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

S p a c i n g

S p a c i n g

Spacing Words

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

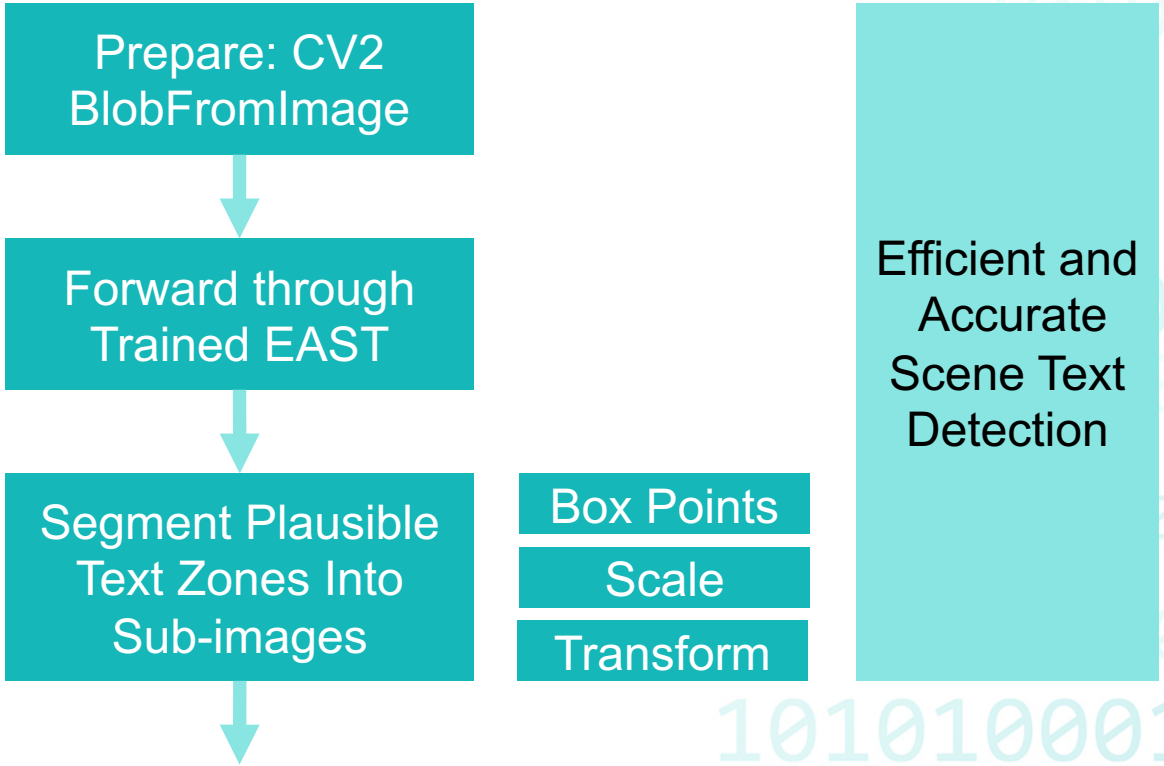
S p a c i n g W o r d s

Real World

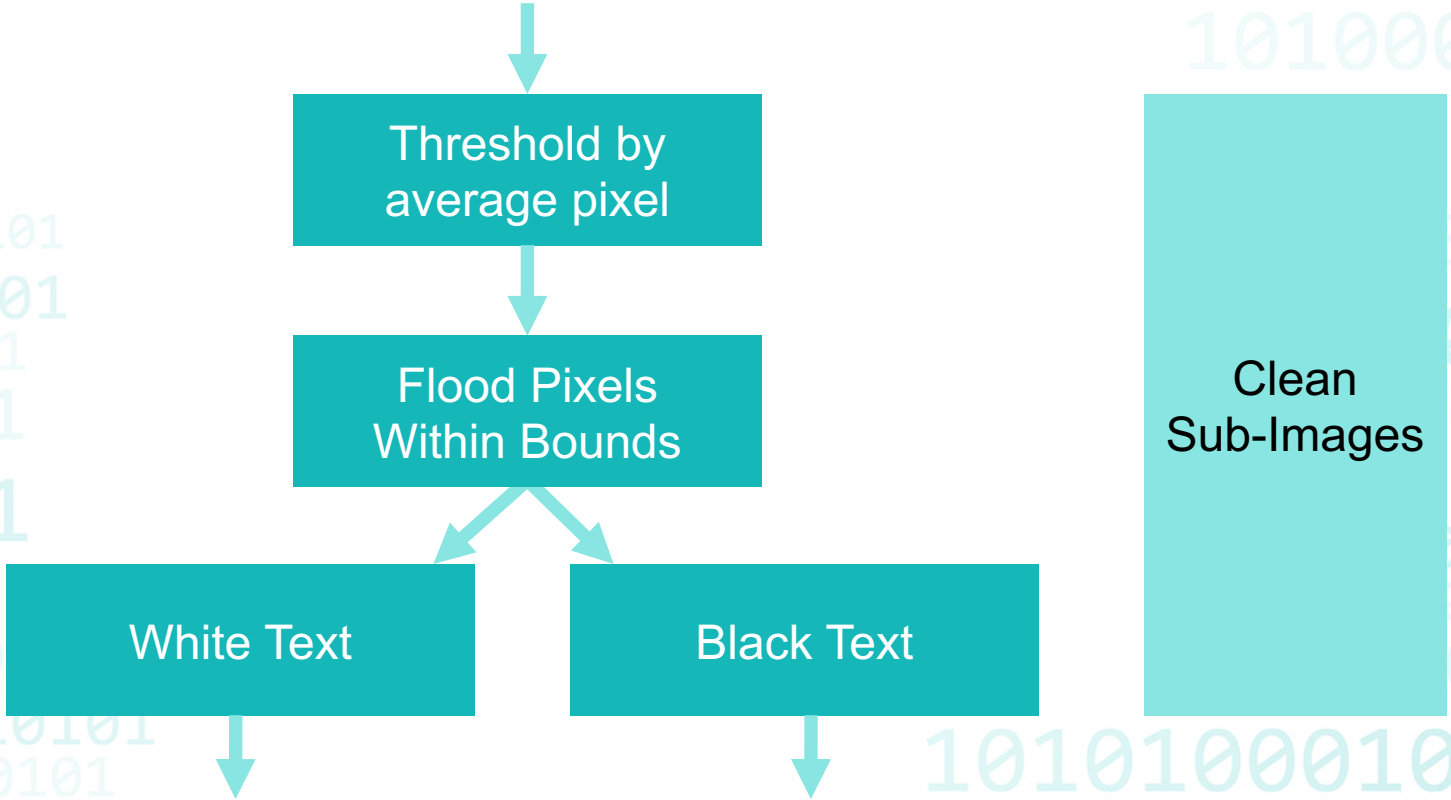

Domino's
OPEN


Twisted Texan
A twist of Texas pride in every design
132

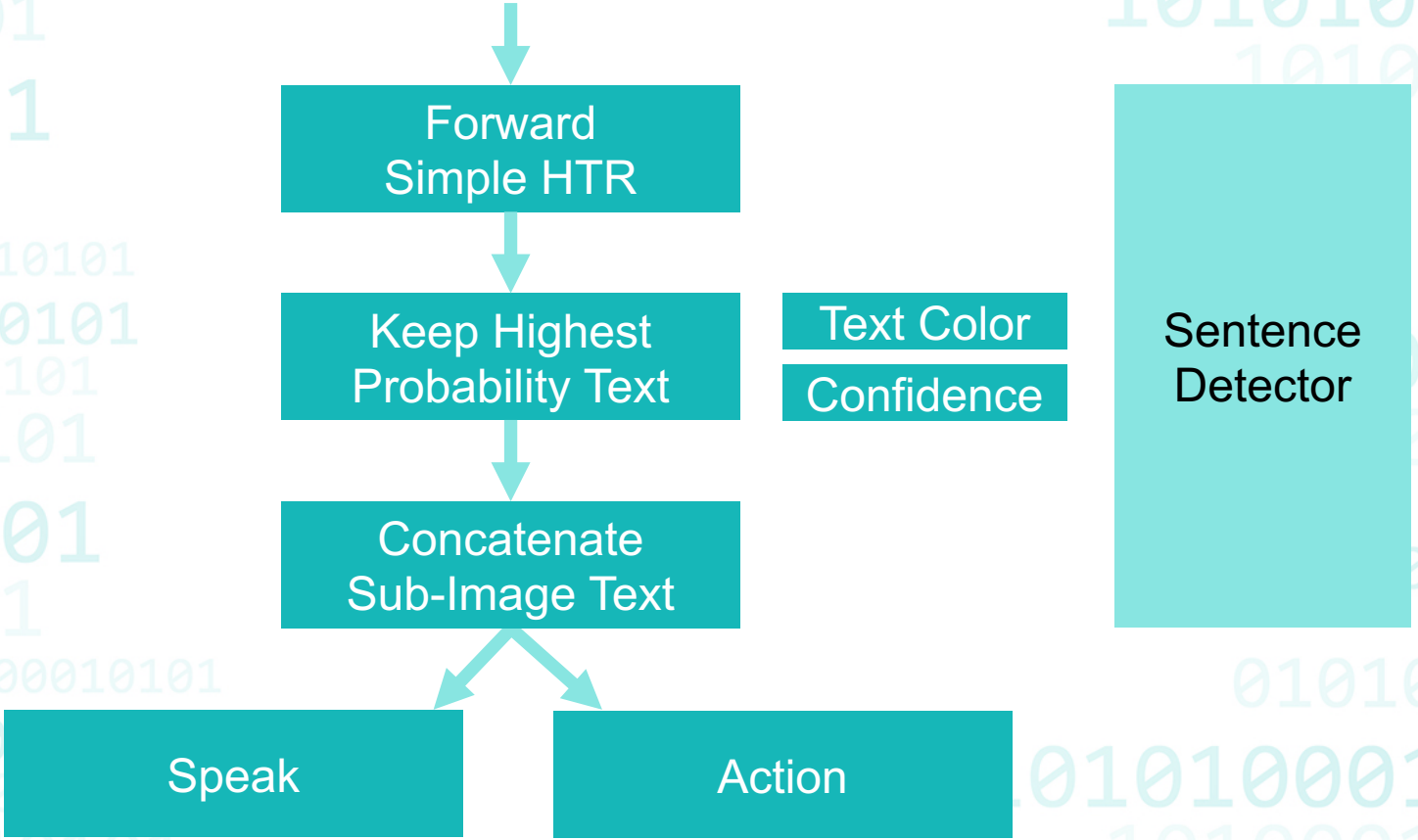
Text Detection Approach



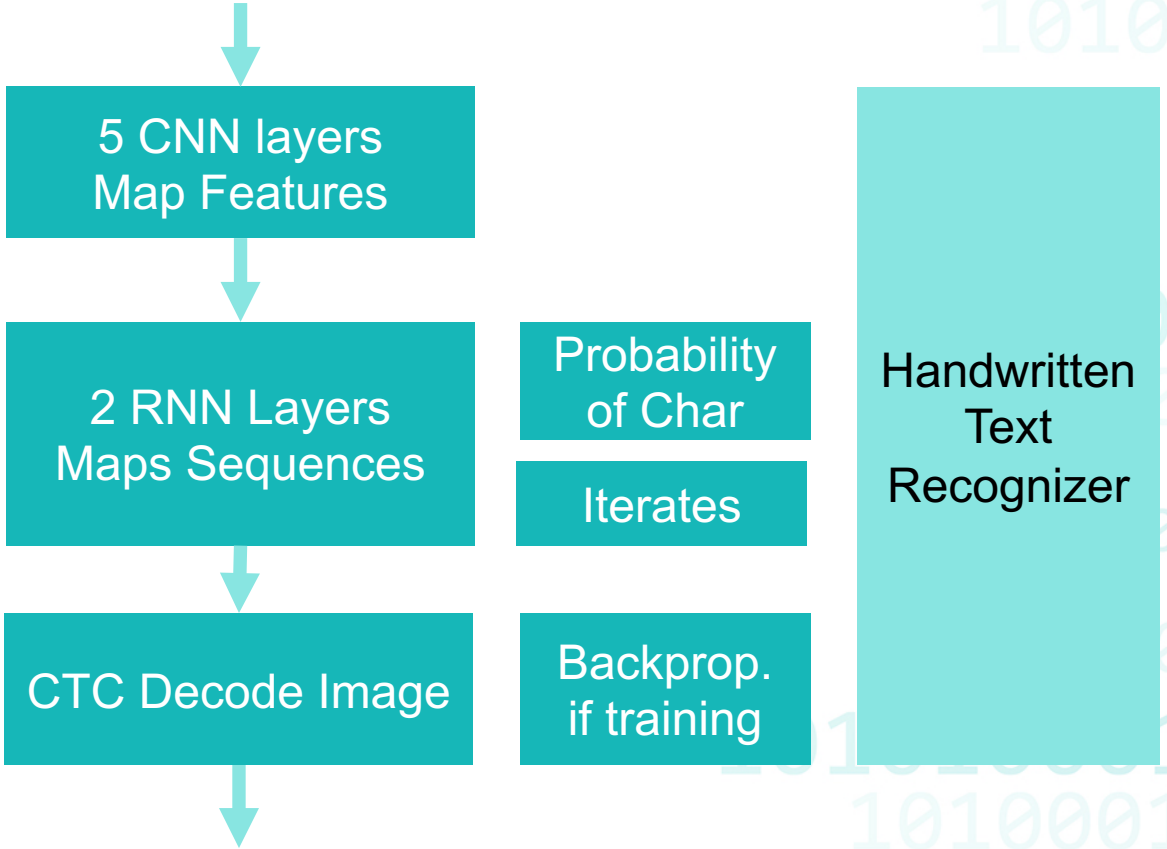
Text Interpreter Preprocess



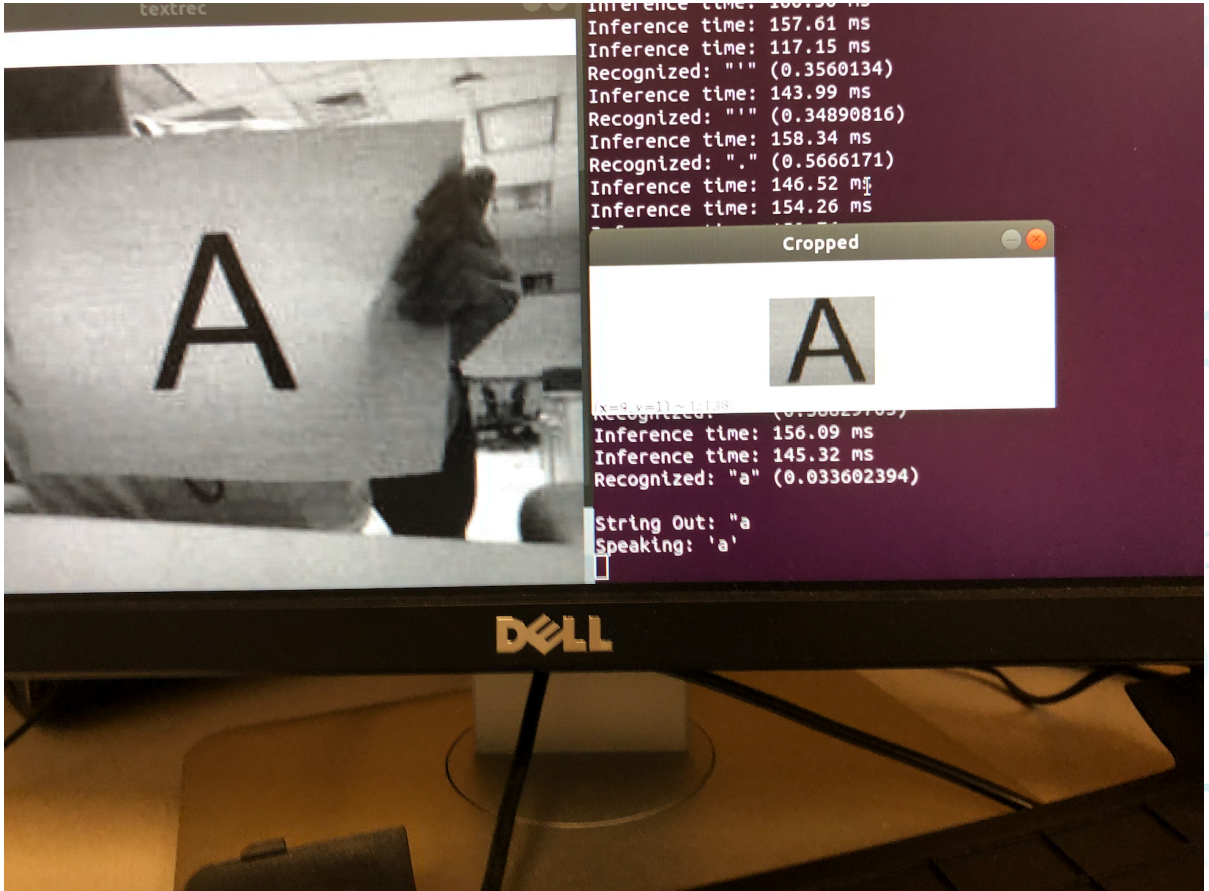
Text Detection Approach



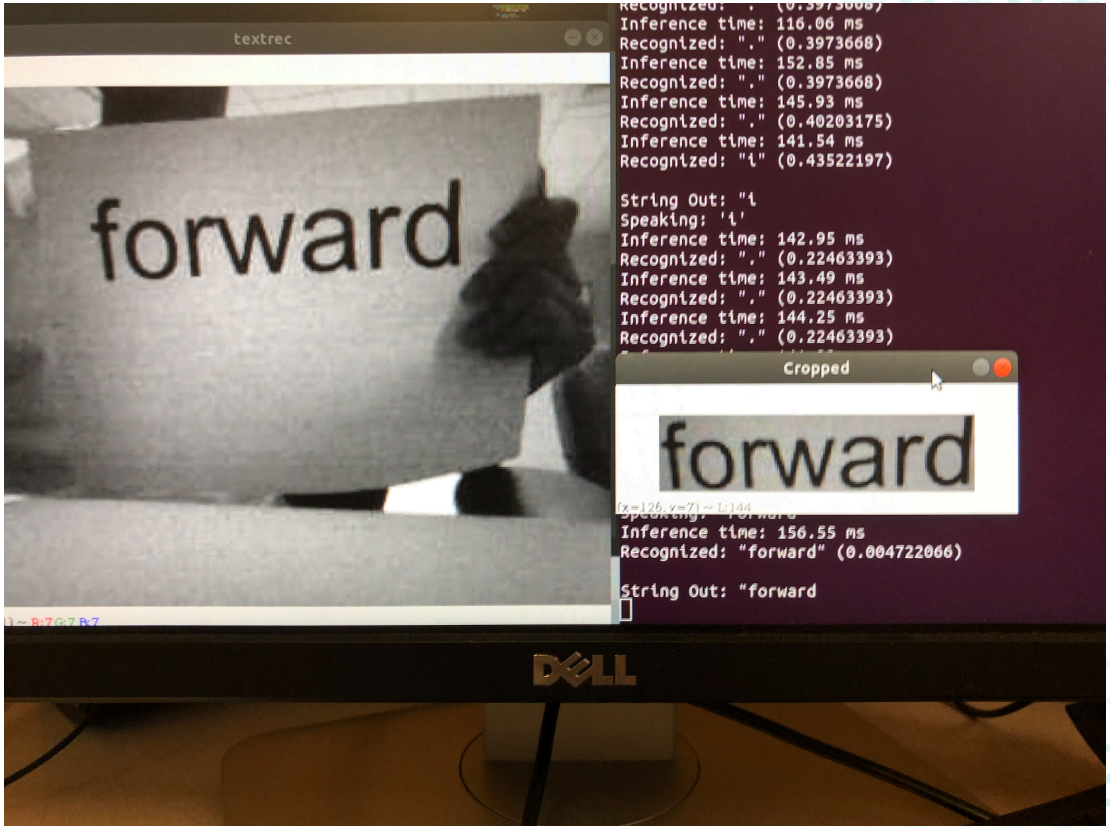
Dive into Text Recognizer Model



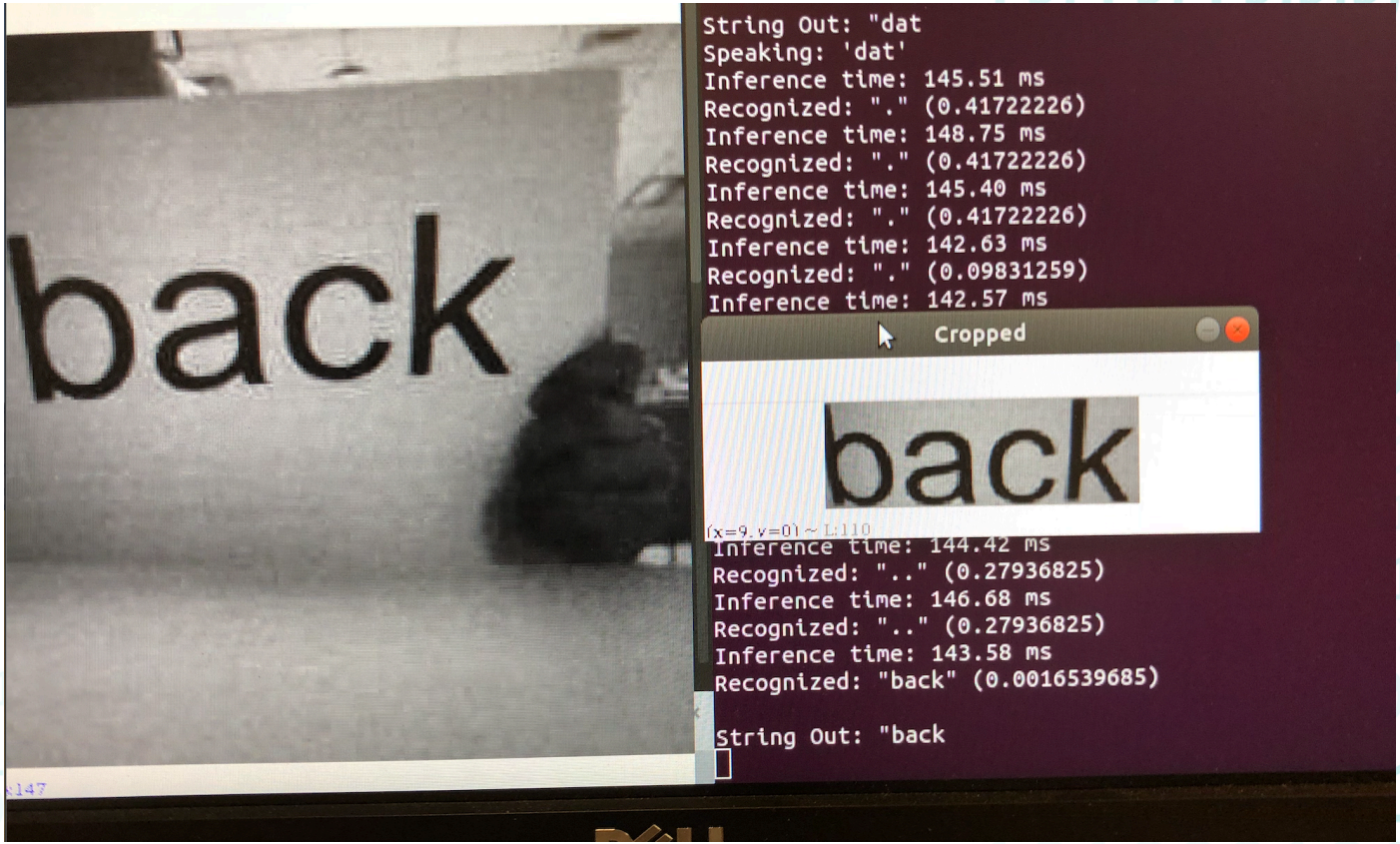
Letter Results



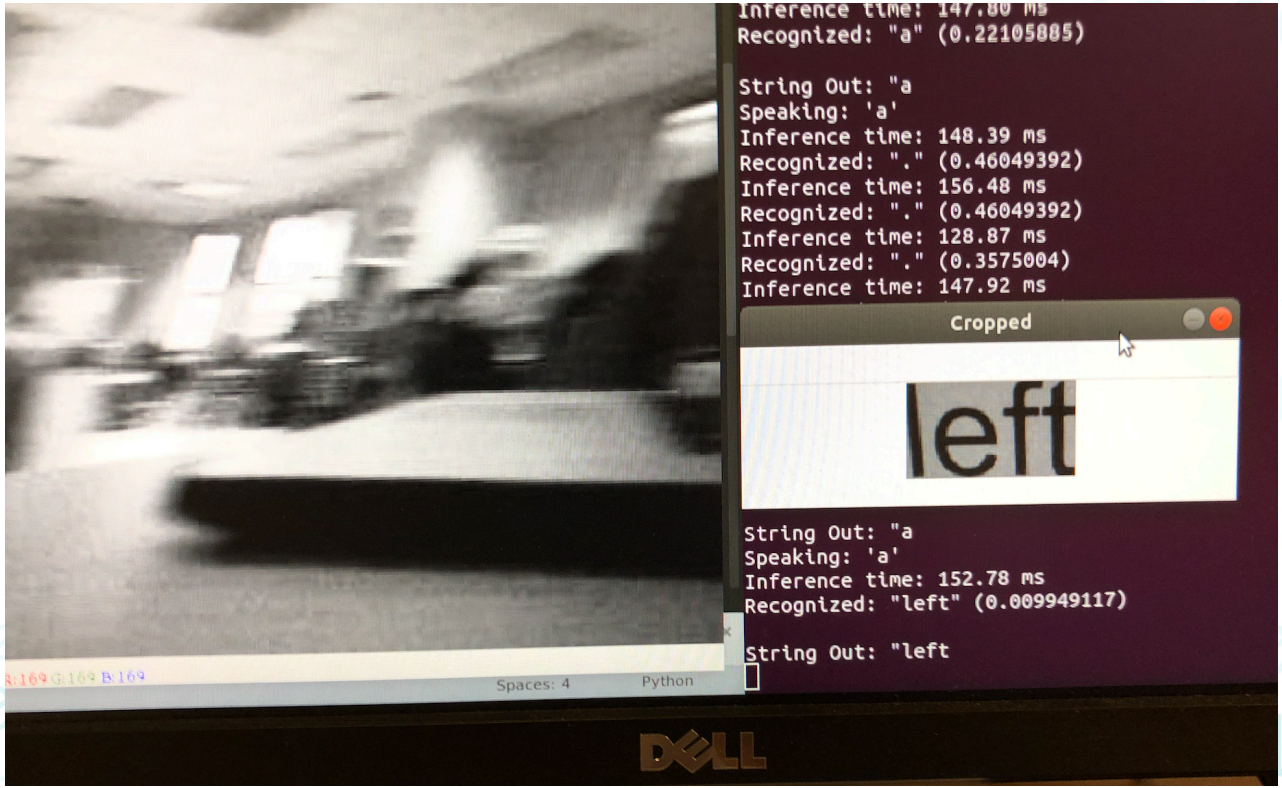
Action Results



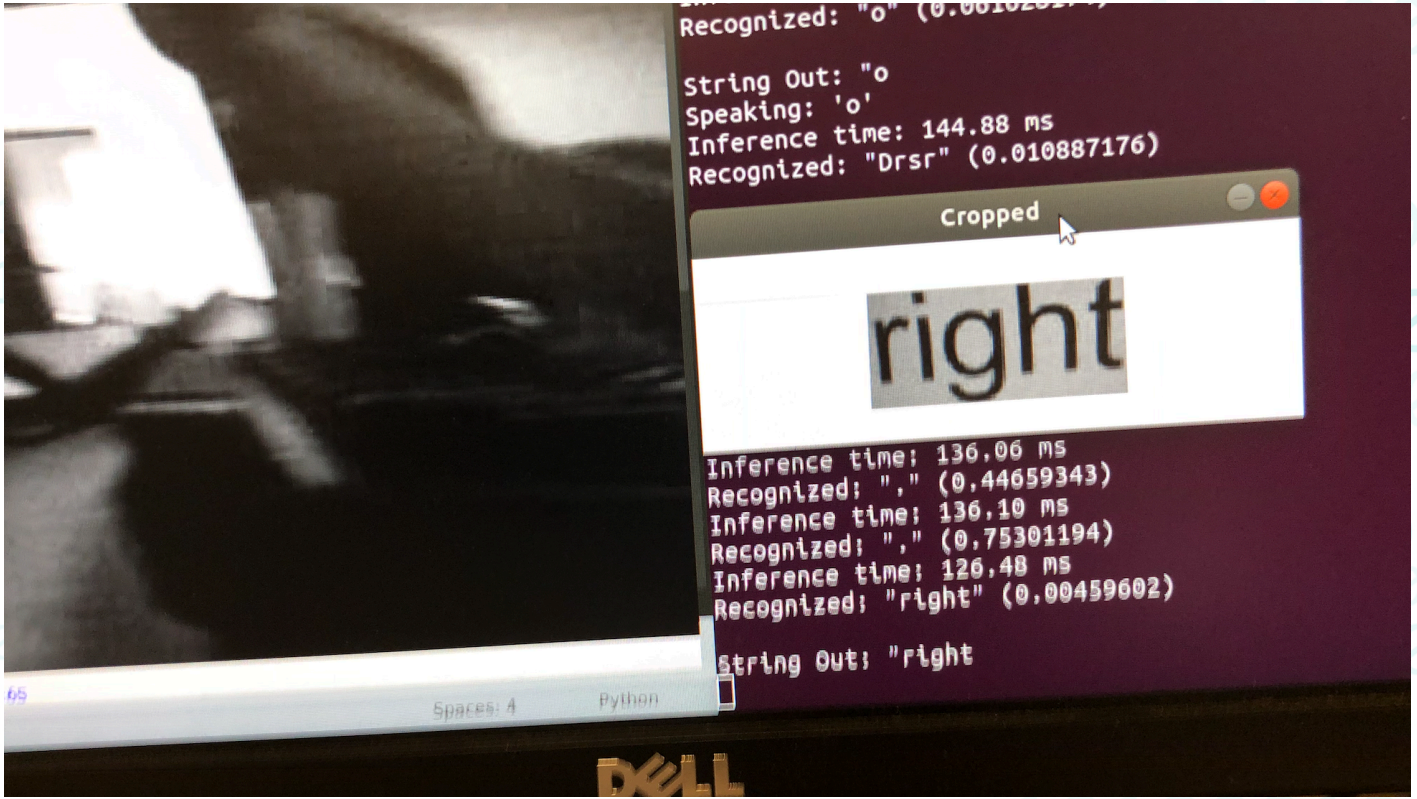
Action Results

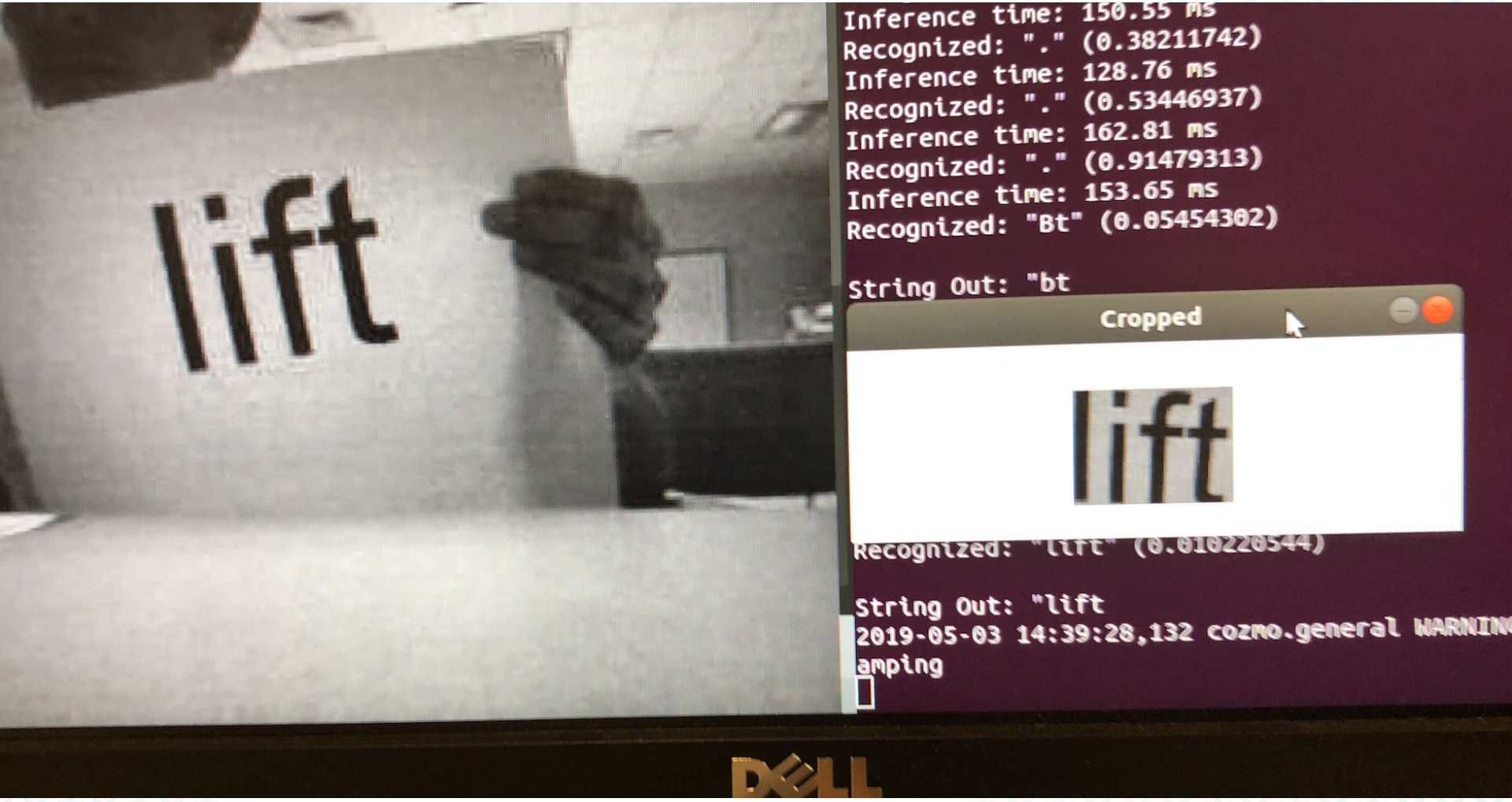


Action Results

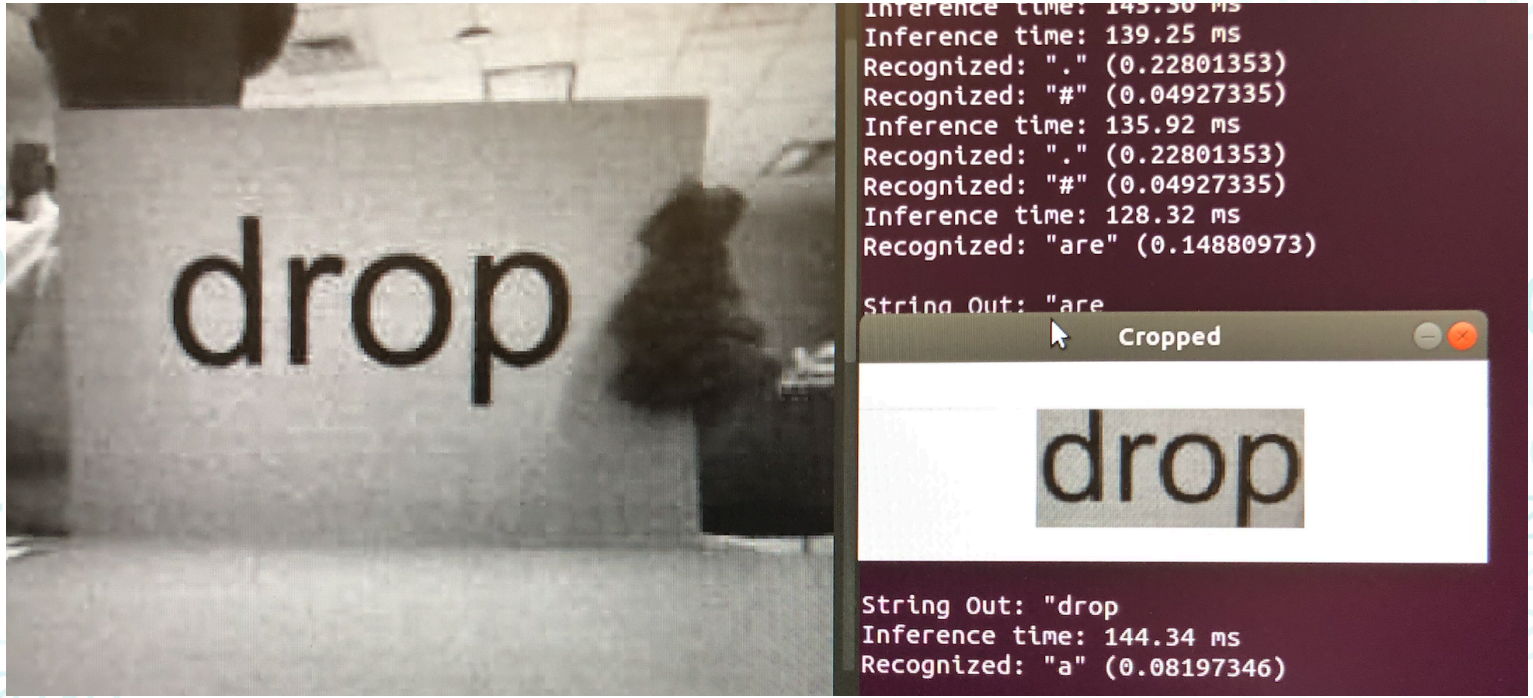


Action Results




lift
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
Cropped
lift
String Out: "lift
2019-05-03 14:39:28,132 cozmo.general
amping

Action Results



Speak Results Signs

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