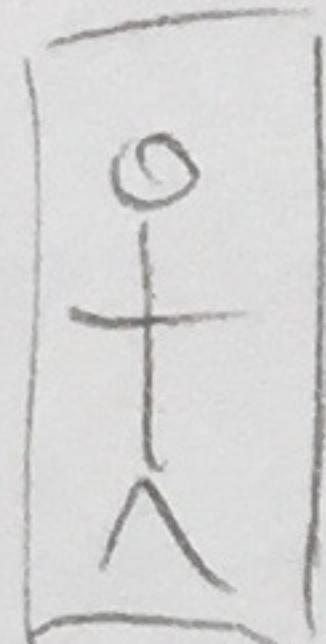


Application Example: OCR in Photos

<sup>image:</sup>  
 $\Rightarrow$  text detection  $\rightarrow$  character segmentation  $\rightarrow$  character classification

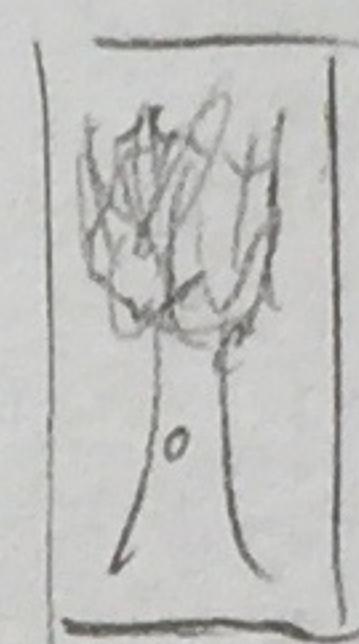
ML PIPELINE

$\hookrightarrow$  break algorithm down into actionable modules

Supervised Learning for Pedestrian Detection:

Fixed aspect ratio (82x36)

$y = 1$

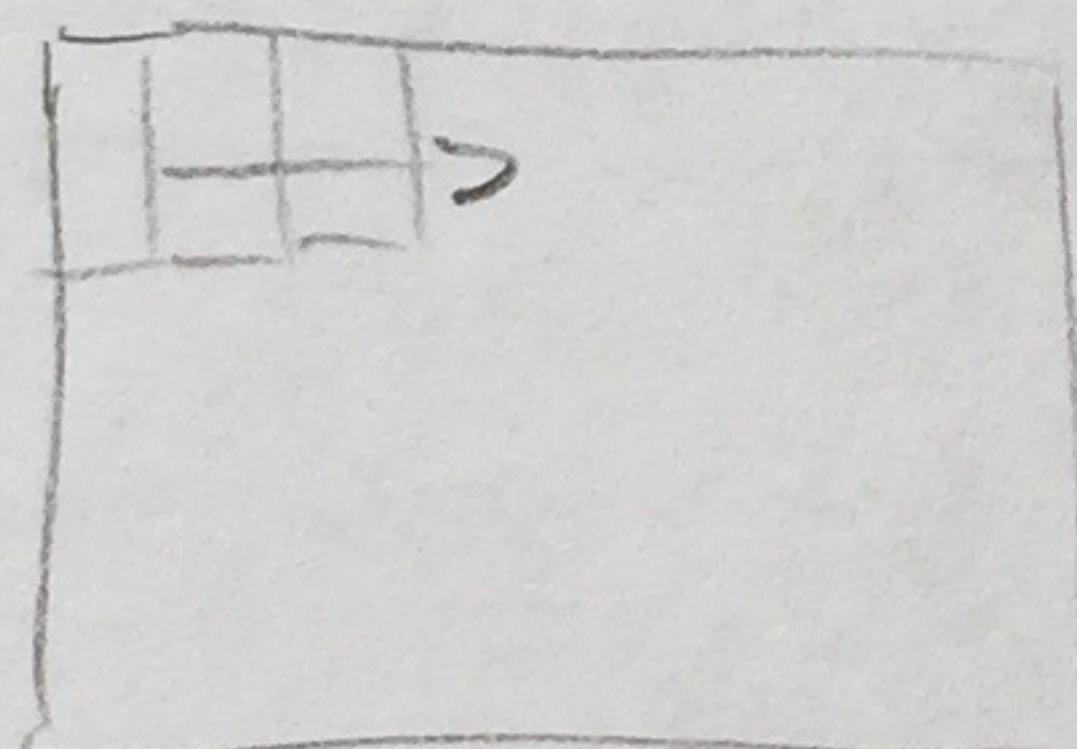


$y = 0$

$\geq c$  = pixels in image patches

step size

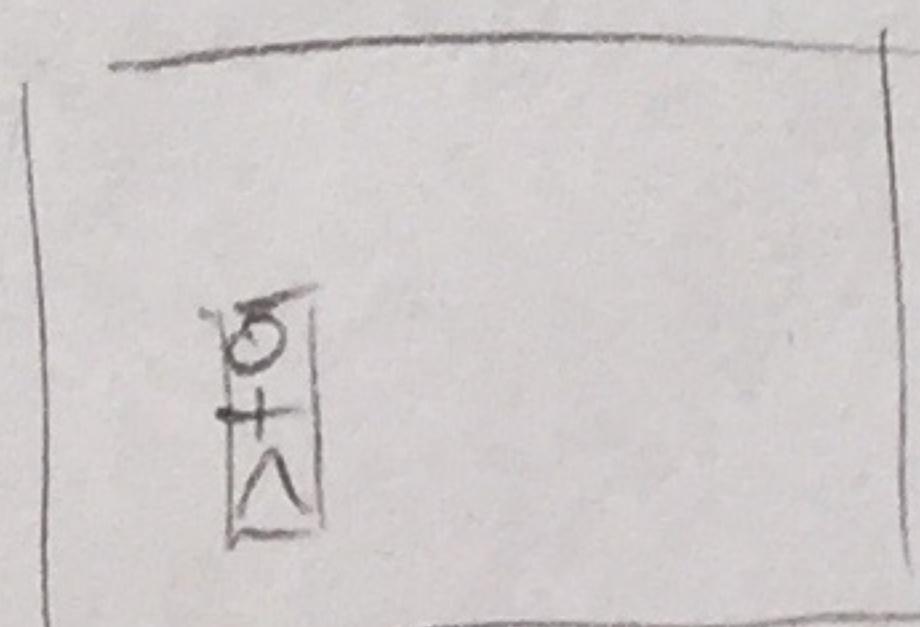
$\Rightarrow$  Sliding Window:



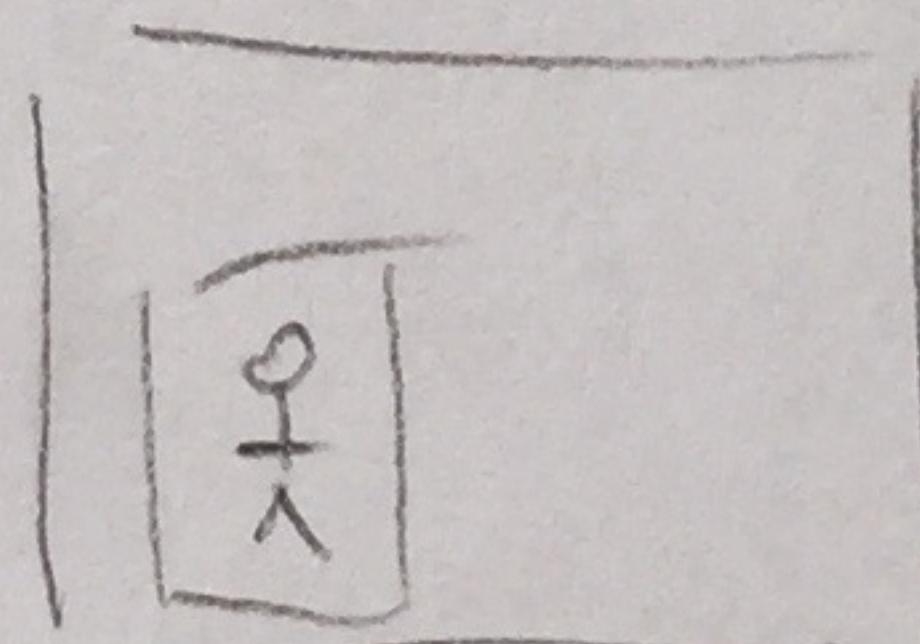
$\Rightarrow$  run each rectangle through trained classifier

$\hookrightarrow$  increase rectangle sliding size:  $\square \rightarrow \square \rightarrow \square$  (resized to 82x36),  
run each size over whole image

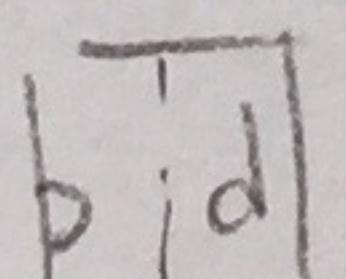
$\Rightarrow$  can do same thing for text!



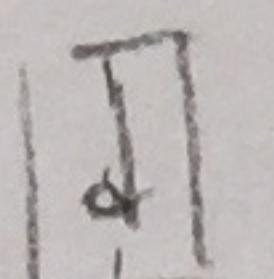
$\Rightarrow$  expansion  $\Rightarrow$



$\Rightarrow$  same thing for character segmentation:



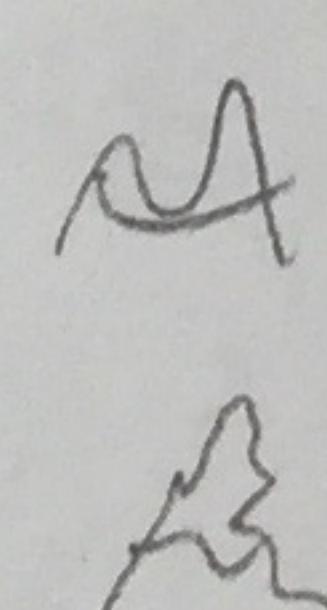
$y = 1$



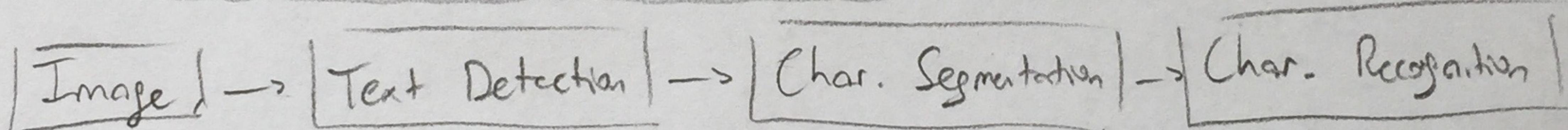
$y = 0$

$\Rightarrow$  run on HI WORLD  
A segment to individual  
characters

$$y = b - \alpha x$$

- ⇒ final step: individual character recognition module
- ⇒ Data acquisition:
  - real (pictures of text)
  - synthetic (from fonts)
    - ↳ take A & distort → A 
    - 4 different examples!
- ⇒ Note: distortions shouldn't be random noise
  - ↳ e.g. for audio, add machinery, background noise rather than synthetic noise
- ⇒ Recall, make sure low bias is present before getting more data ( $\uparrow$  features, hidden units)
- ⇒ Good question: how hard is it to get 10x the data we currently have?
  - ↳ artificial / synthesis, # hours of collection + labeling, crowd source it (e.g. Mechanical Turk)

### Ceiling Analysis: Prioritization of Pipeline Modules



- ⇒ Simulate the system accuracy if certain modules were 100% accurate (ground-truth accuracy)
  - ↳ examine accuracy delta's & prioritize accordingly → "How much could you gain if this module is perfect?"