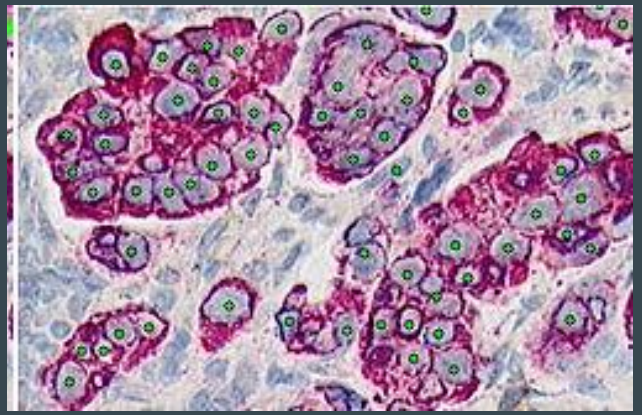
SITUATE : MACHINE LEARNING FOR IMAGE RECOGNITION

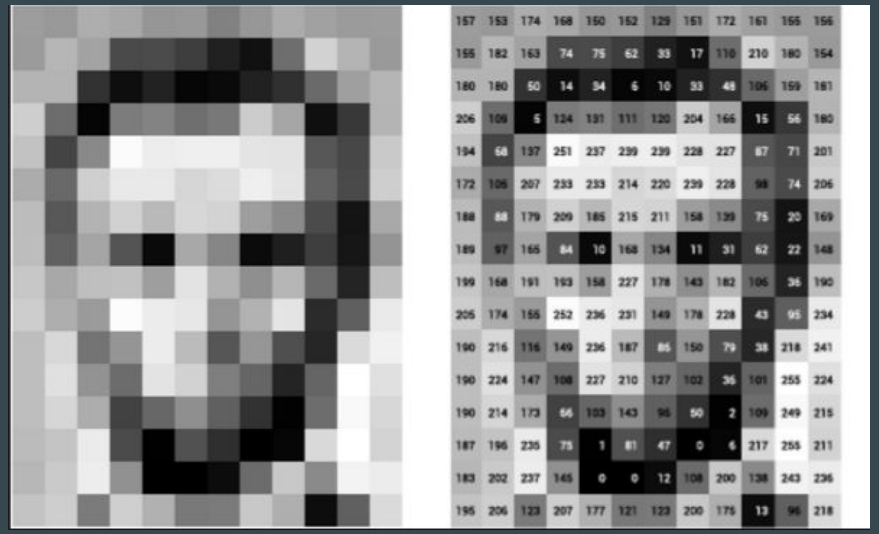
By: Bryan Lee and Robin Tan
ASE 2016

Introduction

Why Image Recognition?



Why is Image Recognition Difficult?



What is Machine Learning?

*An **algorithm** with the ability to learn **without** being **explicitly programmed***

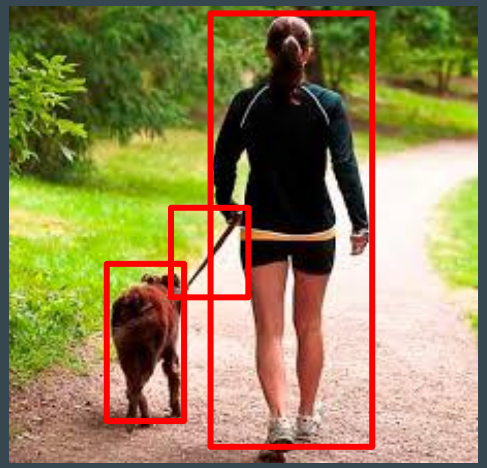
Input Data → **Finds Patterns** → **Predict**

(Training) **(Algorithm- Learning)** **(Testing)**

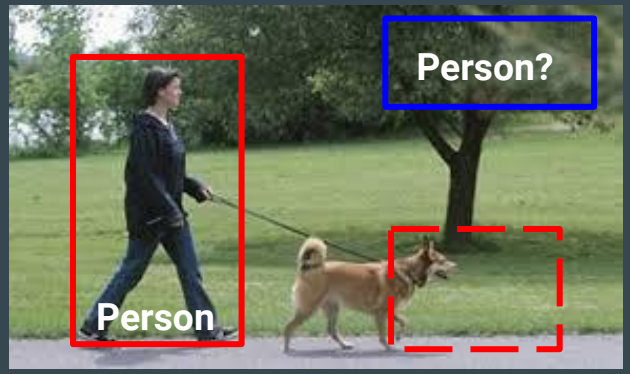
Situate : Information and Goals

Locate objects relevant to situation as quickly as possible

Situate : Overview

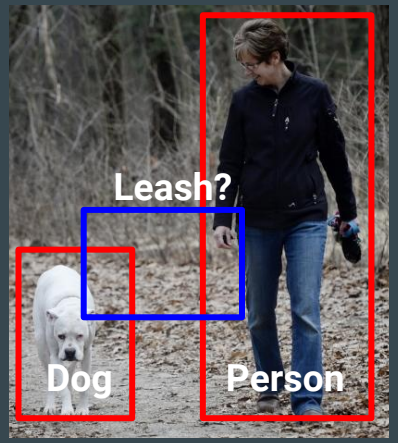


Dog Walking



Bounding Boxes

Object Proposals
Oracle
Intersection-over-Union



Using Context

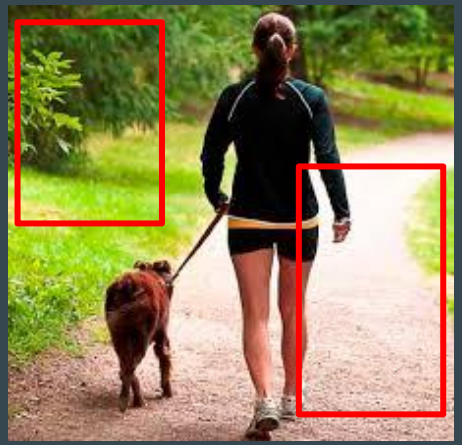
Project Goals

Problem 1

Generalize Situate

Problem 2

Improve Box Proposals

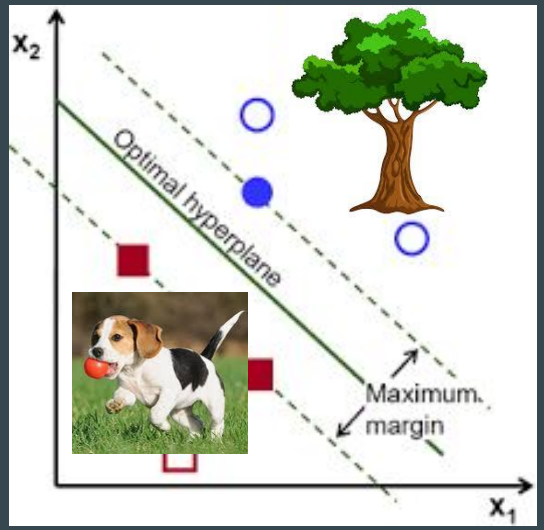


Problem 3

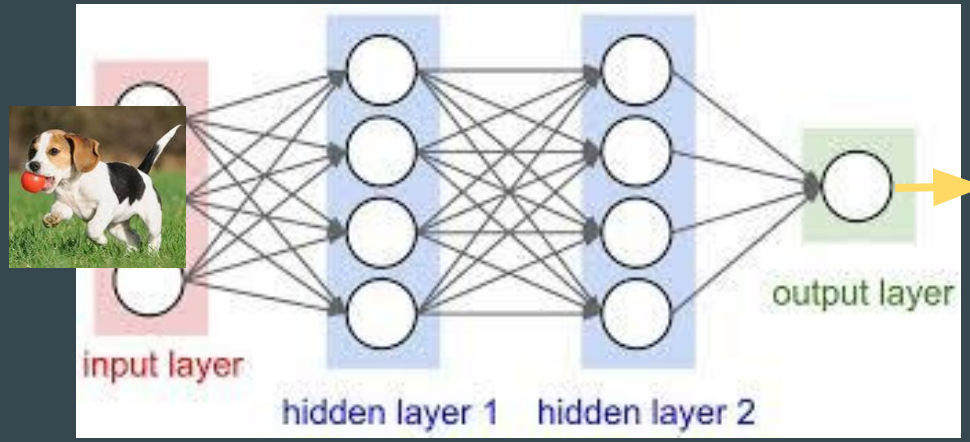
Accurately Identify Object in the box

Machine Learning Methods

Support Vector Machines



Neural Networks



Situate : User Interface

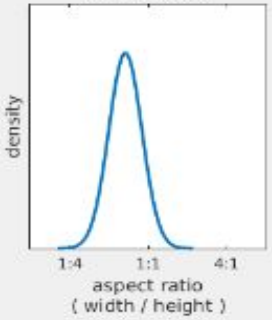
Workspace
handshake88

person_myleft location

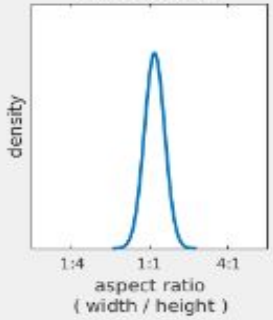
handshake location

person_myright location

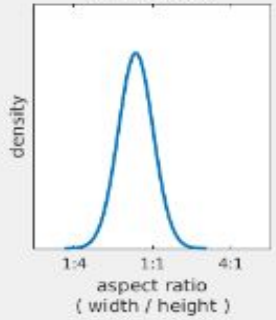
person_myleft box distribution

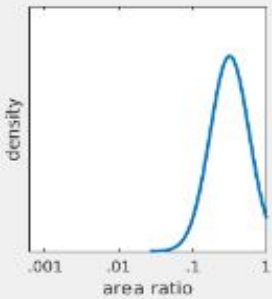


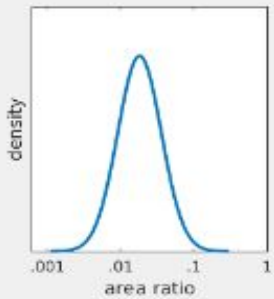
handshake box distribution

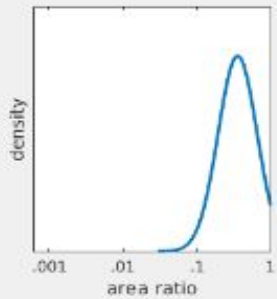


person_myright box distribution









Start | Pause | Step | Next image

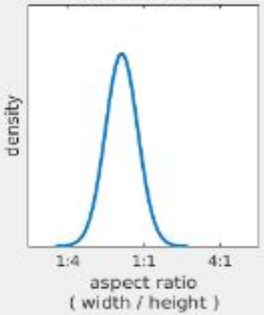
Workspace
handshake88
1/1000

person$_m$y$_l$eft location

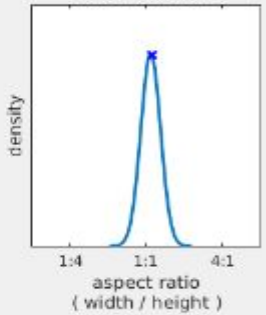
handshake location

person$_m$y$_r$ight location

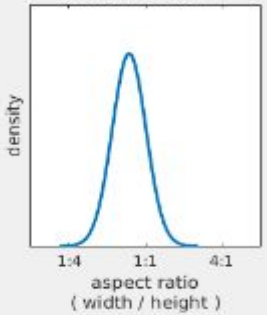
person$_m$y$_l$eft box distribution

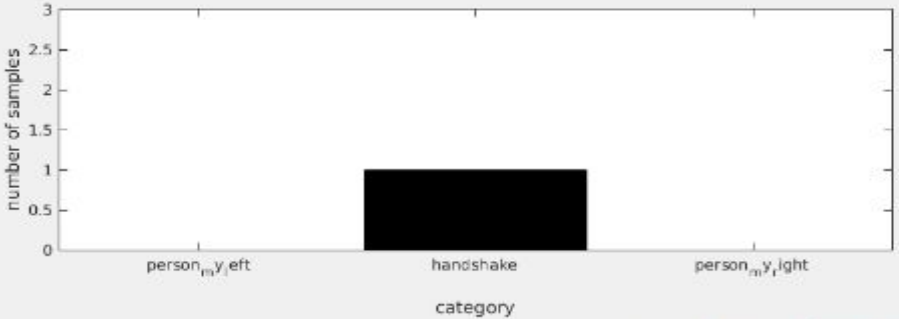


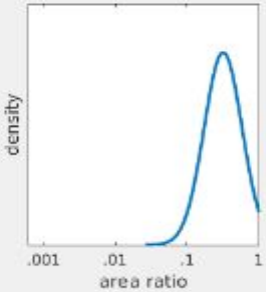
handshake box distribution

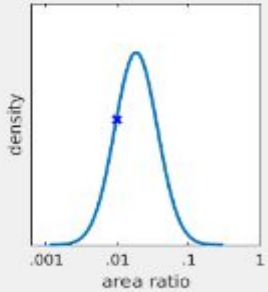


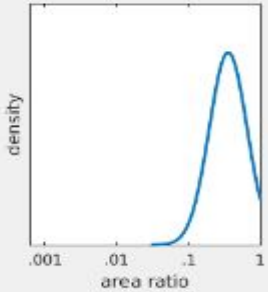
person$_m$y$_r$ight box distribution











Restart | Resume | Step | Next image

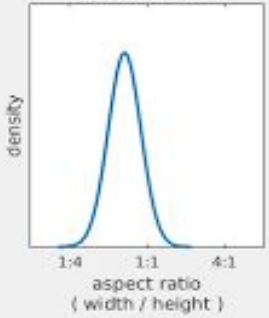
Workspace
handshake88
2/1000

person$_m$y$_l$eft location

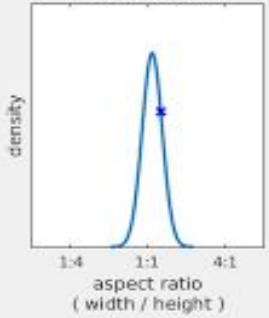
handshake location

person$_m$y$_r$ight location

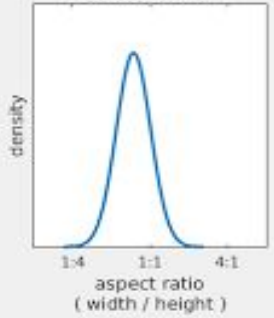
person$_m$y$_l$eft box distribution

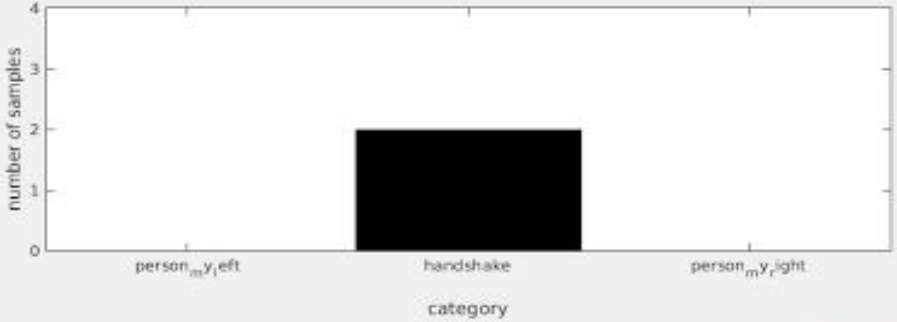


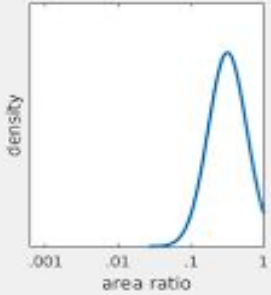
handshake box distribution

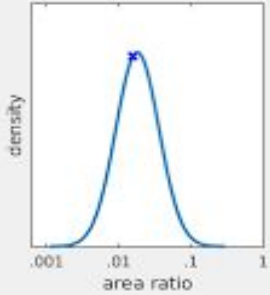


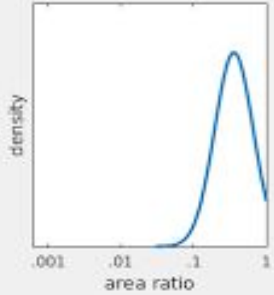
person$_m$y$_r$ight box distribution











Restart | Resume | Step | Next image

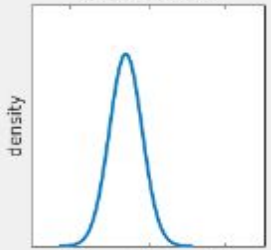
Workspace
handshake88
3/1000

person$_m$y$_l$eft location

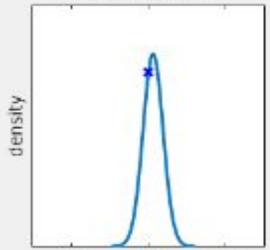
handshake location

person$_m$y$_r$ight location

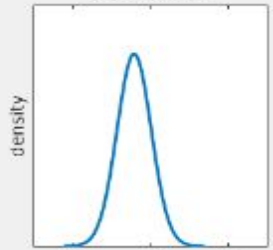
person$_m$y$_l$eft box distribution

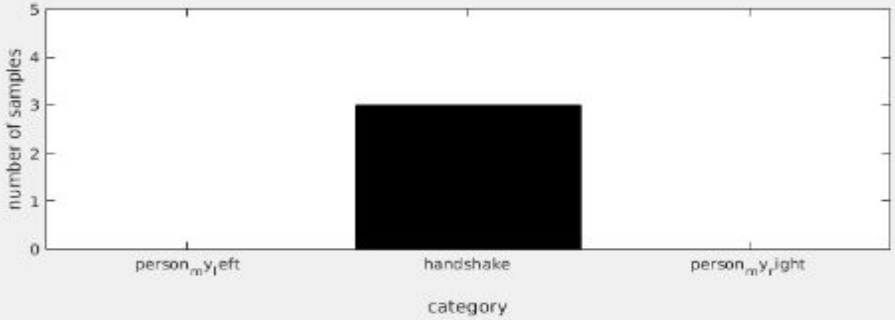


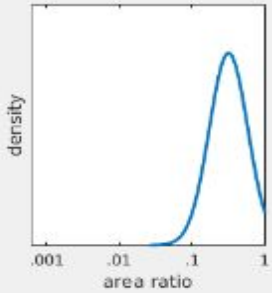
handshake box distribution

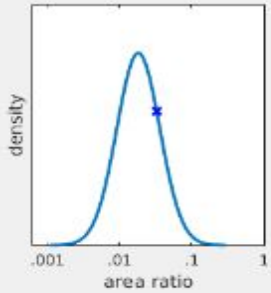


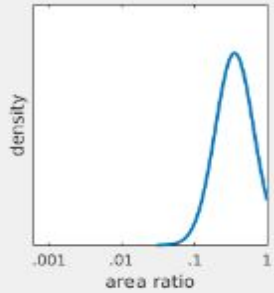
person$_m$y$_r$ight box distribution



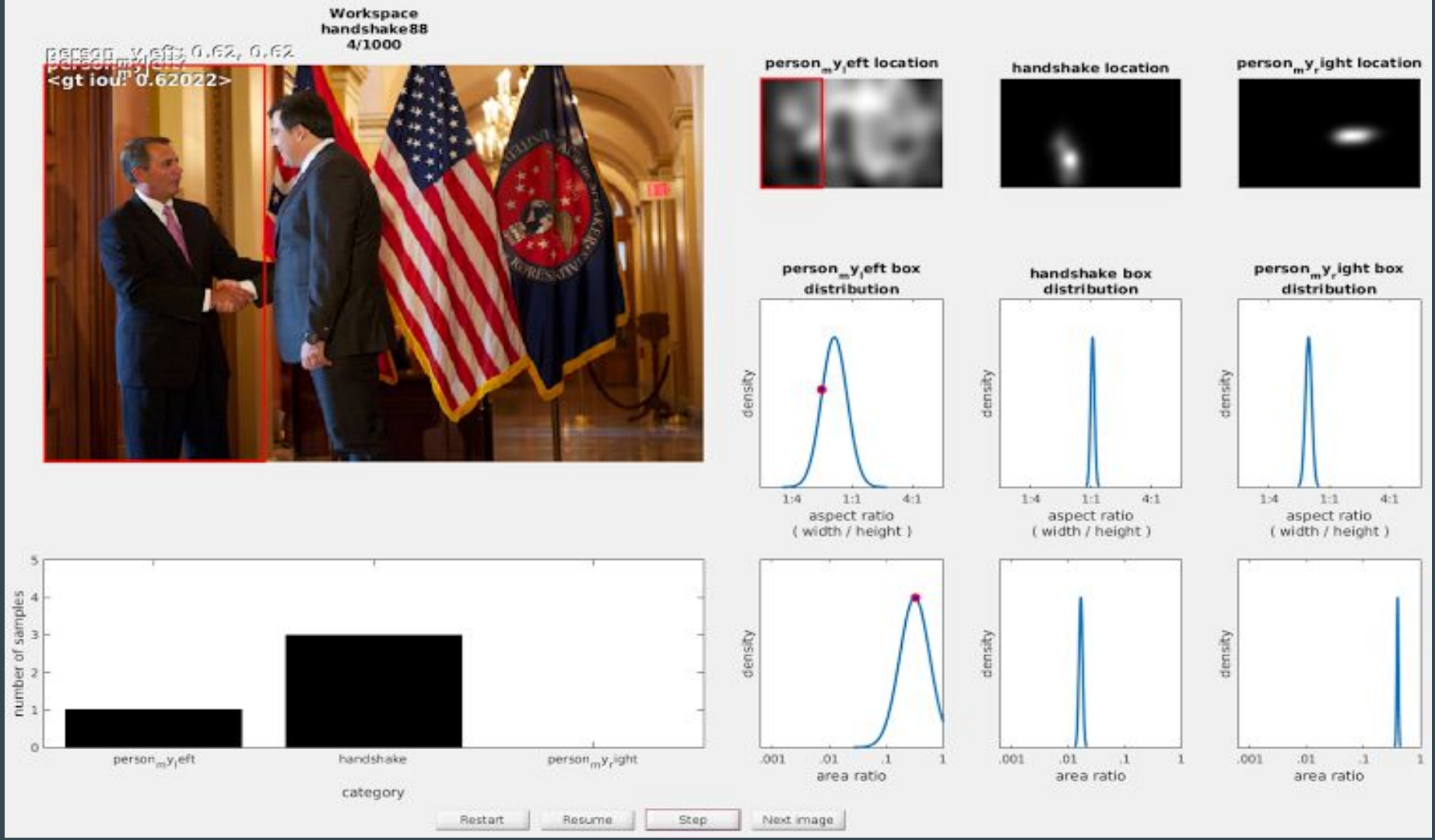


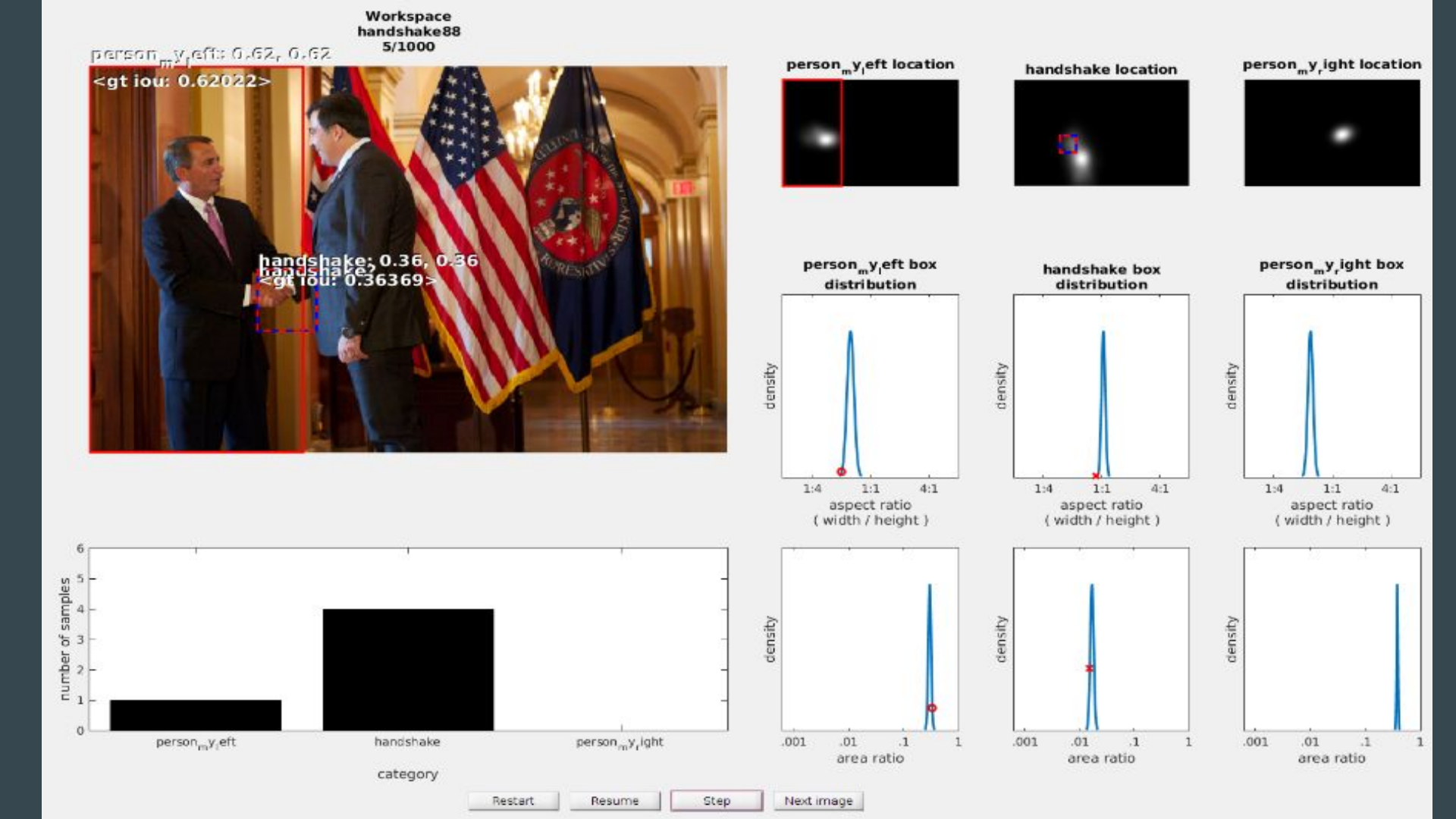


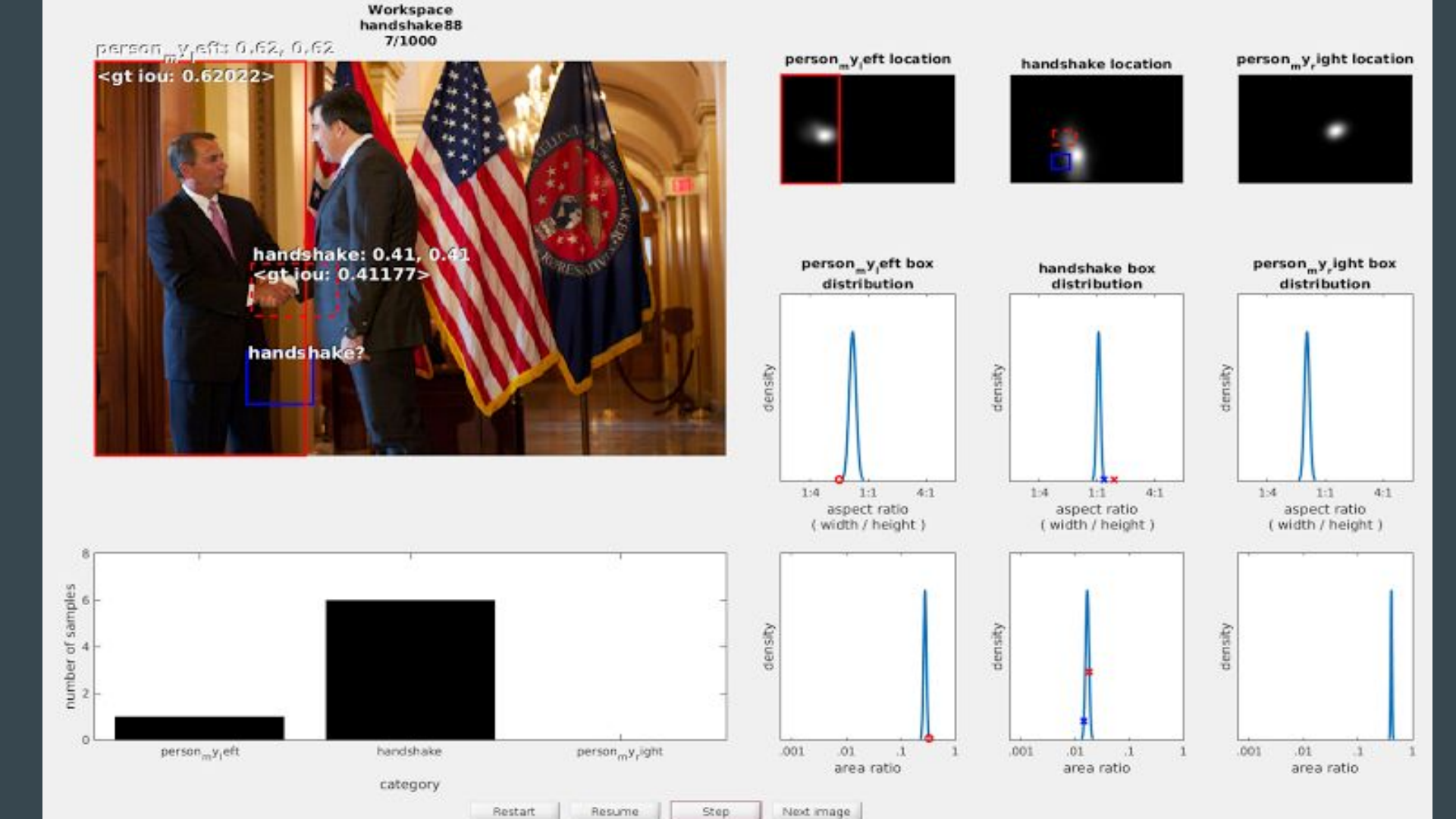




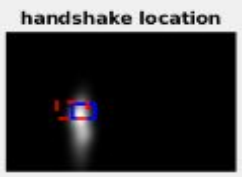
Restart | Resume | Step | Next image

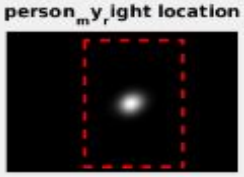

Workspace
handshake88
4/1000
person_y_eft: 0.62, 0.62
person_y_ight
<gt iou: 0.62022>
person_my_eft location
handshake location
person_my_ight location
person_my_eft box distribution
handshake box distribution
person_my_ight box distribution
density
1:4
1:1
4:1
aspect ratio
(width / height)
number of samples
person_my_eft
handshake
person_my_ight
category
.001
.01
.1
1
area ratio
Restart
Resume
Step
Next image

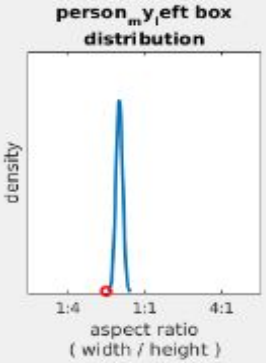

Workspace
handshake88
5/1000
person_myleft: 0.62, 0.62
<gt iou: 0.62022>
handshake: 0.36, 0.36
handshake
<gt iou: 0.36369>
person_myleft location
handshake location
person_myright location
person_myleft box distribution
handshake box distribution
person_myright box distribution
density
1:4
1:1
4:1
aspect ratio
(width / height)
number of samples
person_myleft
handshake
person_myright
category
.001
.01
.1
1
area ratio
Restart
Resume
Step
Next image

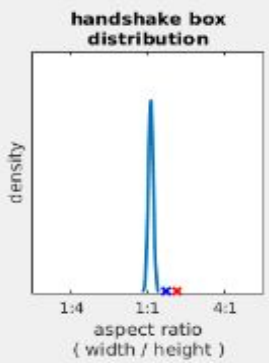

Workspace
handshake88
7/1000
person_my_left: 0.62, 0.62
<gt iou: 0.62022>
handshake: 0.41, 0.41
<gt iou: 0.41177>
handshake?
person_my_left location
handshake location
person_my_right location
person_my_left box distribution
handshake box distribution
person_my_right box distribution
density
1:4
1:1
4:1
aspect ratio
(width / height)
number of samples
person_my_left
handshake
person_my_right
category
.001
.01
.1
1
area ratio
Restart
Resume
Step
Next image

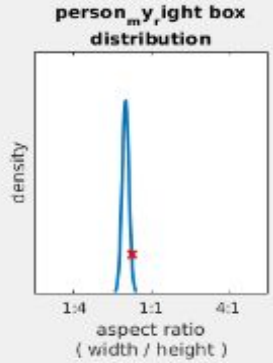
Workspace
handshake88
25/1000

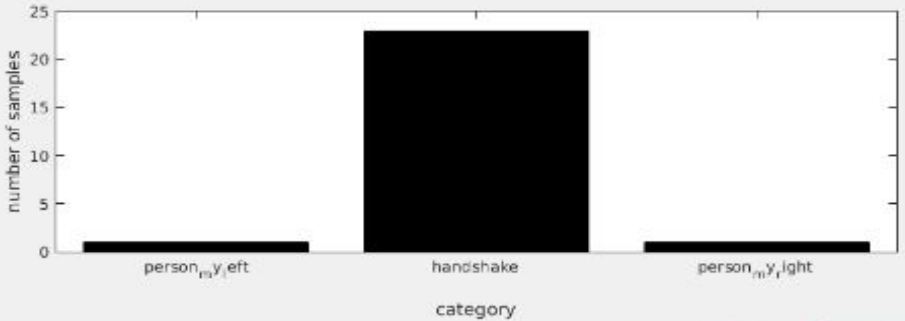


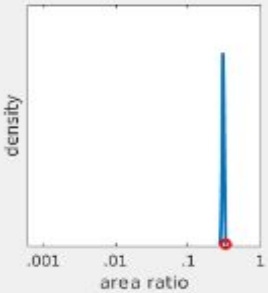


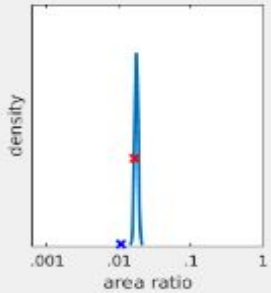


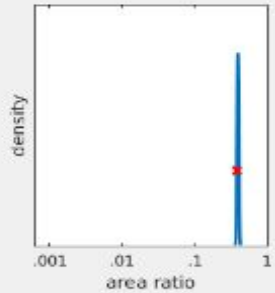




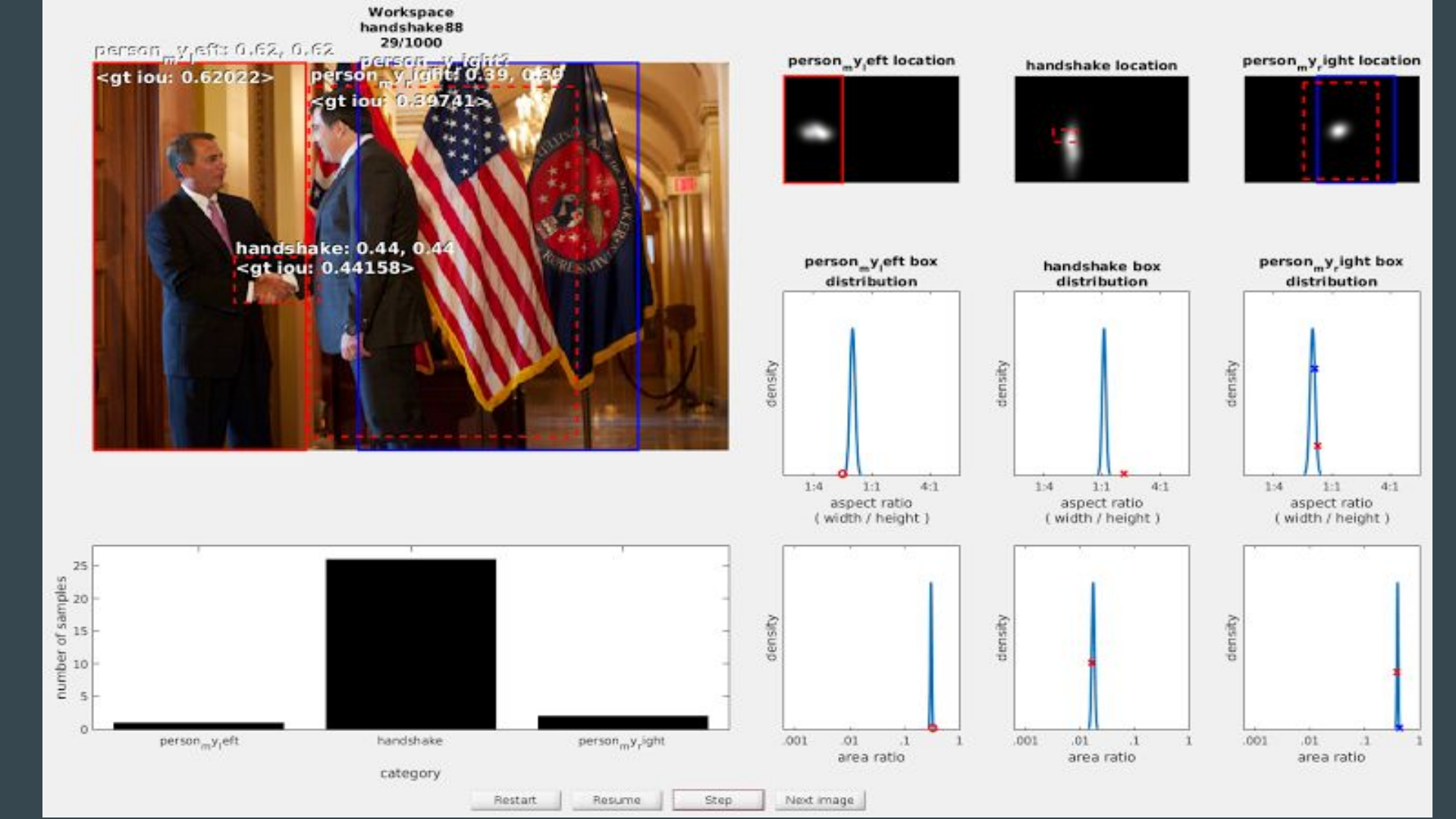


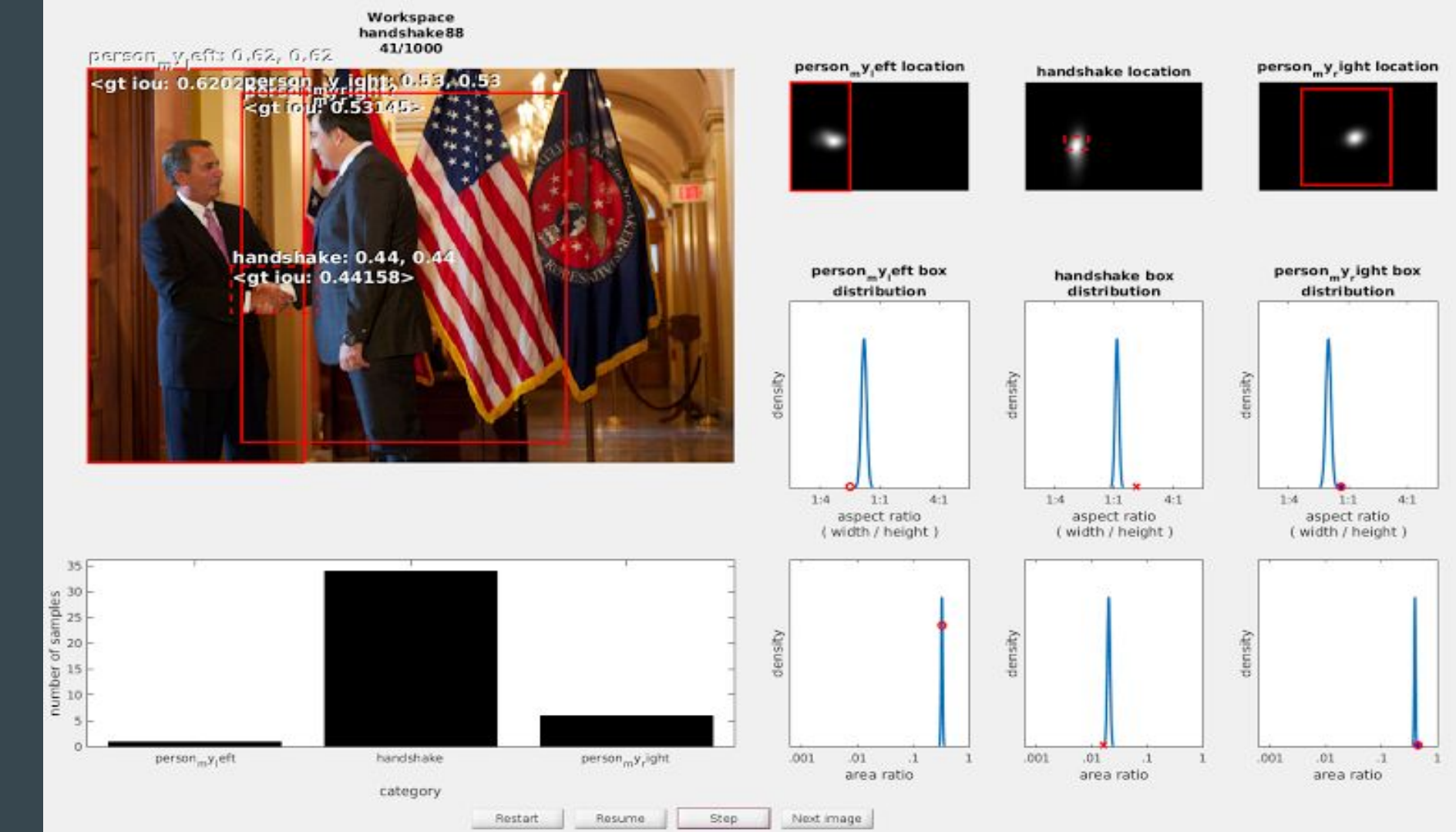




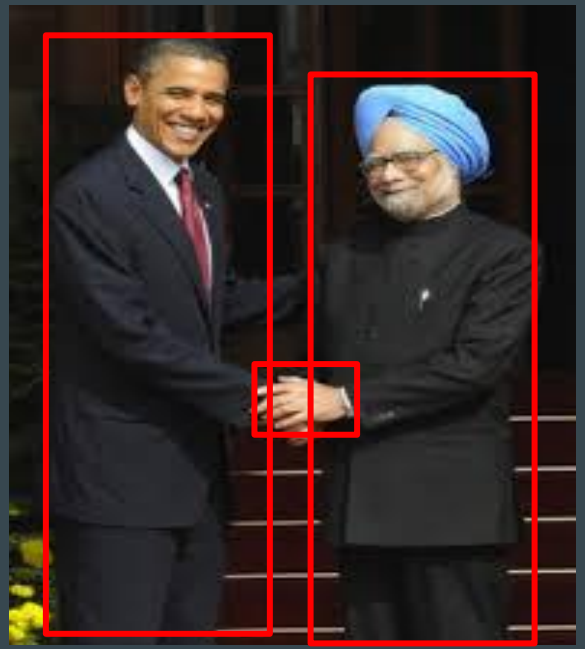


Restart | Resume | Step | Next image

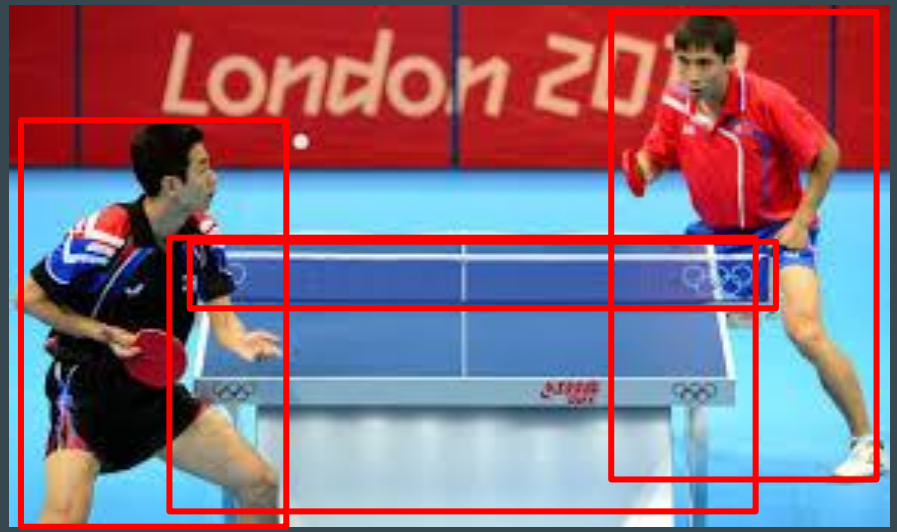

Workspace
handshake88
29/1000
person_myleft: 0.62, 0.62
<gt iou: 0.62022>
person_myright: 0.39, 0.39
<gt iou: 0.39741>
handshake: 0.44, 0.44
<gt iou: 0.44158>
person_myleft location
handshake location
person_myright location
person_myleft box distribution
handshake box distribution
person_myright box distribution
density
1:4
1:1
4:1
aspect ratio
(width / height)
number of samples
person_myleft
handshake
person_myright
category
.001
.01
.1
1
area ratio
Restart
Resume
Step
Next image


Workspace
handshake88
41/1000
person_myleft: 0.62, 0.62
<gt iou: 0.6202
person_myright: 0.53, 0.53
<gt iou: 0.53145>
handshake: 0.44, 0.44
<gt iou: 0.44158>
person_myleft location
handshake location
person_myright location
person_myleft box distribution
handshake box distribution
person_myright box distribution
density
1:4
1:1
4:1
aspect ratio
(width / height)
number of samples
person_myleft
handshake
person_myright
category
.001
.01
.1
1
area ratio
Restart
Resume
Step
Next image

[Goal 1] *Situate* : New Situations



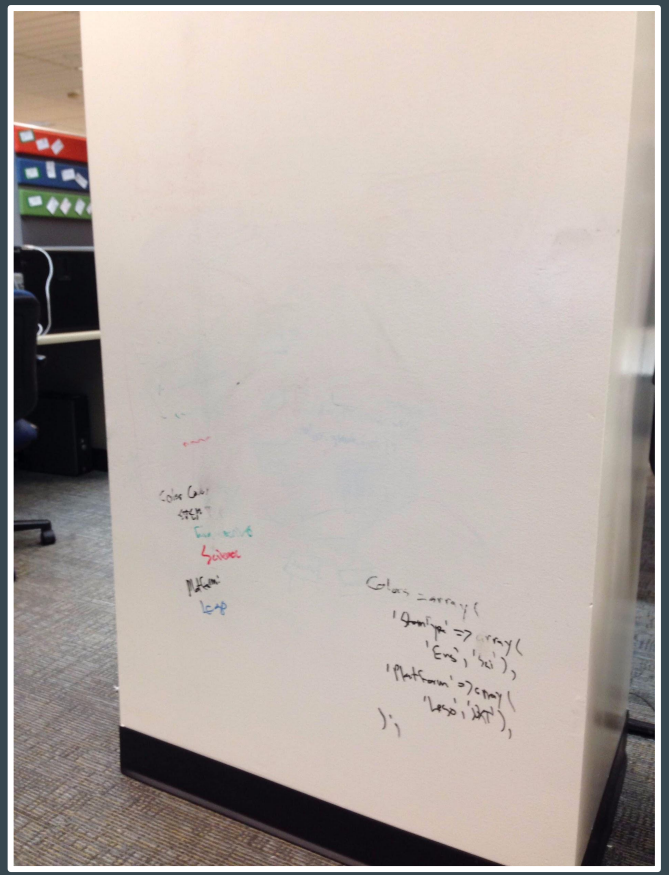
Handshaking



Ping Pong

Goal 1

[Goal 1.5] Wall Cleaning



[Goal 2] Improve Box Proposals

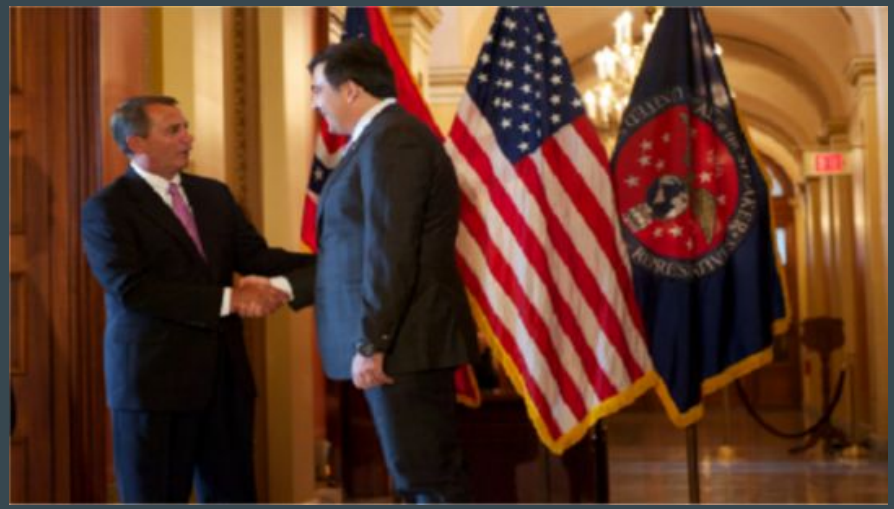
Goal 2

Problem 1: Where to propose the box?

Box Proposals based on

Salience

Prior knowledge on object location

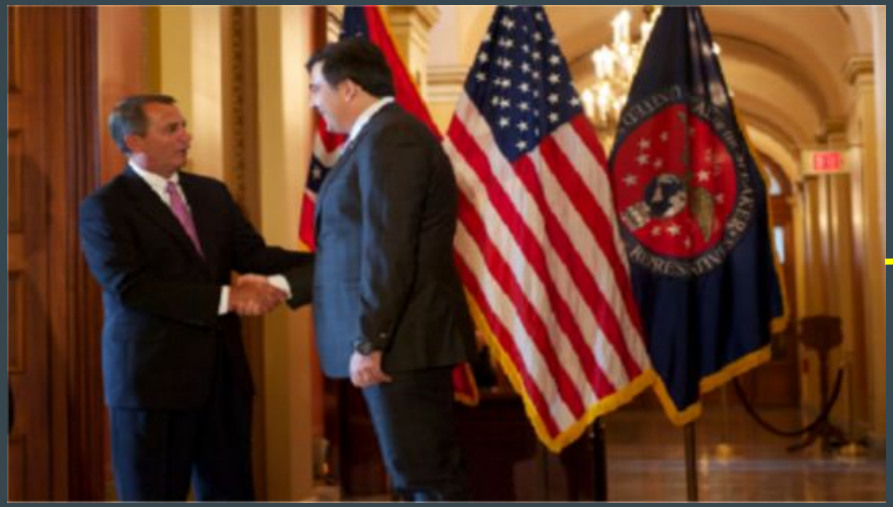


Goal 2

Note: Additional research on partial boxes

person_my_left location
handshake location
person_my_right location

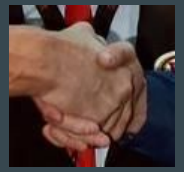
person_my_left location
handshake location
person_my_right location

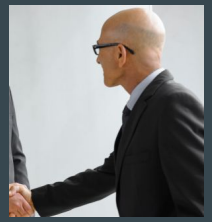


Person-Left

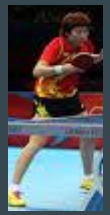
Goal 2

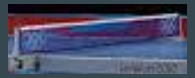
Problem 2: What size should the box be?



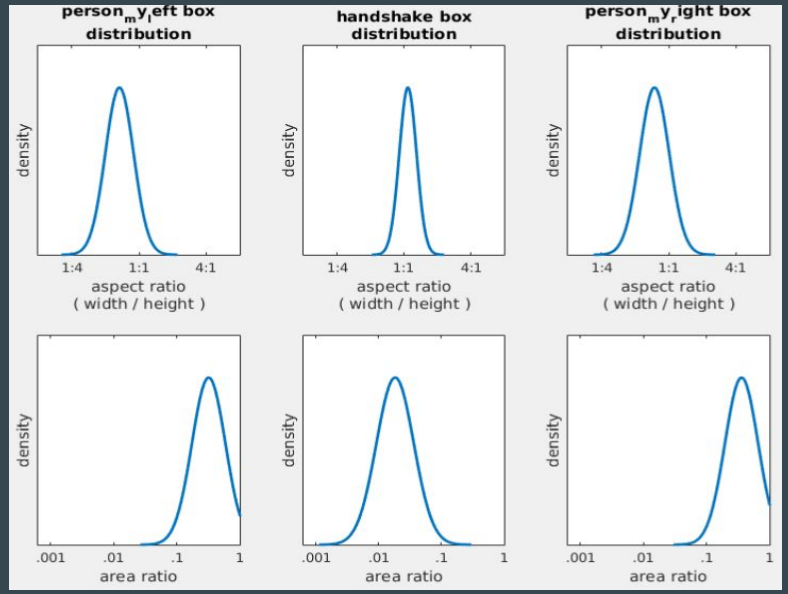


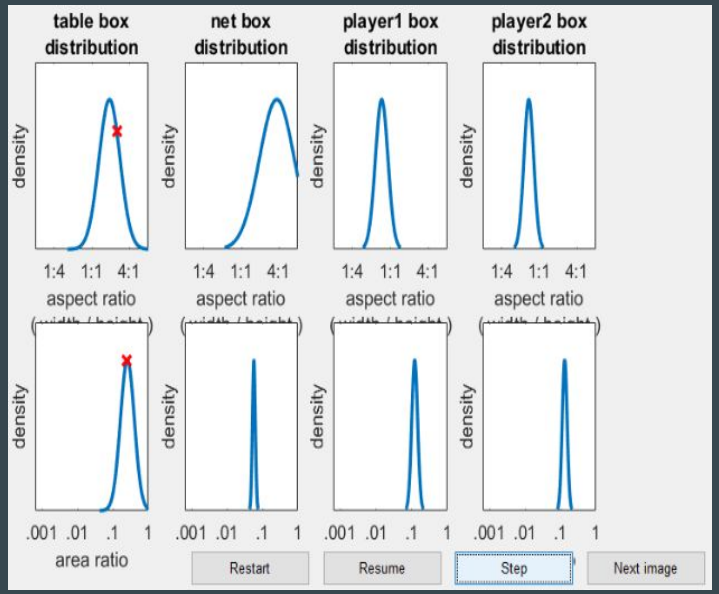
Handshaking





Ping Pong Playing





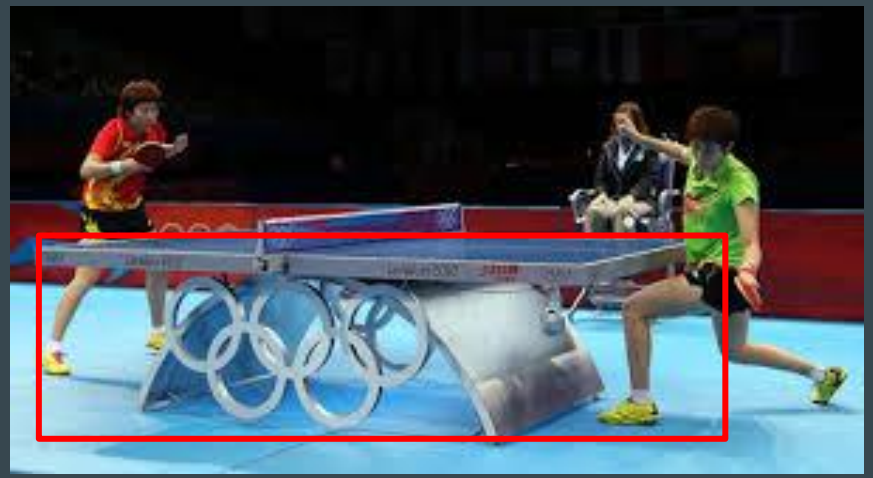
Problem 3: What object do we propose?

Handshaking

Person-left

	PL	PR	HS
Probability of object proposal	15%	15%	70%

Ping Pong



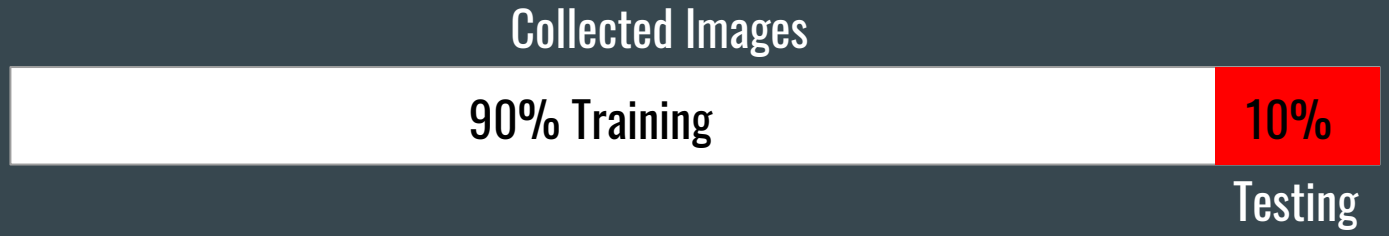
Table

	P1	P2	T	N
Probability of object proposal	27%	27%	40%	6%

Goal 2

Testing

Overview:

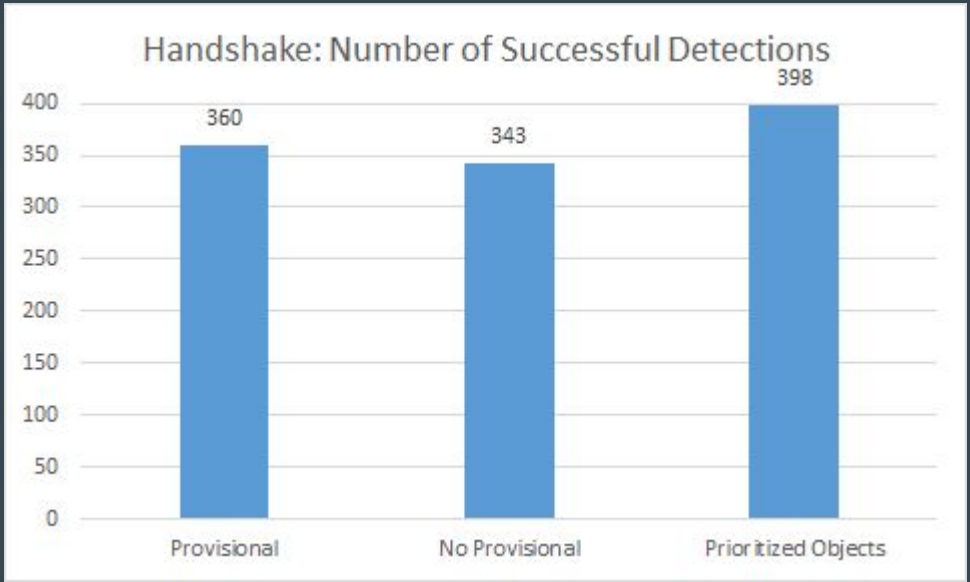


	Handshaking	Ping Pong
Training Set:	360 Images	270 Images
Test Set:	40 Images	30 Images

Repeated 10 times

Goal 2

Results: Handshaking



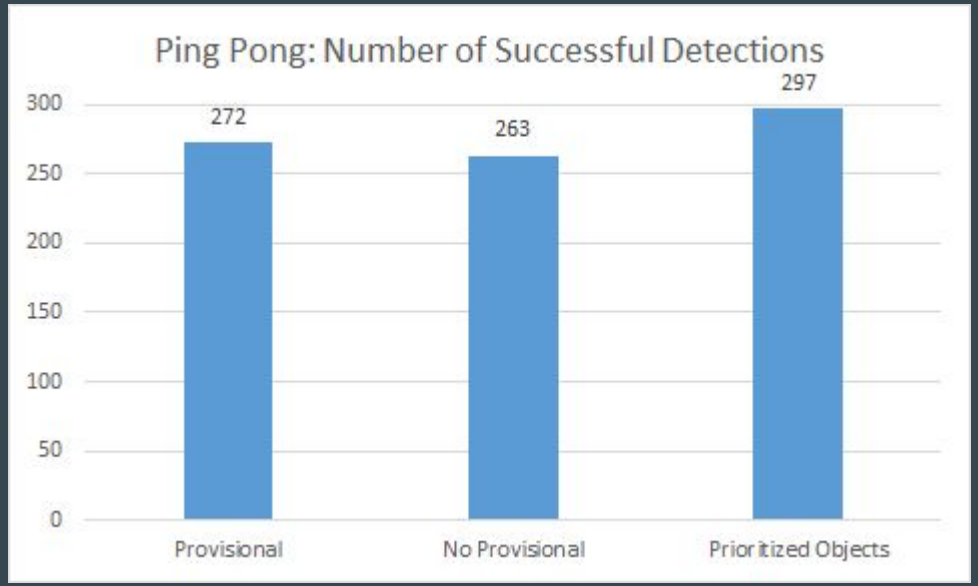
Percentage Successful

	Handshake
Provisional	90%
No Provisional	85%
Prioritized Objects	99%

(Successful = all objects localized within 1000 iterations)

Goal 2

Results: Ping Pong



Percentage Successful

	Ping Pong
Provisional	90%
No Provisional	87%
Prioritized Objects	99%

(Successful = all objects localized within 1000 iterations)

Goal 2

[Goal 3] Improve Object Recognition

We used a CNN-SVM to identify the object within a box

Goal 3

Conclusion

New Situations

- Successfully expanded *Situate* to two new situations.

Handshaking

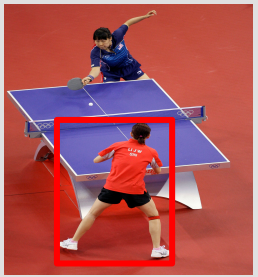
Ping Pong Playing

Improve Box Proposals

- Demonstrated *Situate*'s improved box proposal methods.
- Improved results with prioritized object detection.

Improve Object Recognition

- Implemented object identification without relying on ground truth boxes



Acknowledgements

Thank you for listening

Do you have any questions?

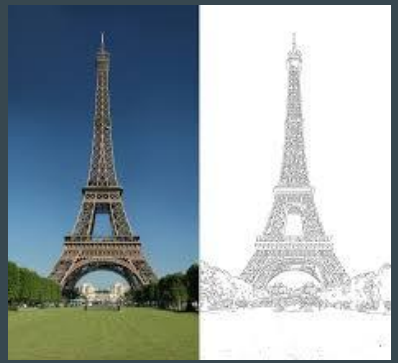
Object Recognition: Brief Overview

What is Object Recognition?

Motivation

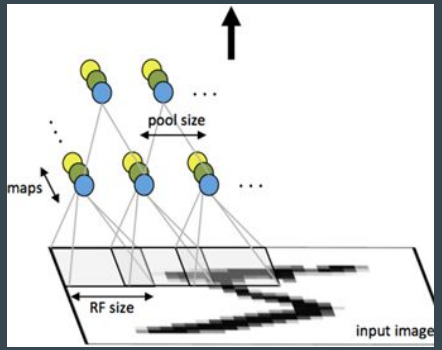
Convolutional Neural Network (CNN)

Scans Input Image for Features
(Corners, Edges, Shapes, etc.)



→

Inputs to a “Neural Network”
(Combines information from features)



→

Outputs a Classification
(Person Left, Person Right, Handshake)

“Not an Object”

Goal 2