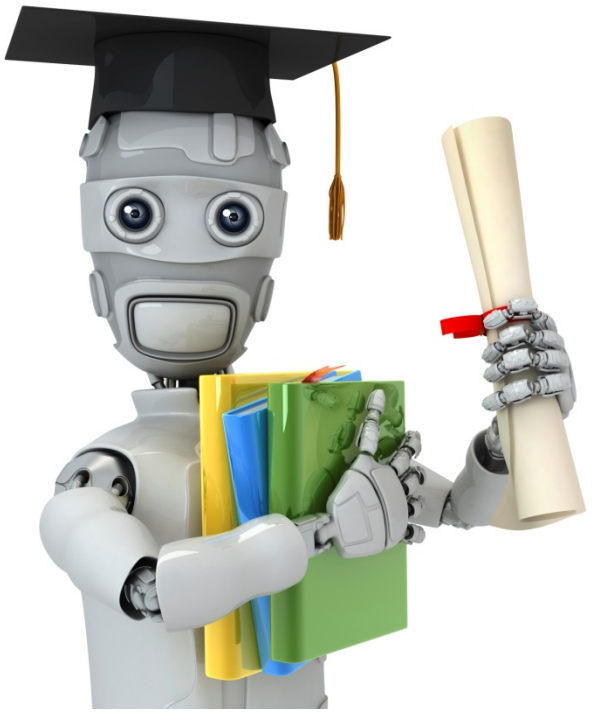
Discussion on getting more data

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)



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)

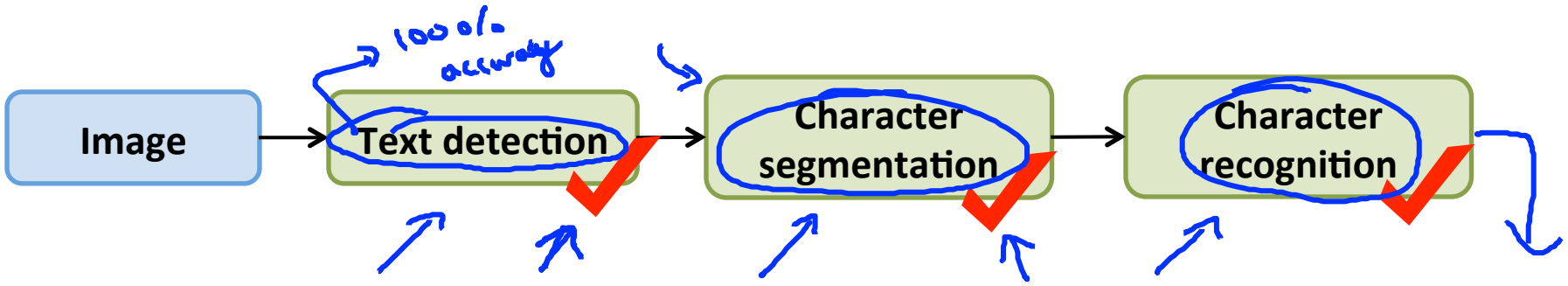


Machine Learning

Application example: Photo OCR

Ceiling analysis: What part of the pipeline to work on next

Estimating the errors due to each component (ceiling analysis)



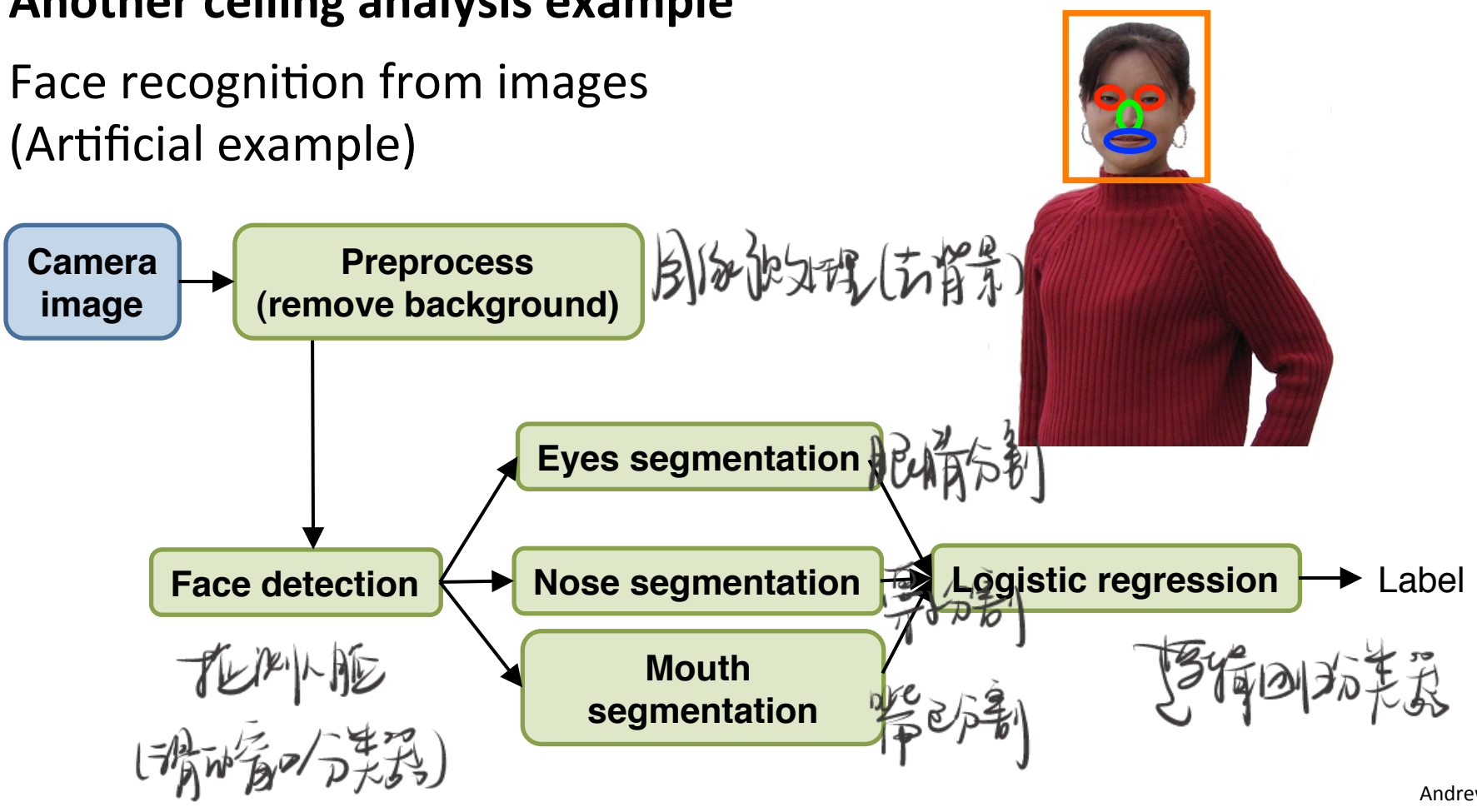
What part of the pipeline should you spend the most time trying to improve?

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

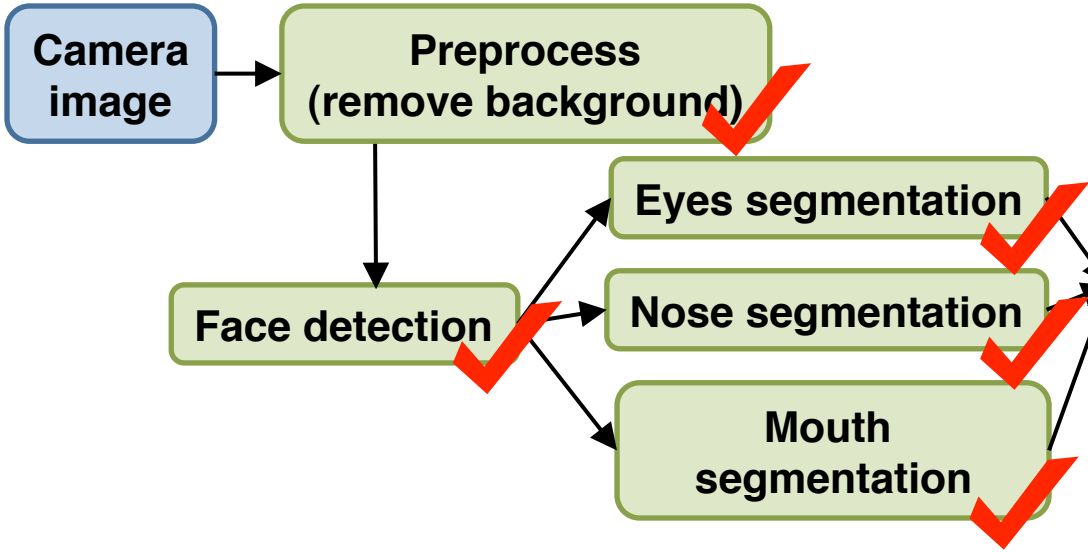
Andrew Ng

Another ceiling analysis example

Face recognition from images
(Artificial example)



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%
Logistic regression	100%

0.1%
5.9%
4%
1%
1%
3%

Summary: Main topics

→ Supervised Learning $(x^{(i)}, y^{(i)})$
- Linear regression, logistic regression, neural networks, SVMs

→ Unsupervised Learning $x^{(i)}$
- K-means, PCA, Anomaly detection

→ Special applications/special topics
- Recommender systems, large scale machine learning.

→ Advice on building a machine learning system
- Bias/variance, regularization; deciding what to work on next: evaluation of learning algorithms, learning curves, error analysis, ceiling analysis.