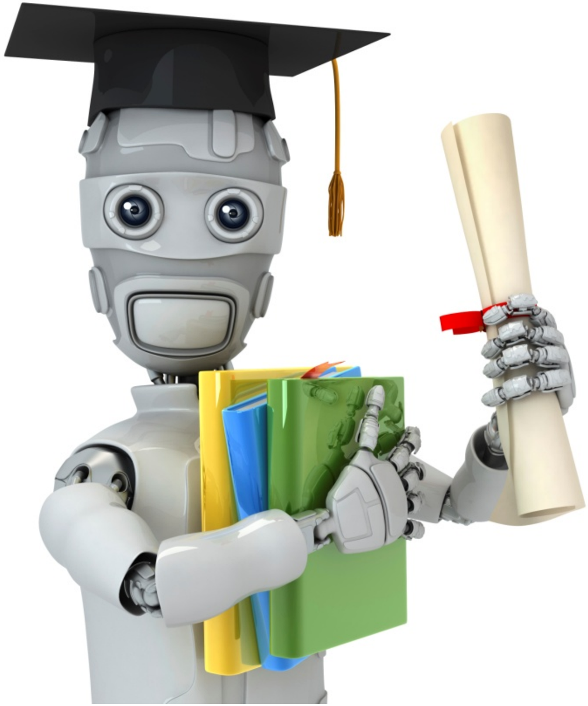


Machine Learning

Application example: Photo OCR

Problem description and pipeline

Photo optical character recognition

The Photo OCR problem

how to get the computer/camera to read the text/pictures better

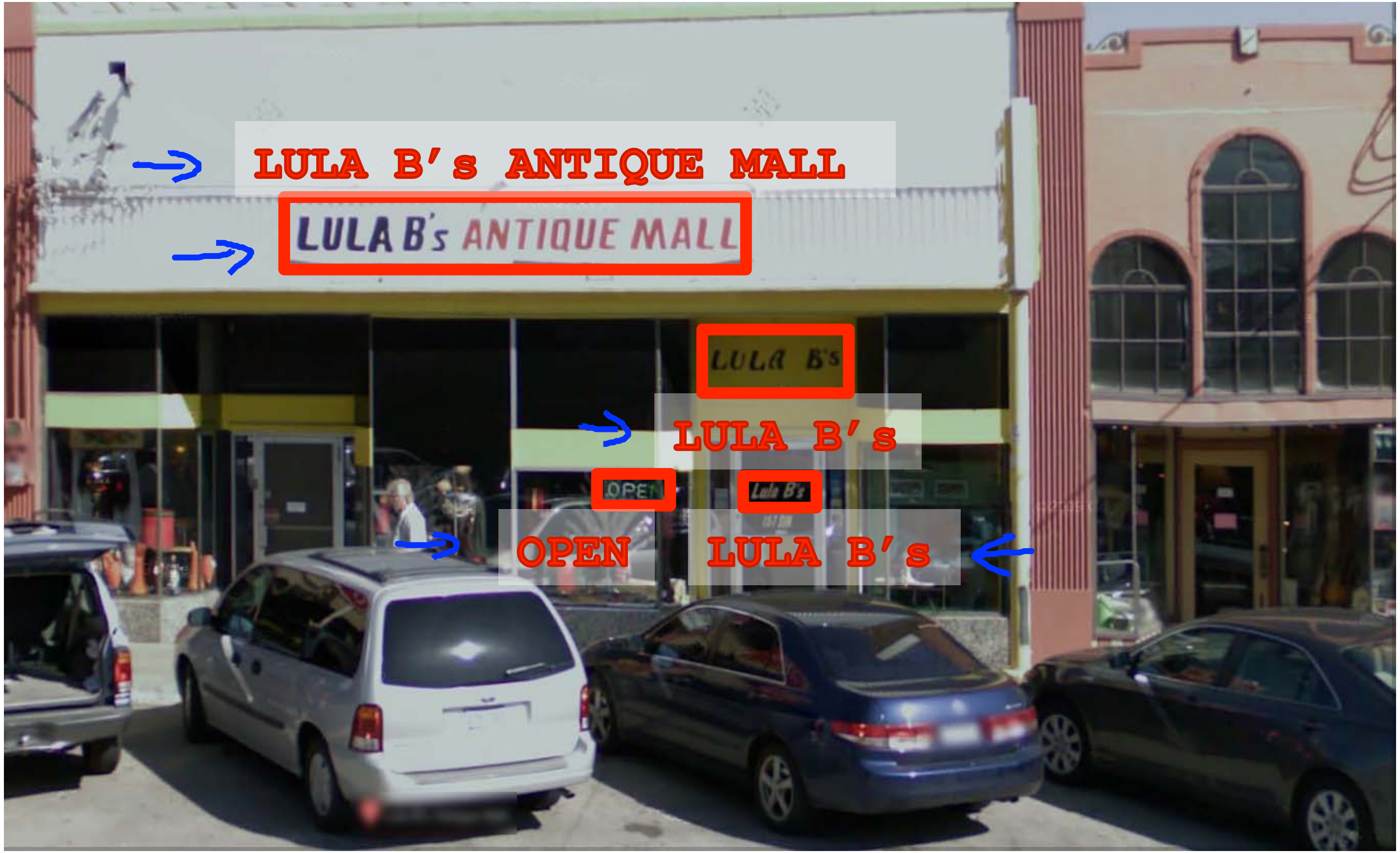
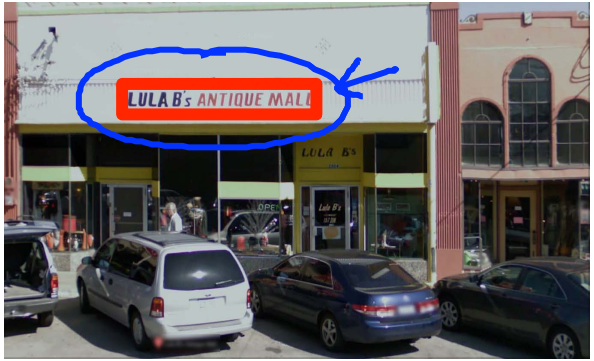
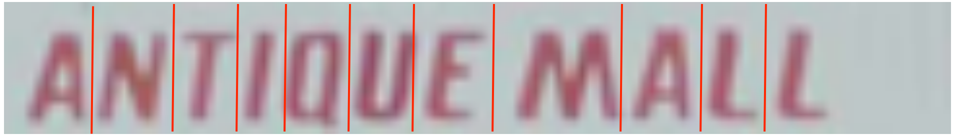


Photo OCR pipeline

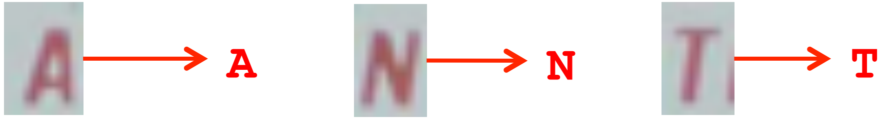
→ 1. Text detection



→ 2. Character segmentation



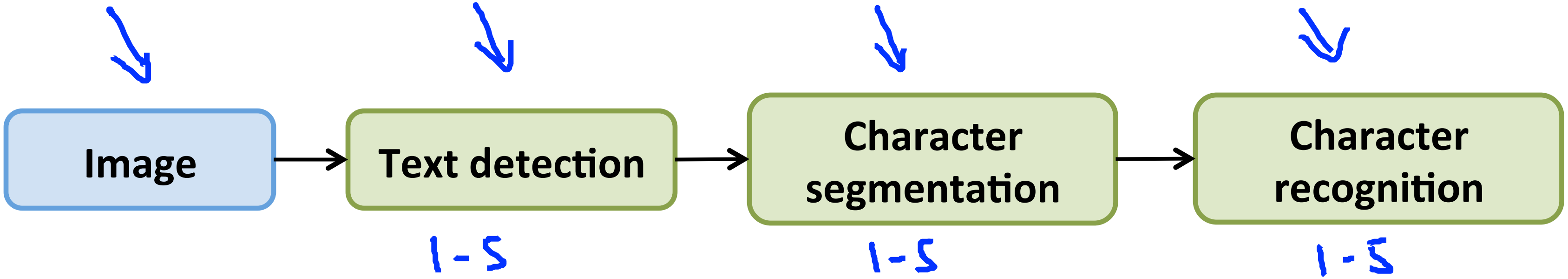
→ 3. Character classification

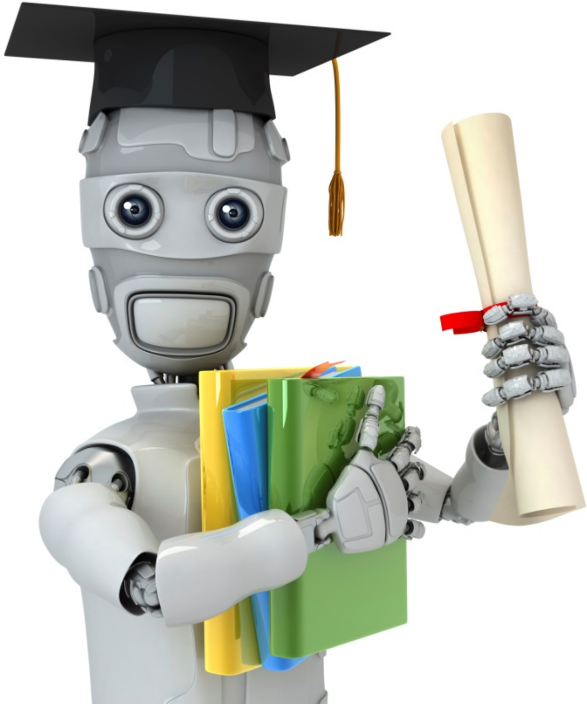


C1eaning → Cleaning

Photo OCR pipeline

each of this module can be a machine learning problem



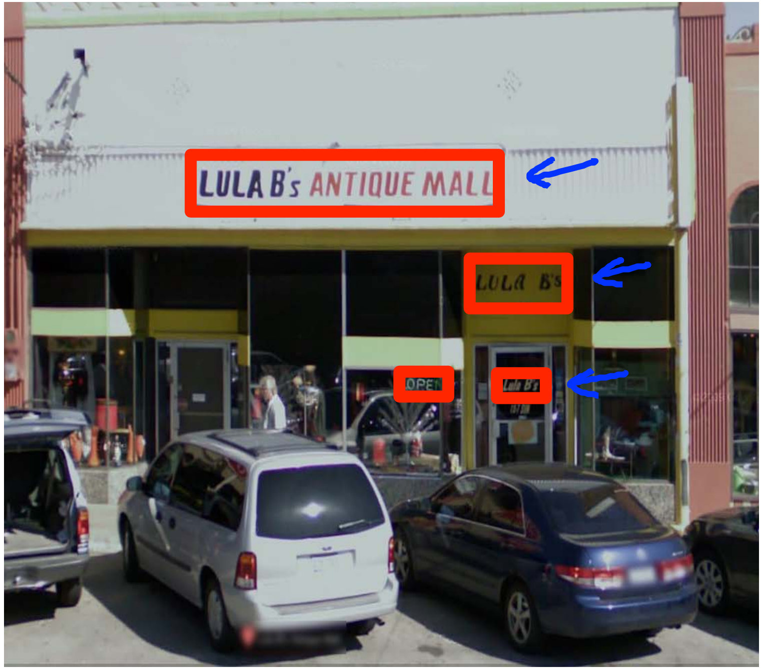


Machine Learning

Application example: Photo OCR

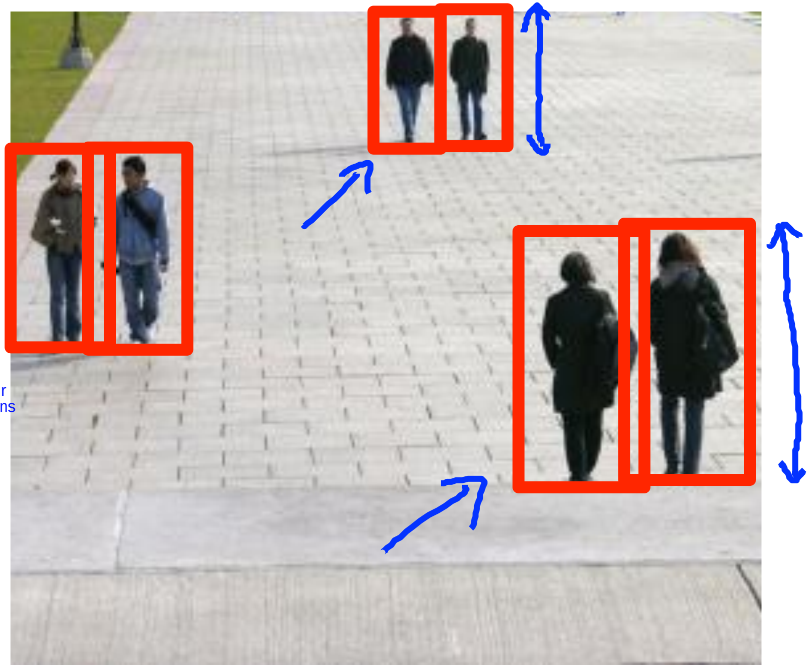
Sliding windows

Text detection



A simpler case

Pedestrian detection

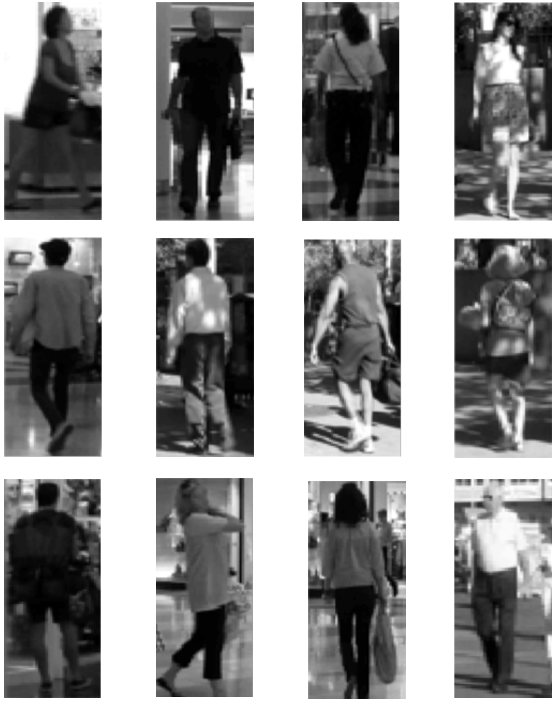


aspect ratio
is similar for
all pedestrians

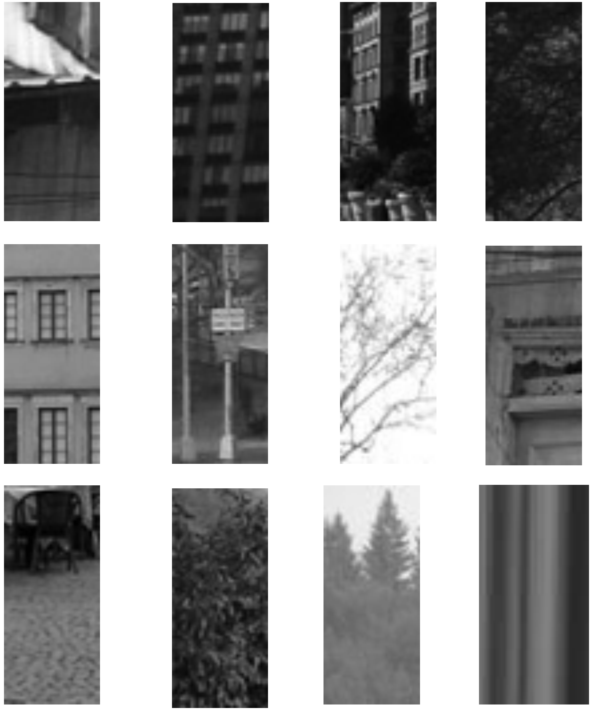
Supervised learning for pedestrian detection

1,000
10,000
⋮

$x =$ pixels in 82x36 image patches



Positive examples $(\underline{y =} 1)$



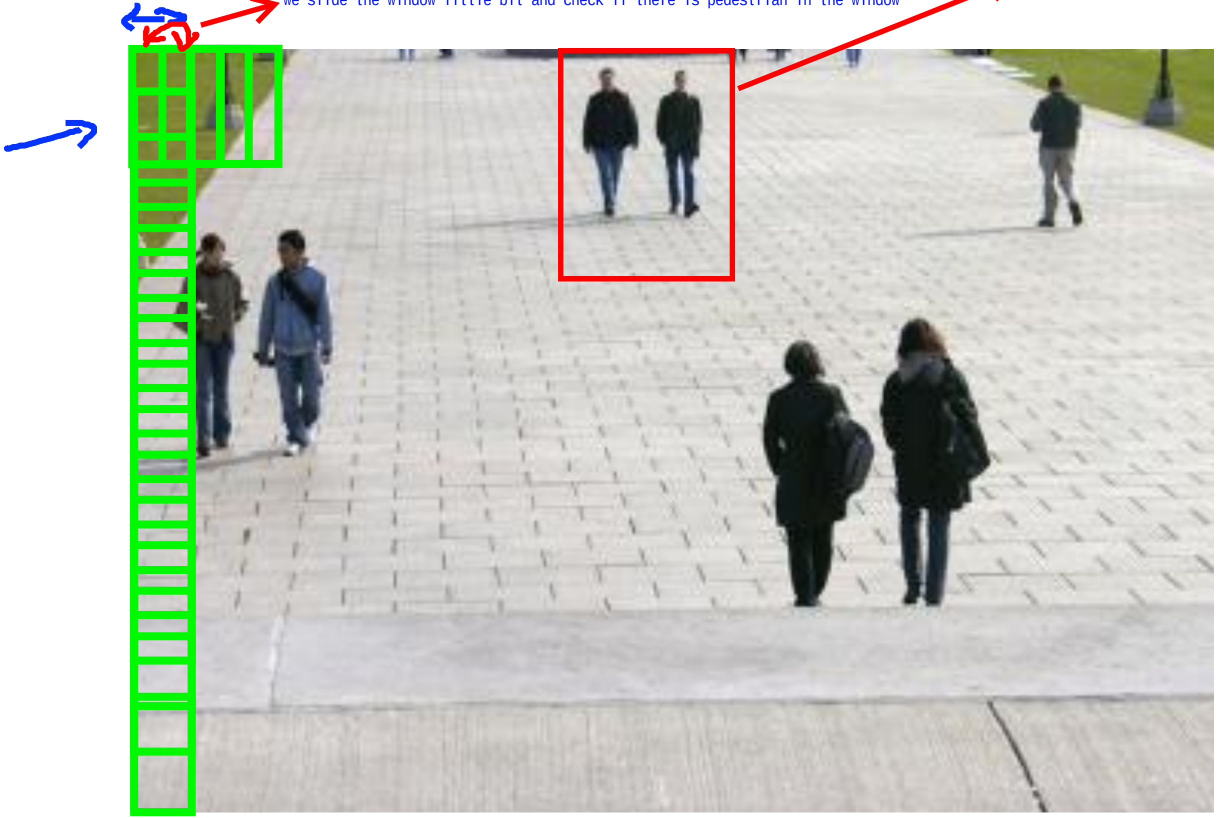
Negative examples $(\underline{y = 0})$

Andrew Ng

We can use supervised learning algorithm

Sliding window detection

step-size /stride

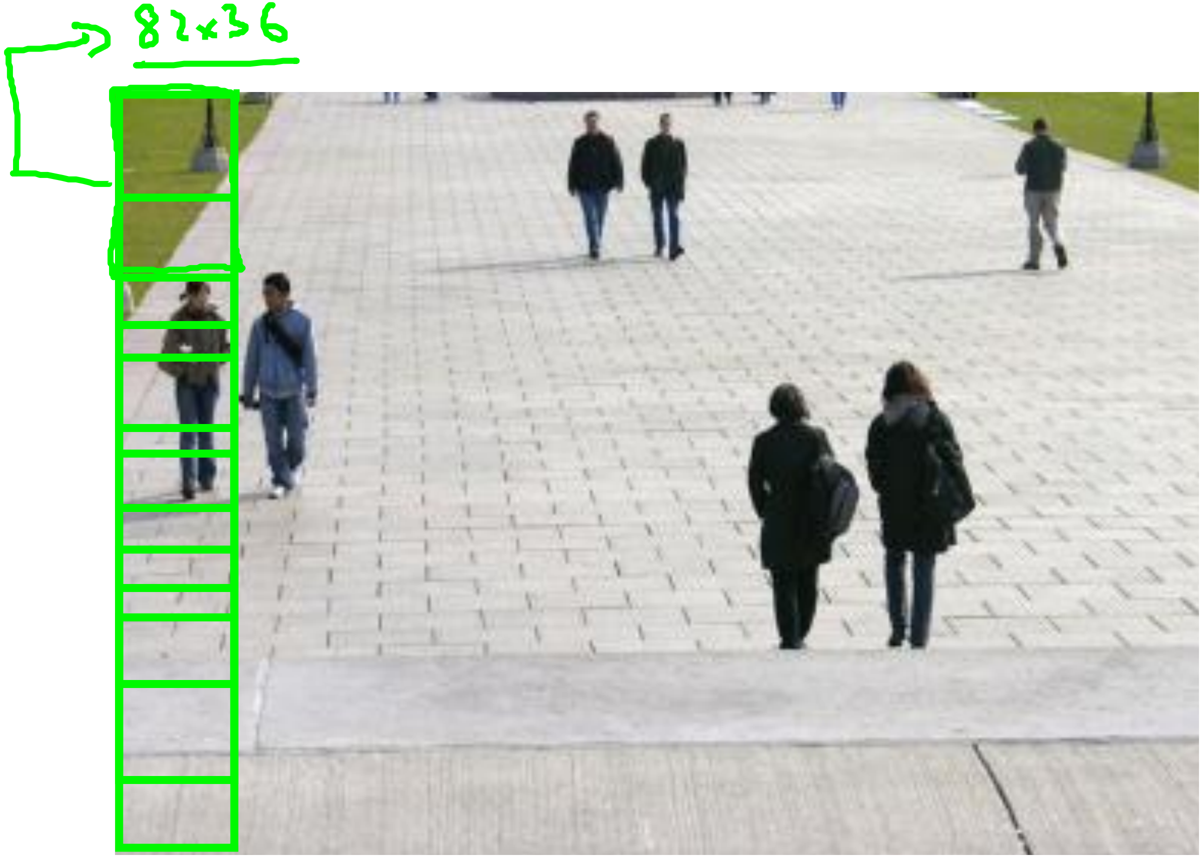


we slide the window little bit and check if there is pedestrian in the window

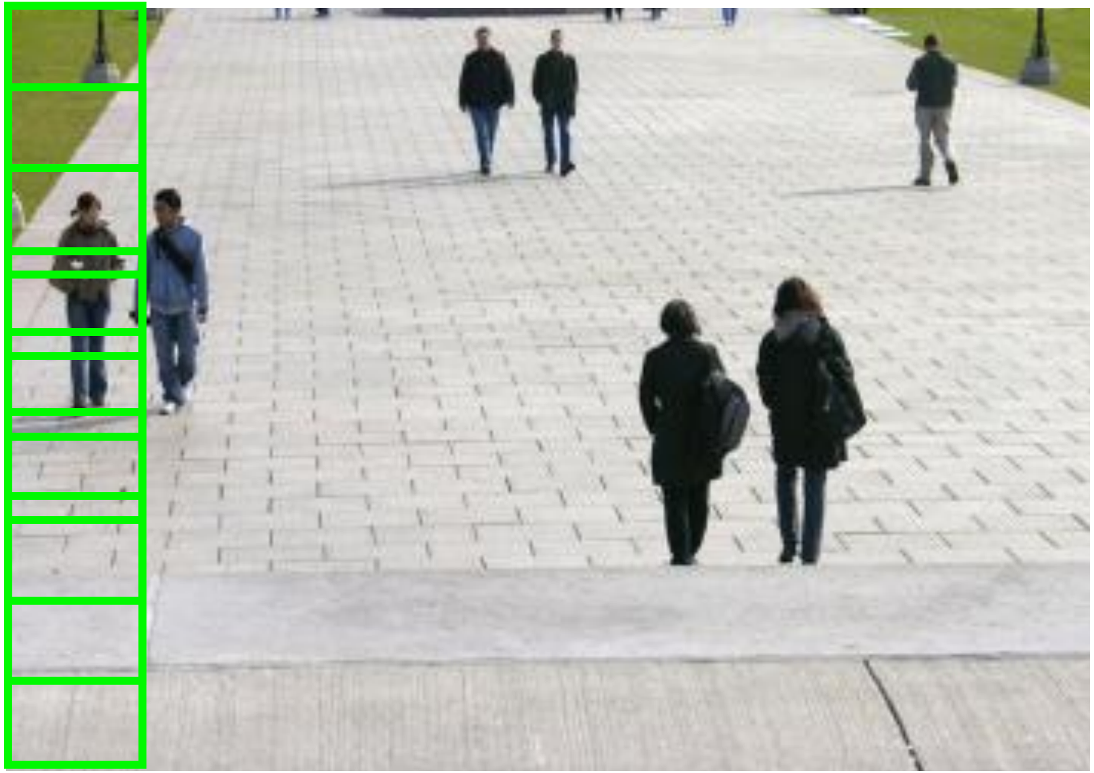
large image patch

Andrew Ng

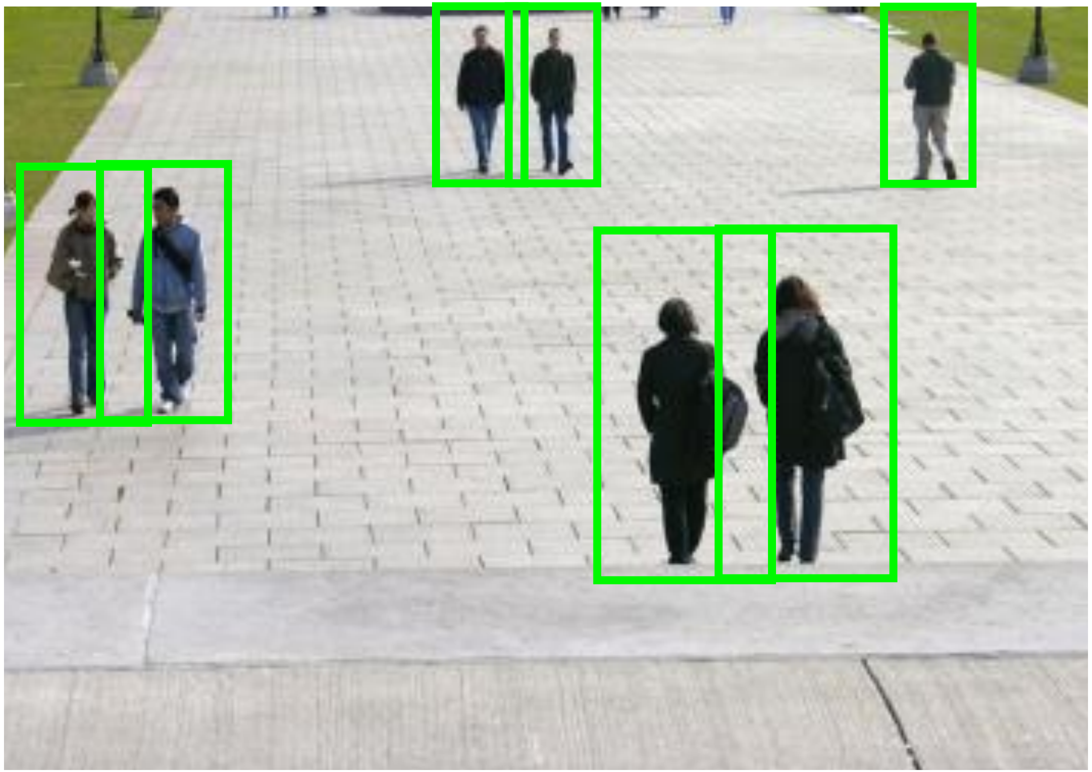
Sliding window detection



Sliding window detection

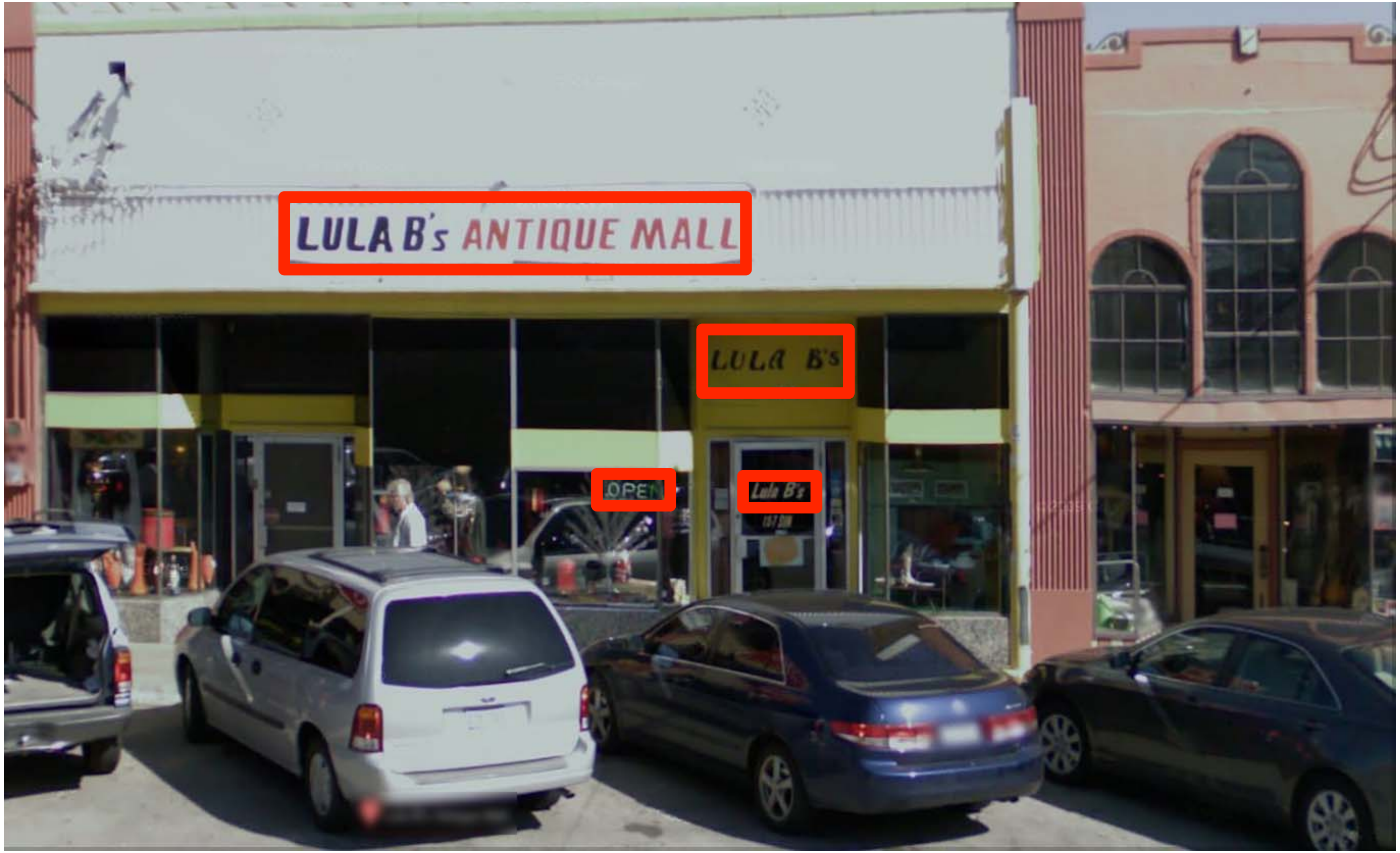


Sliding window detection



Andrew Ng

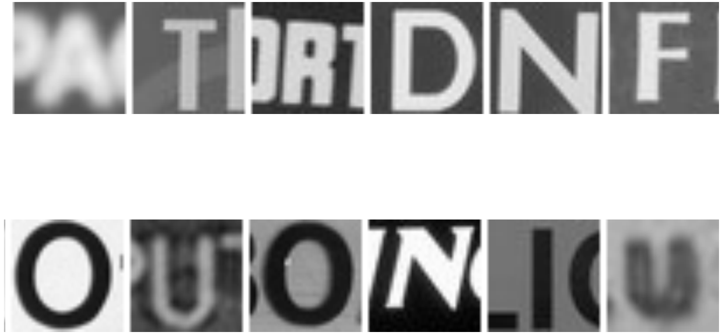
Text detection



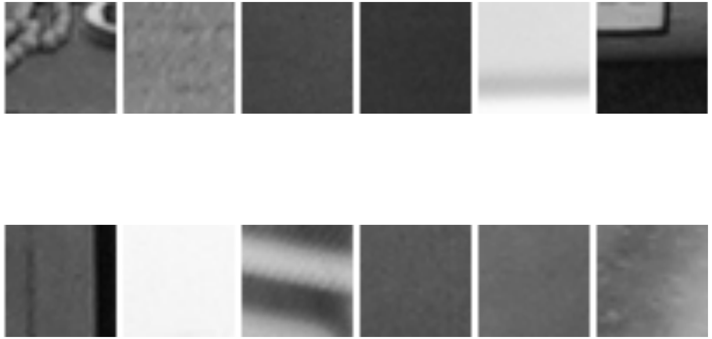
Andrew Ng

Now instead of detecting pedestrians, we try to detect text

Text detection

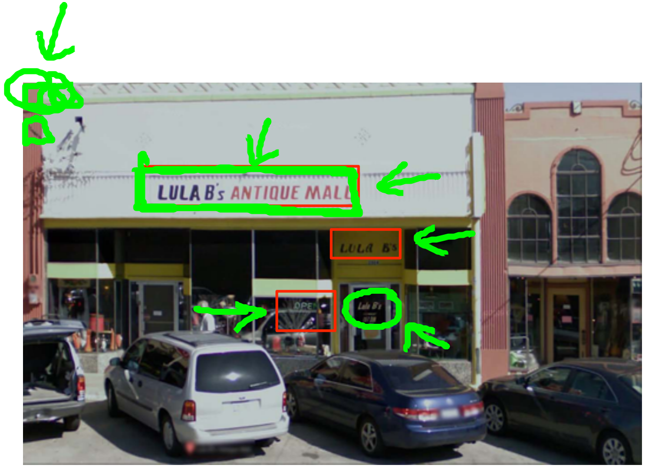


Positive examples $\underline{(y=1)}$

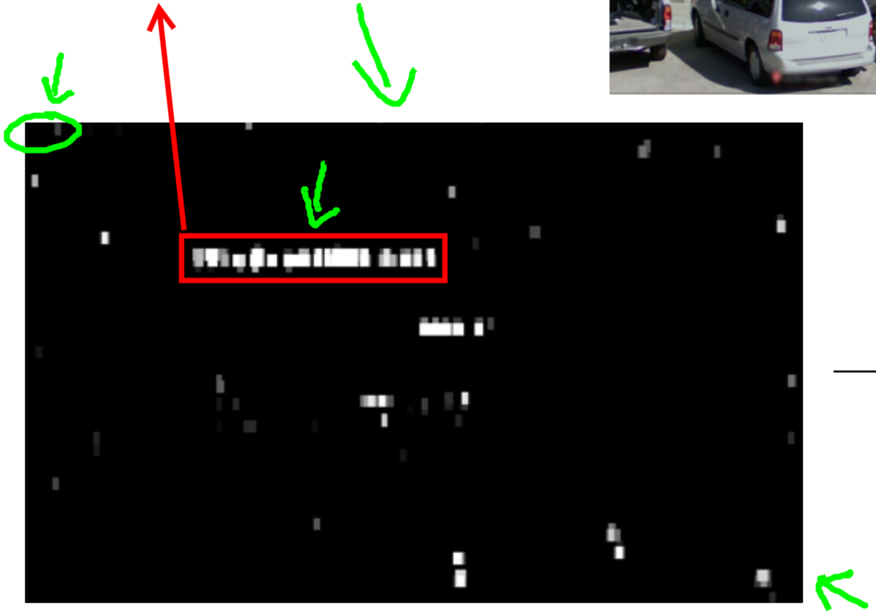


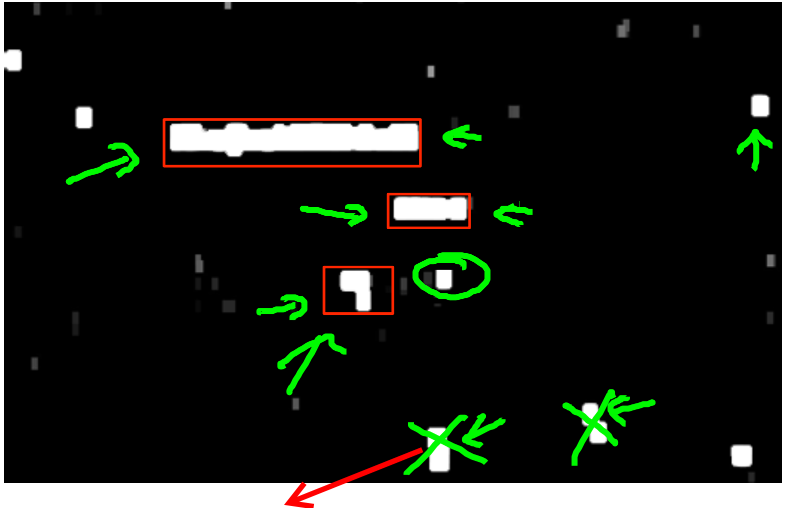
Negative examples $\underline{(y=0)}$

Text detection



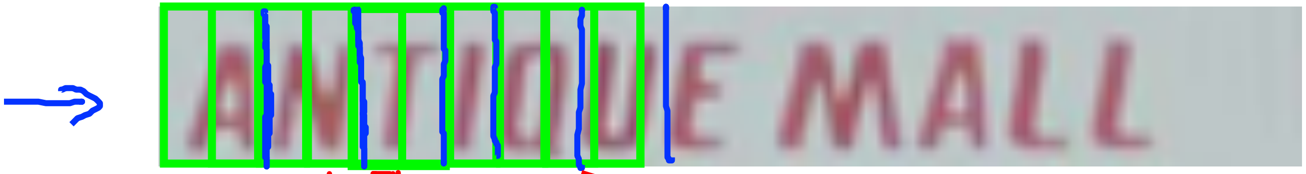
"expansion"


text region


aspect ratio looks funny for text

[David Wu]

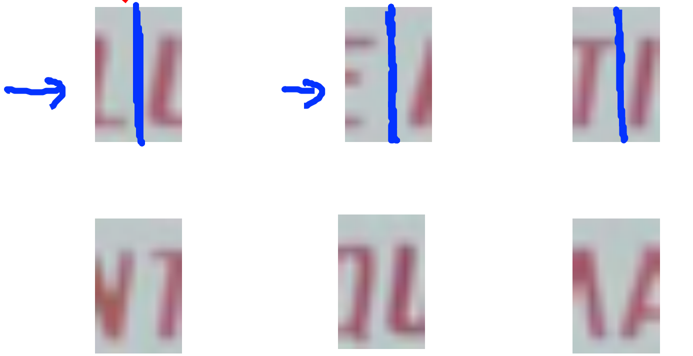
1D Sliding window for character segmentation



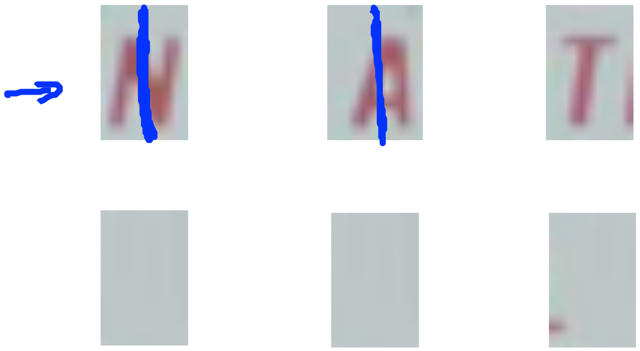
Then we use sliding window to check if there is a split between two characters

slide the window over

find this split



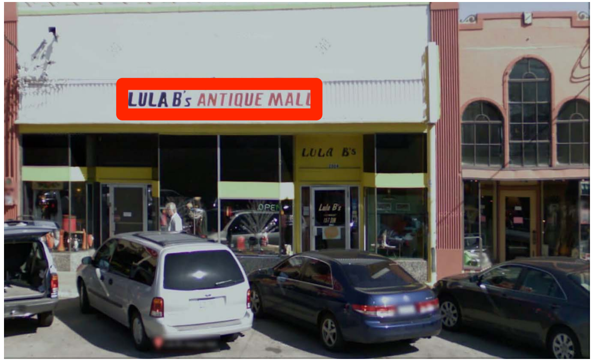
Positive examples $(y = 1)$



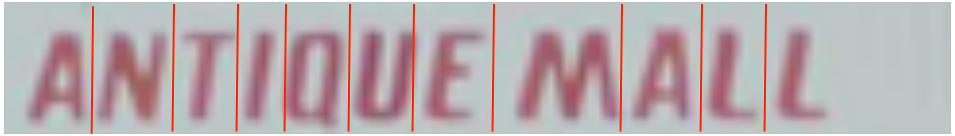
Negative examples $(y = 0)$

Photo OCR pipeline

1. Text detection

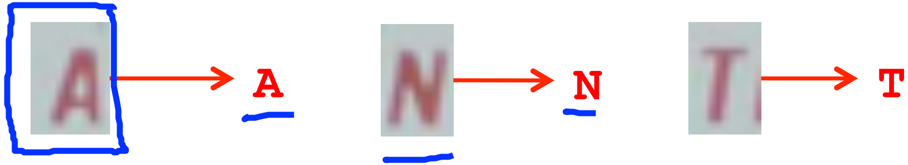


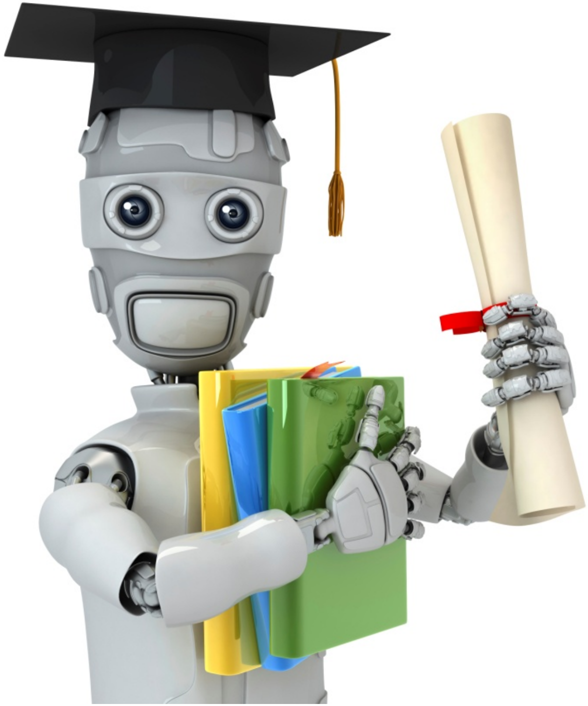
2. Character segmentation



3. Character classification

just apply some supervised learning algorithm





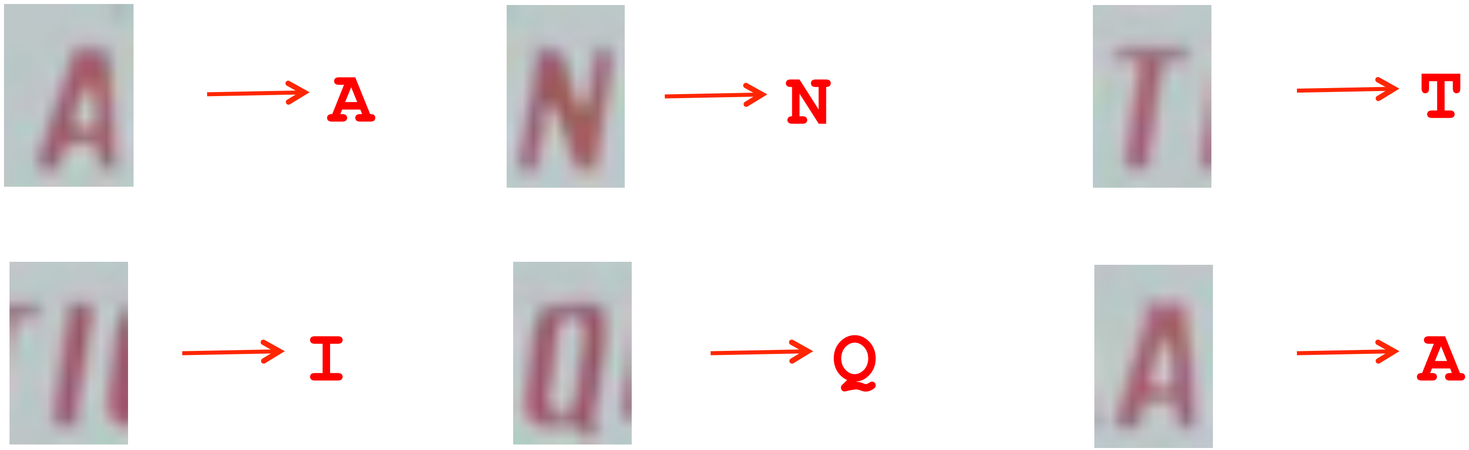
Machine Learning

Application example: Photo OCR

Getting lots of data: Artificial data synthesis

two main variations

Character recognition



→ A

→ N

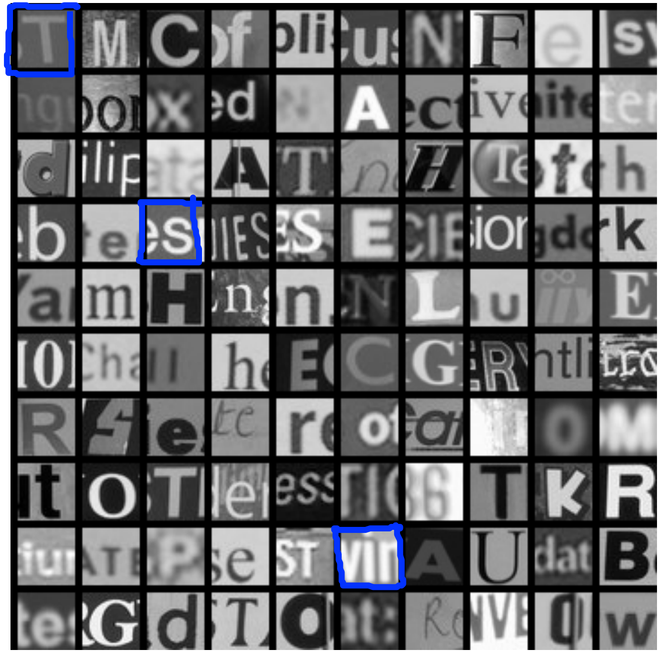
→ T

→ I

→ Q

→ A

Artificial data synthesis for photo OCR



Real data

Abcdefg

Abcdefg

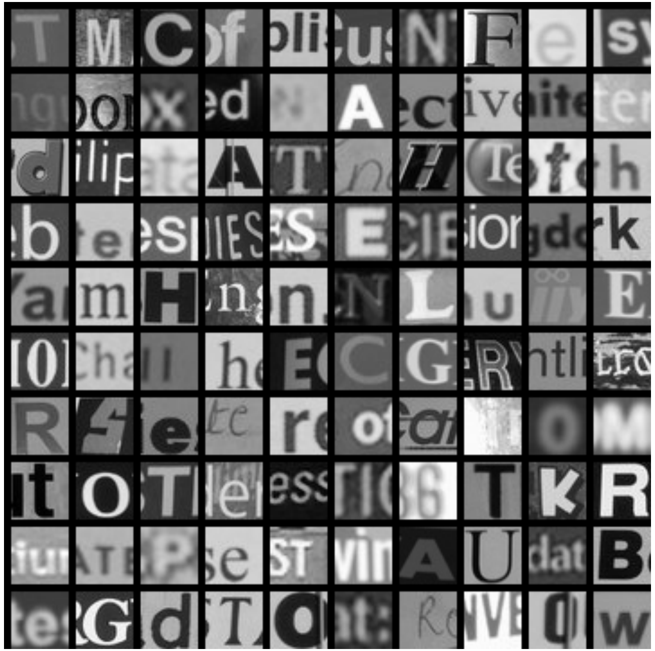
Abcdefg

Abcdefg

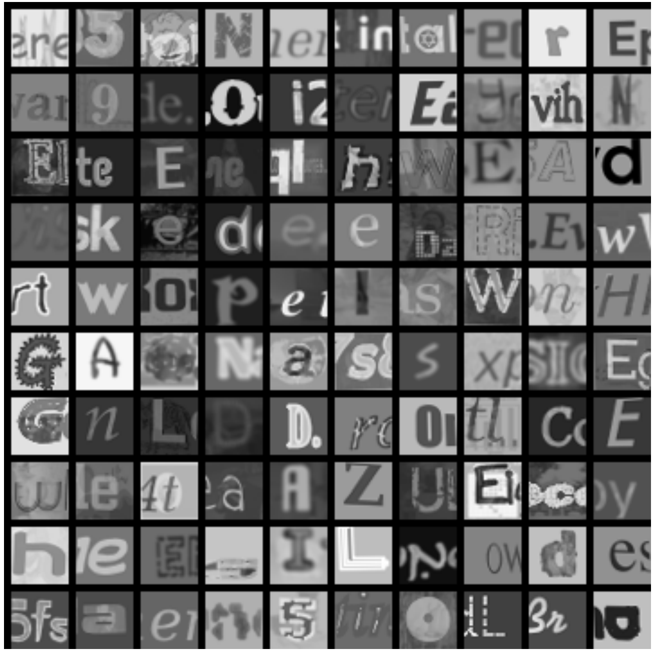
Abcdefg

[Adam Coates and Tao Wang]

Artificial data synthesis for photo OCR



Real data

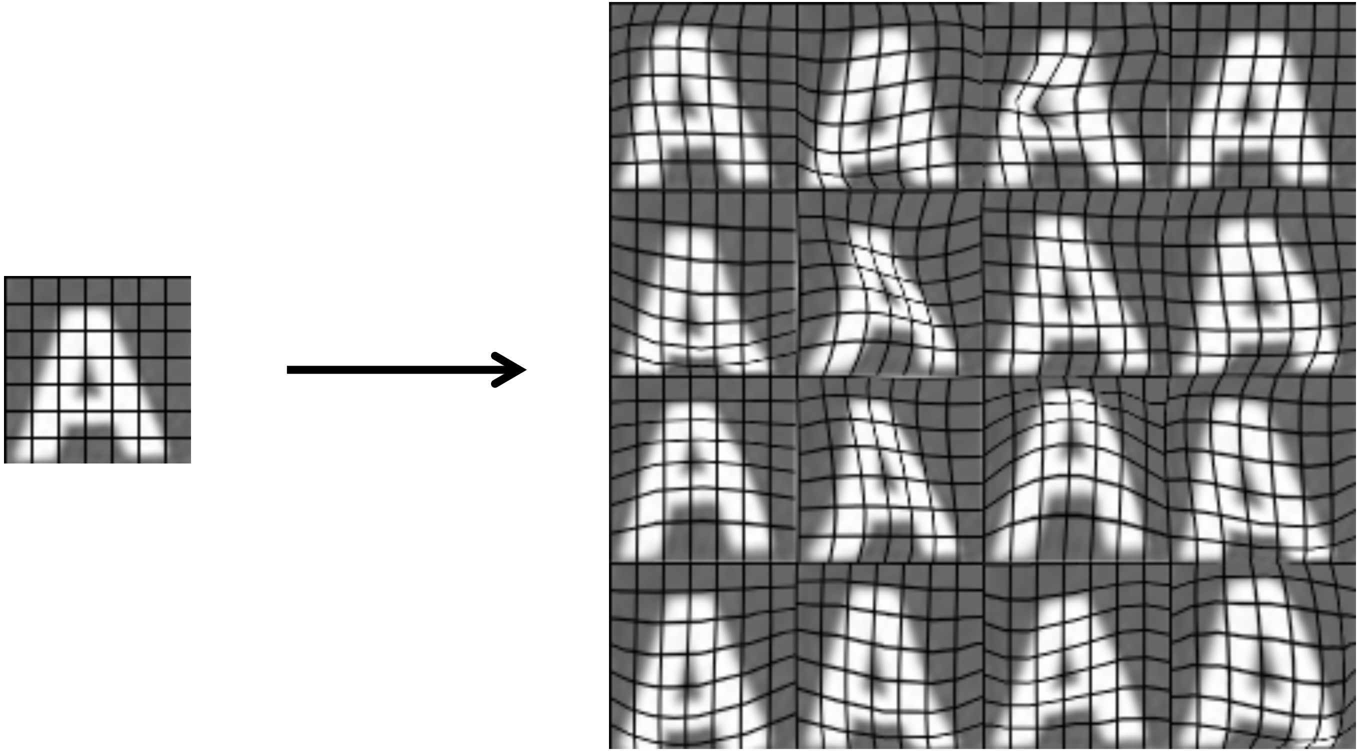


Synthetic data

Creating new data from scratch

[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

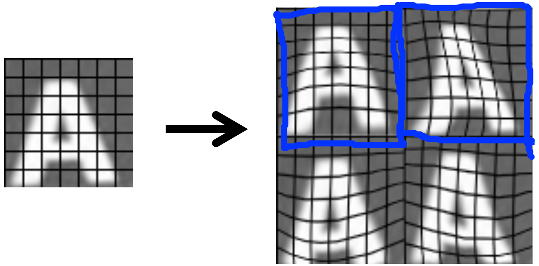
🔊 Noisy background: Crowd

🔊 Noisy background: Machinery

[www.pdsounds.org]

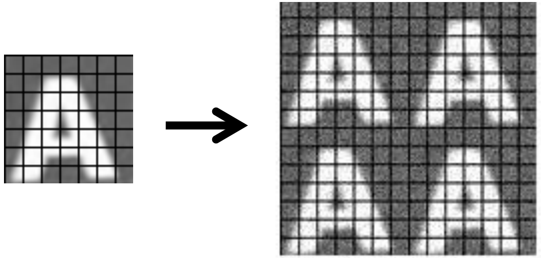
Synthesizing data by introducing distortions

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



- Audio:
  Background noise,
  bad cellphone connection

- Usually does not help to add purely random/meaningless noise to your data.



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

meaningless noise is less meaningful

[Adam Coates and Tao Wang]

Andrew Ng

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?”
   ?
   → #hours?
   - Artificial data synthesis
   - Collect/label it yourself
   - “Crowd source” (E.g. Amazon Mechanical Turk)

m = 1,000

→ 10 secs / example ←

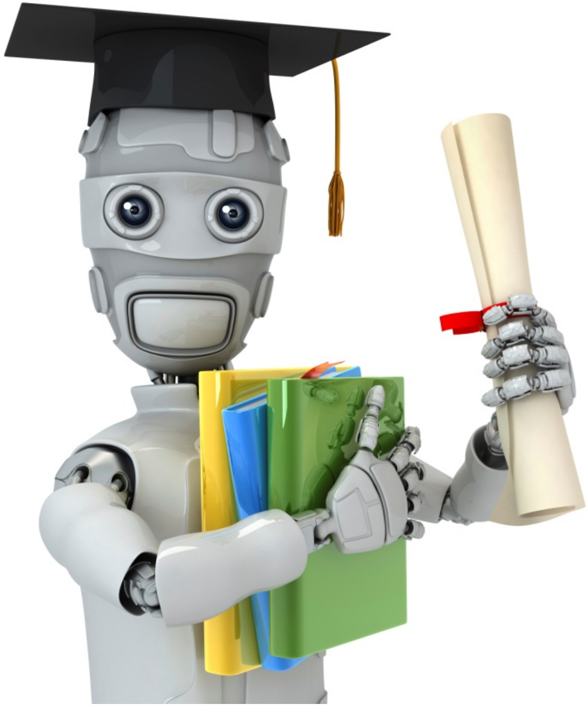
m = 10,000

May be the most popular source

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)

众包

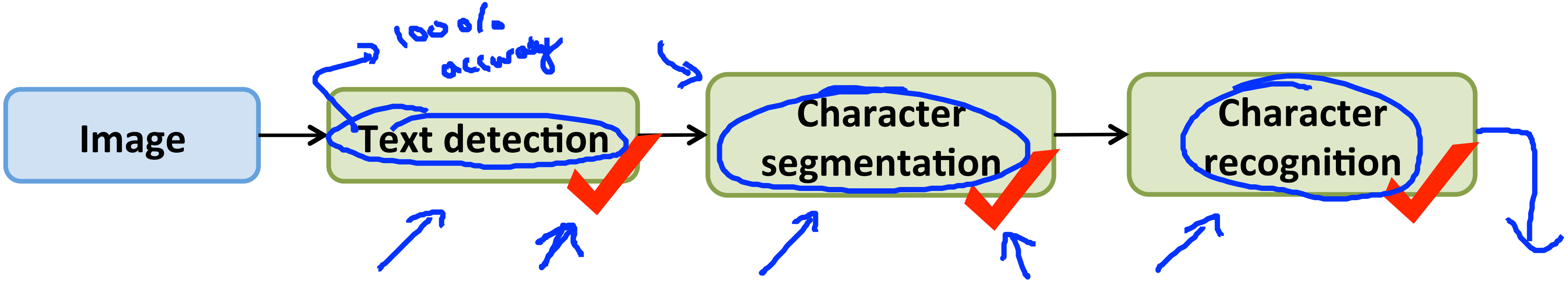


Machine Learning

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 manually set that all text are correctly detected	89%
Character segmentation same idea	90%
Character recognition same	100%

17%

1%

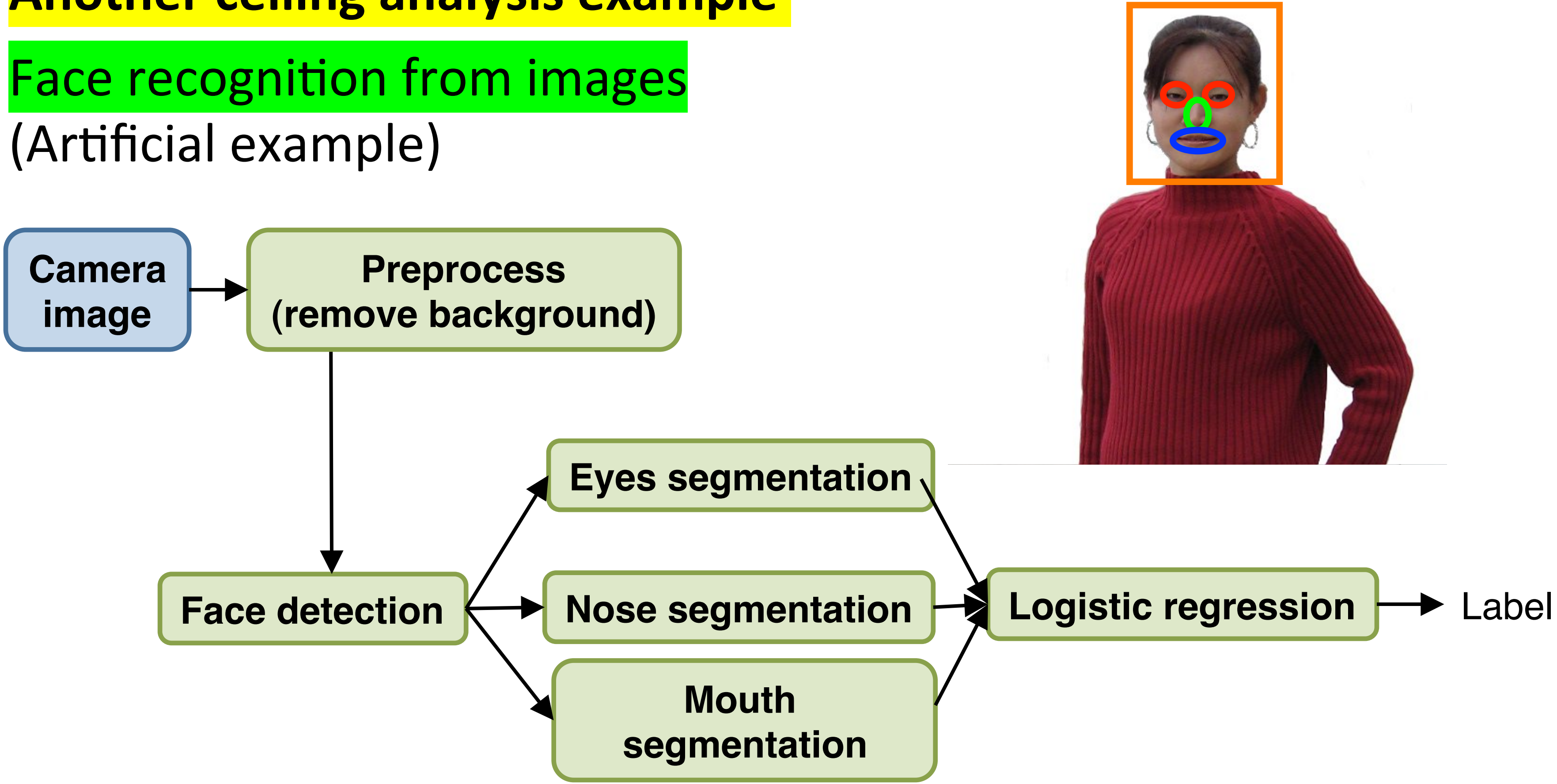
only 1% improvement (do not spend too much time here!)

10%

Another ceiling analysis example

Face recognition from images

(Artificial example)



Camera image

Preprocess (remove background)

Face detection

Eyes segmentation

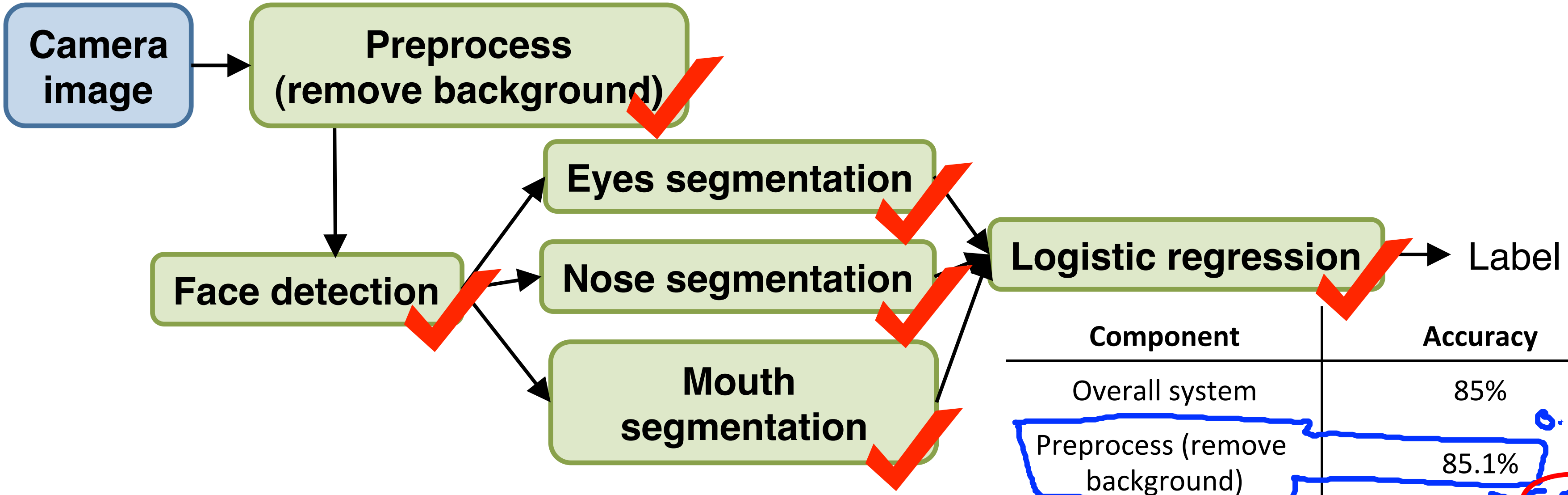
Nose segmentation

Mouth segmentation

Logistic regression

Label

Another ceiling analysis example



Component	Accuracy
Overall system	85%
	0.1%
Preprocess (remove background)	85.1%
	5.9%
Face detection	91%
	4%
Eyes segmentation	95%
	1%
Nose segmentation	96%
	1%
Mouth segmentation	97%
	3%
Logistic regression	100%