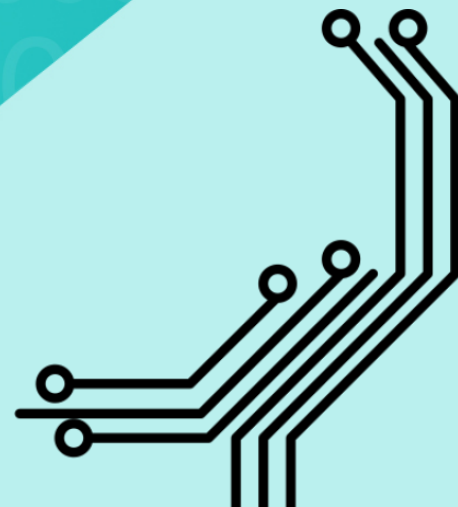


Cognitive Robotics

Dylan Vanmali



Problem Statement

"Speak what you seek until you see what you've said!"



How does the brain interpret textual images into sentences?

Why is Text Detection Challenging?



Flat



Angled

Black on White

White on Black

Font v1

Font v2

Font v3

Font v4

Font v5

Font v6

Spacing

Spacing

Spacing

Spacing Words

S p a c i n g W o r d s

S p a c i n g W o r d s

Real World



Text Detection Approach

Prepare: CV2
BlobFromImage



Forward through
Trained EAST



Segment Plausible
Text Zones Into
Sub-images



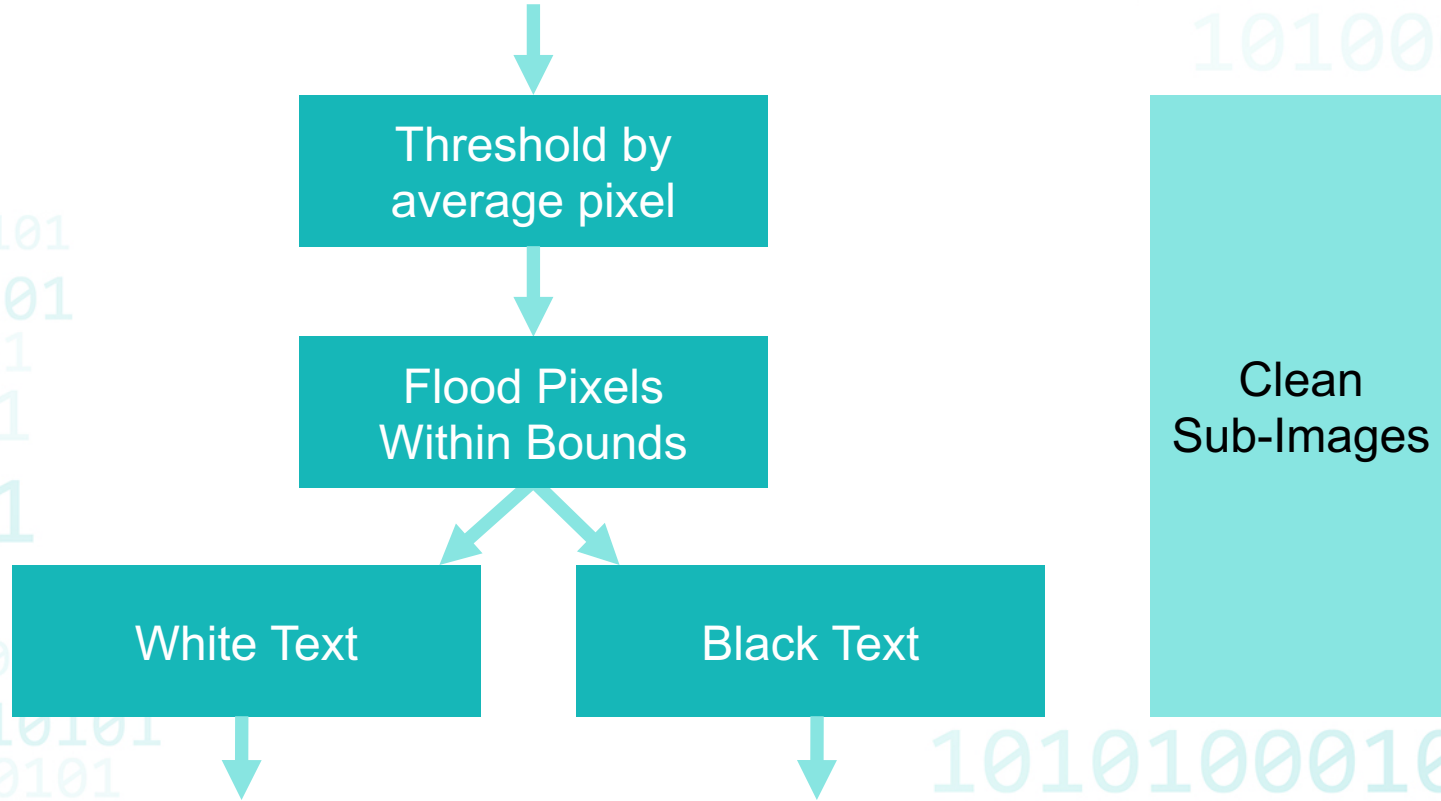
Box Points

Scale

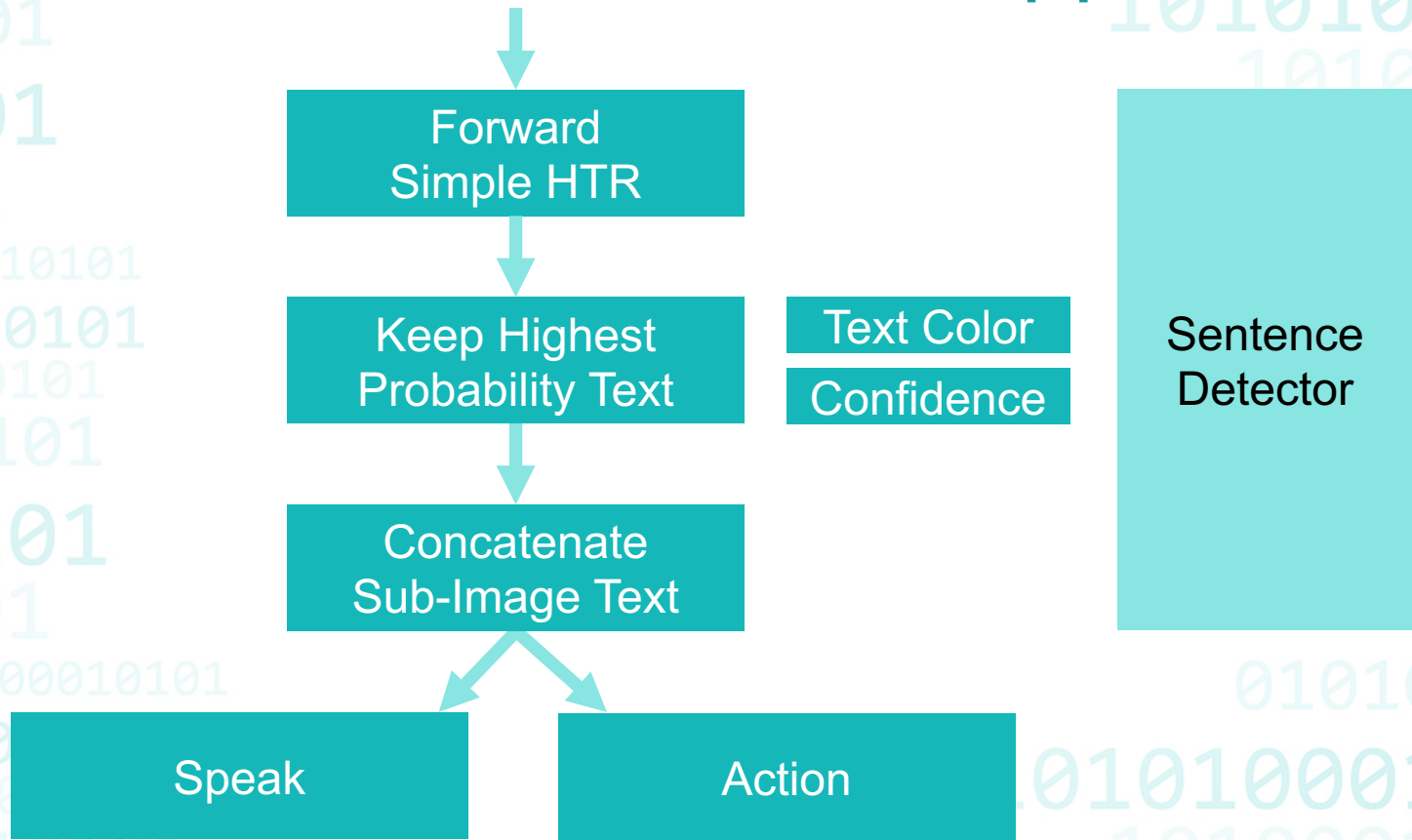
Transform

Efficient and
Accurate
Scene Text
Detection

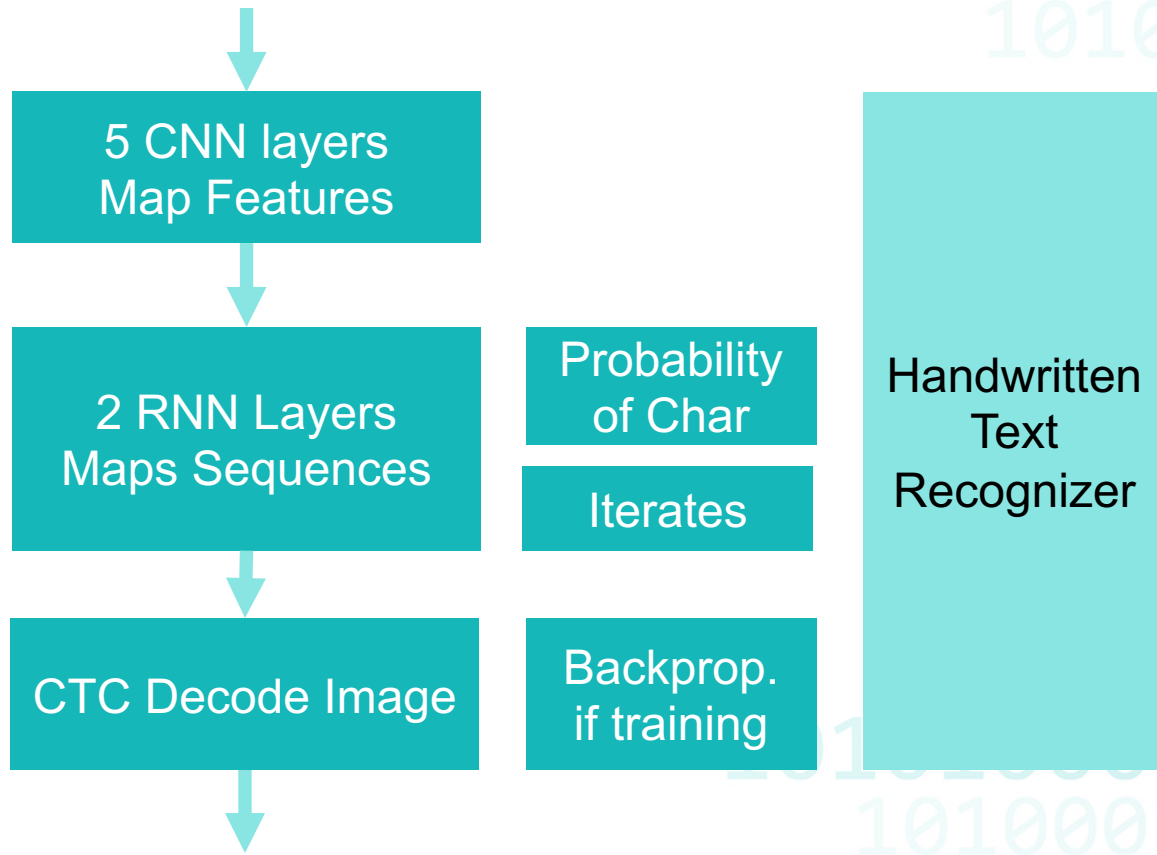
Text Interpreter Preprocess



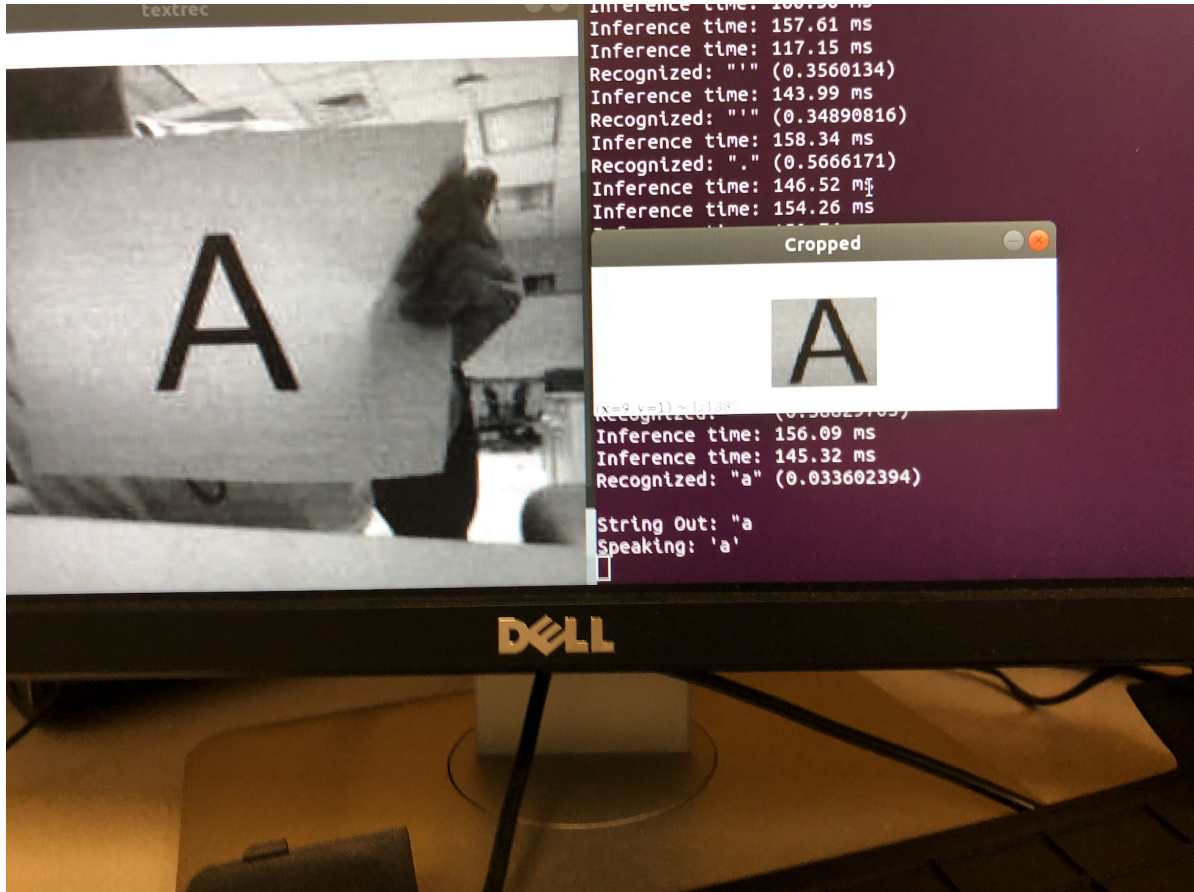
Text Detection Approach



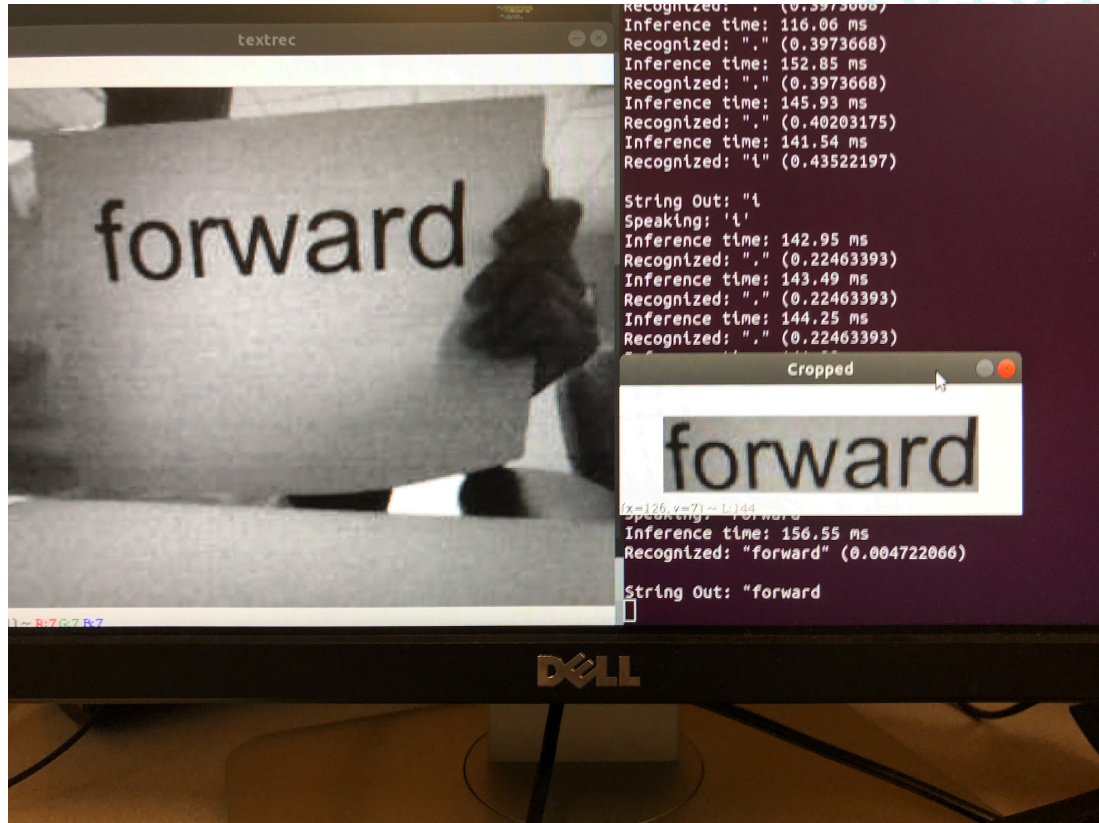
Dive into Text Recognizer Model



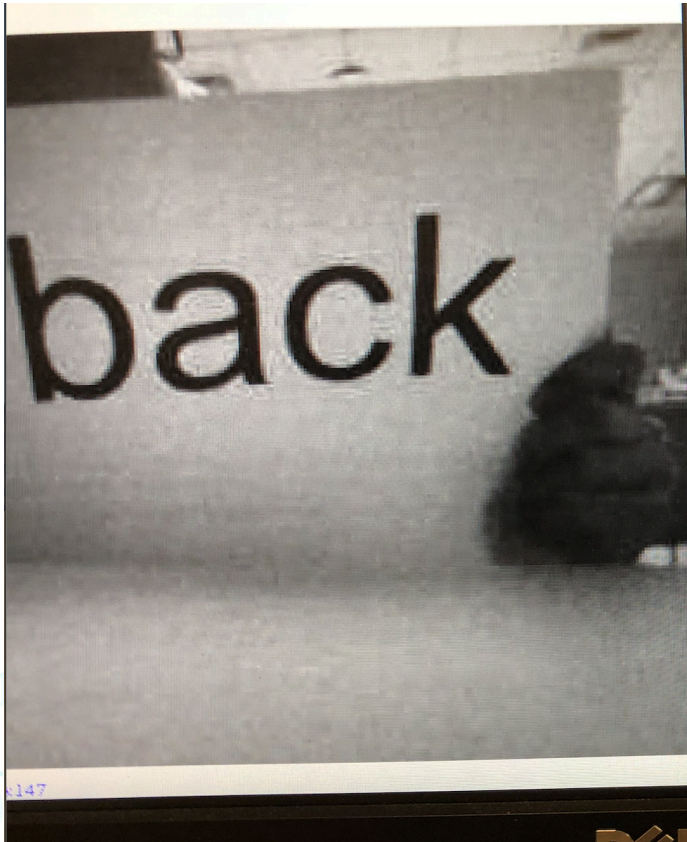
Letter Results



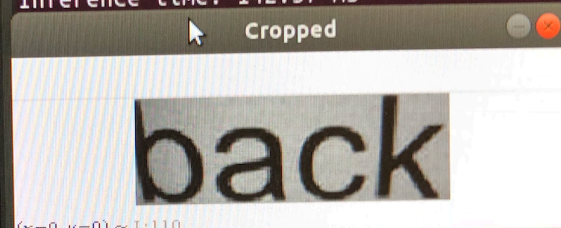
Action Results



Action Results



```
String Out: "dat  
Speaking: 'dat'  
Inference time: 145.51 ms  
Recognized: "." (0.41722226)  
Inference time: 148.75 ms  
Recognized: "." (0.41722226)  
Inference time: 145.40 ms  
Recognized: "." (0.41722226)  
Inference time: 142.63 ms  
Recognized: "." (0.09831259)  
Inference time: 142.57 ms
```



```
(x=9, y=0) ~ L:110  
Inference time: 144.42 ms  
Recognized: ".." (0.27936825)  
Inference time: 146.68 ms  
Recognized: ".." (0.27936825)  
Inference time: 143.58 ms  
Recognized: "back" (0.0016539685)
```

```
String Out: "back"
```

Action Results

The image shows a video analysis application interface. On the left, a blurred video frame is displayed. On the right, a terminal window shows the following output:

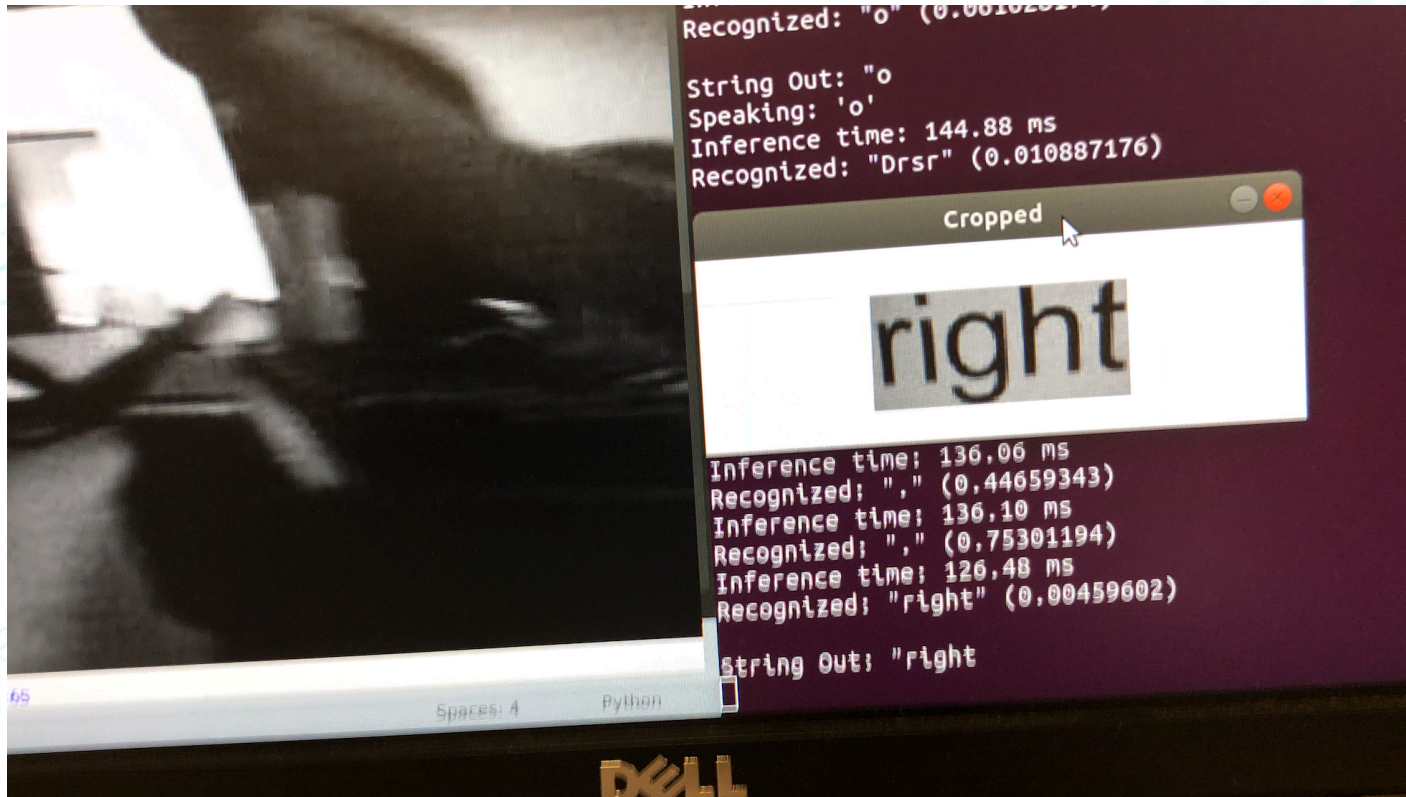
```
Inference time: 147.80 ms  
Recognized: "a" (0.22105885)  
  
String Out: "a"  
Speaking: 'a'  
Inference time: 148.39 ms  
Recognized: "." (0.46049392)  
Inference time: 156.48 ms  
Recognized: "." (0.46049392)  
Inference time: 128.87 ms  
Recognized: "." (0.3575004)  
Inference time: 147.92 ms
```

A cropped window titled "Cropped" is overlaid on the terminal, showing the word "left" in a black font on a white background.

```
String Out: "a"  
Speaking: 'a'  
Inference time: 152.78 ms  
Recognized: "left" (0.009949117)  
  
String Out: "left"
```

At the bottom of the terminal window, the text "Spaces: 4 Python" is visible. The Dell logo is located at the bottom center of the image.

Action Results





```
Inference time: 150.55 ms  
Recognized: "." (0.38211742)  
Inference time: 128.76 ms  
Recognized: "." (0.53446937)  
Inference time: 162.81 ms  
Recognized: "." (0.91479313)  
Inference time: 153.65 ms  
Recognized: "bt" (0.05454302)
```

String Out: "bt"



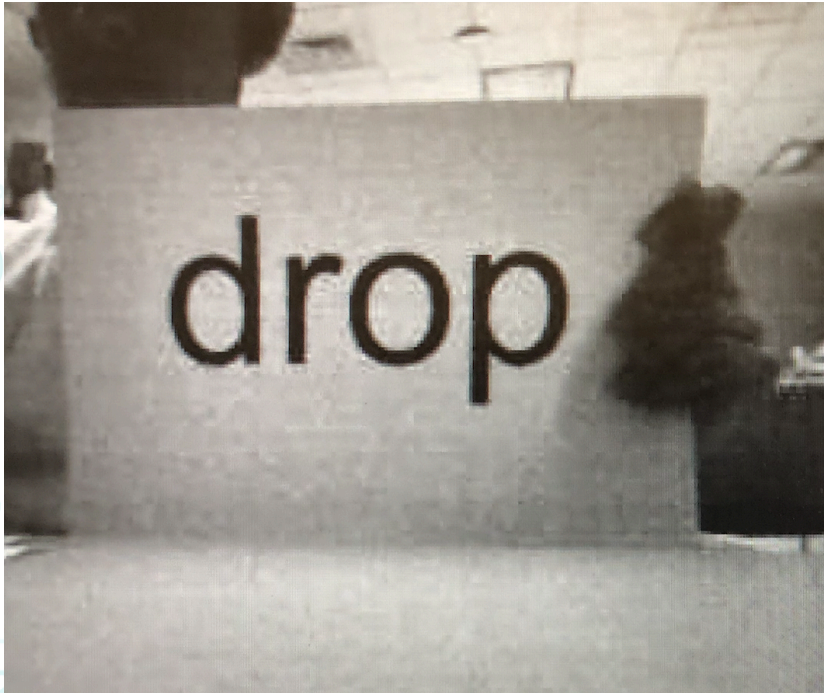
```
Recognized: "lift" (0.010220544)
```

String Out: "lift"

```
2019-05-03 14:39:28,132 cozno.general WARNING  
amping
```

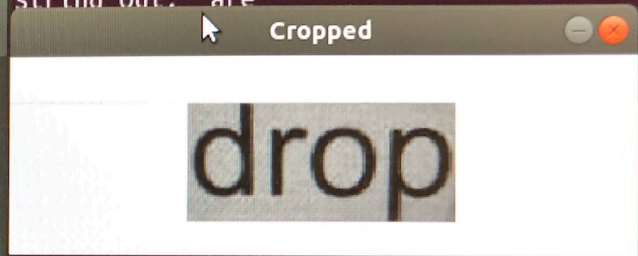


Action Results



```
Inference time: 143.50 ms  
Inference time: 139.25 ms  
Recognized: "." (0.22801353)  
Recognized: "#" (0.04927335)  
Inference time: 135.92 ms  
Recognized: "." (0.22801353)  
Recognized: "#" (0.04927335)  
Inference time: 128.32 ms  
Recognized: "are" (0.14880973)
```

String Out: "are"



```
String Out: "drop"  
Inference time: 144.34 ms  
Recognized: "a" (0.08197346)
```

Speak Results Signs

The image shows a screenshot of a textrec application window. The main window displays a photograph of a 'STOP ALTO' sign. Below the sign, there is a 'Web Price' section showing '\$92.96 / each' and a 'Shipping' section. To the right of the main window, there is a vertical stack of three smaller windows:

- The top window, titled 'Cropped', shows a close-up of the 'STOP' text from the sign.
- The middle window, titled 'Processed', shows the 'STOP' text rotated 90 degrees counter-clockwise.
- The bottom window shows the recognition result: 'Recognized: "sTOP" (0.088267654)'. A small white box is visible below the text.

At the bottom left of the main window, there is a bounding box coordinate: `x=146, v=591 ~ R:184 G:184 B:184`. At the top right of the main window, there is a terminal window with the following code snippet:

```
import ttp
/usr/local/lib/python3.6/dist-packages/te
inspect.getargspec() is deprecated, use in
return _inspect.getargspec(target)
Validation character error rate of saved
/usr/local/lib/python3.6/dist-packages/te
```



Demo

What I Learned

- Noisy Images with low resolution do not respond well in trained datasets with high resolution
- Python code is very abstract, most optimization has to be done on between cv2 calls
 - Lots of Parameter and Threshold Fitting
- Image and Bounding Box Dimensions do not fix all results
- Do not assume SDK works as intended, fiddle with it!
- Character Recognition Difficult
- String Recognition Very Difficult
- Sentence Recognition Extremely Difficult

Future Work

- Retrain classifier with Cozmo camera text images
- Only text values that should be trained on would be the actions rather than all text
 - Stop
 - Forward
 - Backward
 - Left
 - Right
 - Up
 - Down
 - Lift
 - Drop



Thank You

