Lab 12
Learning Picking

Song Chaoyang

Assistant Professor

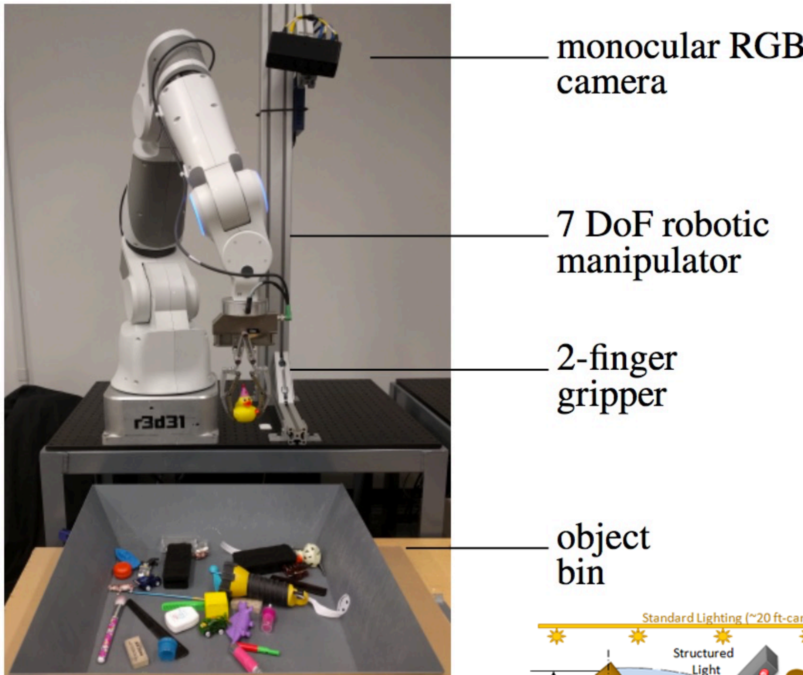
Department of Mechanical and Energy Engineering

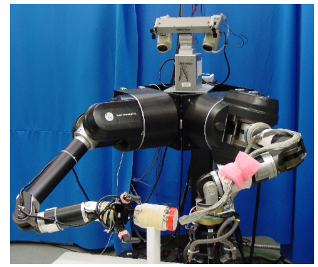
songcy@sustech.edu.cn

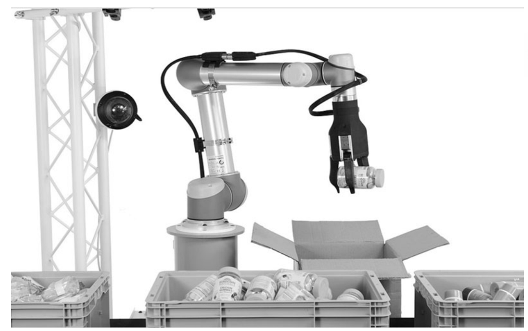
SUSTech
Southern University of Science and Technology

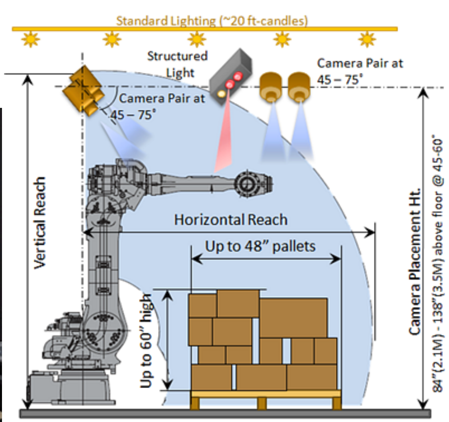
Common Setups of Vision-based Robotic Grasping

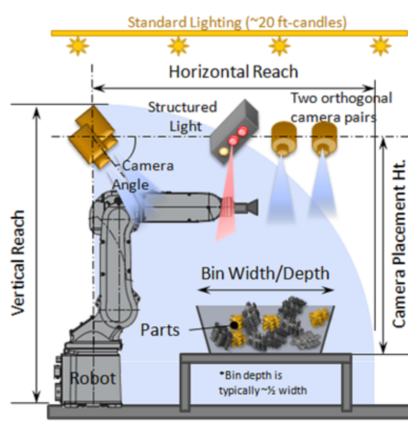
From academic research to industrial applications

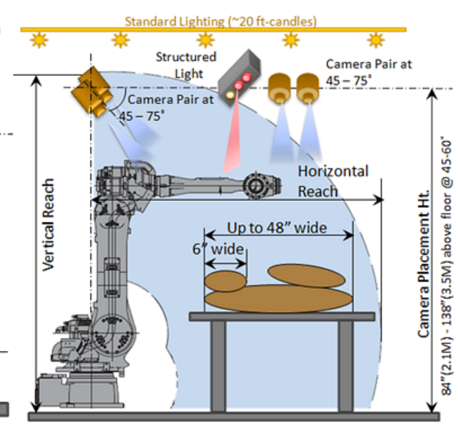












Challenges in Setting Up A Robotic Cell for Learning

Shareability and Reproducibility

- **H**ardware: still expensive
  - Arm: AUBO (~90K), UR5/Franka (~180K), UR10 (~400K)
  - Camera: Photoneo (~120K)
  - Gripper: Robotiq 3F (~250K)/2F (~50K), OnRobot (~50K)
  - Cheaper Alternatives:
    - Open Arms (~?), Realsense (~1.5K), Suction Cup (~1K)
- **S**oftware: steep learning curve & expensive
  - ROS, manufacturer specific, RoboDK, …
- **I**ntegration: time consuming to setup right (~H+S)
  - Cost of data per sample
    - Total Cost (H+S+I) / Cycle Time (~5-10sec): not negligible, yet

AncoraSIR.com

SUSTech
Southern University of Science and Technology

Cost Breakdown of Collecting 100k Samples

A *relatively cheap* version in an *ideal* setup

- Robot System (when everything is well-planned):
  - AUBO i5 + Suction Cup + Realsense + Mechanical + Man Hour
  - 90K+1K+1.5K+5K +5K < ¥100K per cell
- Yearly operation time per robot: ~3840 h/year (in motion)
  - 2 shifts of high use or three shifts of moderate use with moderate demands (a portion of the robot's motions with full load, a portion with tool only)
- Cost per robot cell on an hourly basis: ~ ¥ 26 cell/hour
  - Assume 10 sec per grasp
    - 1 hour usage = 60min/h * 60sec/min / 10sec/grasp = 360 grasp/h
- Cost per cell per grasp: ~ ¥ 0.07/grasp
  - Disregard electricity and maintenance, and assume every grasp counts

- To generate a dataset of 100K samples
  - 1 robot cell: 100K sample/dataset * ¥ 0.07/grasp = ¥ 7K/dataset
  - Cost of Time/Cell?: 100K sample/dataset / 360 grasp/h ~ 277h ~ 1mth ☹

AncoraSIR.com

SUSTech Southern University of Science and Technology

Cost Breakdown of Collecting 100k Samples

A *relatively cheap* version in an *ideal* setup

- Robot System (when everything is well-planned):
  - AUBO i5 + Suction Cup + Realsense + Mechanical + Man Hour
  - 90K+1K+1.5K+5K +5K < ¥100K per cell

What if

- A different robot arm is needed?
- A different end-effector is needed?
- A different camera system is needed?
- A new sensor is needed?

For example:

UR5 (*2) + OnRobot RG2 (*50) + OnRobot FT (+50K) + Photoneo M (*80) + Man Hour (*2)

- 90K*2+1K*50+(50K)+1.5K*80+5K +5K*2 ~ 415K (>100K***4**)

What if it is the case of an industrial arm?

Which part could possibly be remain unchanged?

AncoraSIR.com

Southern University of Science and Technology

Cost Breakdown of Collecting 100k Samples

A *relatively cheap* version in an *ideal* setup

- Robot System (when everything is well-planned):
  - AUBO i5 + Suction Cup + Realsense + Mechanical + Man Hour
  - 90K+1K+1.5K+5K +5K < ¥100K per cell
- Yearly operation time per robot: ~3840 h/year (in motion)
  - 2 shifts of high use or three shifts of moderate use with moderate demands (a portion of the robot's motions with full load, a portion with tool only)

Robots don't ask for benefits, but still need maintenance

Daily Checks:
- Clean sensors and optics within the robot cell and visually check component parts for damage before daily operation
- Check Mechanical Unit for grease/oil leakage or exudation

First 1-month (320 hours) Checks:
- Ventilation of the robot controller' cooling fans – keep these clean for efficient air flow

First 3-month (960 hours) Checks:
- Check Mechanical Unit Cables – free from severe kinks, pinch points, cuts or tears in the wire insulation, check for secure terminations
- Retighten external main mounting bolts
- Clean chips and debris from Mechanical Unit
- Check the end effector or end of arm tool and its' mounting bolts or screws, tighten appropriately
- Check robot connection cables (T.P. Cable, EE Cable, Vision Cables) – free from severe kinks, pinch points, cuts or tears in the wire insulation, check for secure terminations

1-year (3,840 hours) Checks:
- Grease balancer housing (if so equipped)

1.5-year (5,760 hours) Checks:
- Replace batteries in the mechanical unit (these are backup batteries to maintain the robot's factory alignments – don't let these batteries die or you may be in for some unpleasant work to re-zero your robot)

3-year (11,520 hours) Checks:
- Replace the grease of each axis. This is not an idle warning – use only the grease recommended by your robot's manufacturer – the robot controls are sensitive and grease can affect robot motion and operation.

4-year (15,360 hours) Checks:
- Replace the Mechanical Unit cables
- Replace the lithium battery in the CPU

https://motioncontrolsrobotics.com/how-often-do-robots-need-preventative-maintenance/

AncoraSIR.com

Southern University of Science and Technology

Cost Breakdown of Collecting 100k Samples

A *relatively cheap* version in an *ideal* setup

- Robot System (when everything is well-planned):
  - AUBO i5 + Suction Cup + Realsense + Mechanical + Man Hour
  - 90K+1K+1.5K+5K +5K < ¥100K per cell
- Yearly operation time per robot: ~3840 h/year (in motion)
  - 2 shifts of high use or three shifts of moderate use with moderate demands (a portion of the robot's motions with full load, a portion with tool only)
- Cost per robot cell on an hourly basis: ~ ¥ 26 cell/hour
  - Assume 10 sec per grasp
    - 1 hour usage = 60min/h * 60sec/min / 10sec/grasp = 360 grasp/h

Cost per robot cell on an hourly basis: ~ ¥ 26(*4) cell/hour

Cycle time is task-specific, and may require advanced control

- Advanced motion planning (additional cost, like mujin)
- Waiting for other machines to finish workflow (convey belt, CNC machines, …)

- In short, most robotic cells waits longer to execute each grasp

AncoraSIR.com

Southern University of Science and Technology

Cost Breakdown of Collecting 100k Samples

A *relatively cheap* version in an *ideal* setup

- Robot System (when everything is well-planned):
  - AUBO i5 + Suction Cup + Realsense + Mechanical + Man Hour
  - 90K+1K+1.5K+5K +5K < ¥100K per cell
- Yearly operation time per robot: ~3840 h/year (in motion)
  - 2 shifts of high use or three shifts of moderate use with moderate demands (a portion of the robot's motions with full load, a portion with tool only)
- Cost per robot cell on an hourly basis: ~ ¥ 26 cell/hour
  - Assume 10 sec per grasp
    - 1 hour usage = 60min/h * 60sec/min / 10sec/grasp = 360 grasp/h
- Cost per cell per grasp: ~ ¥ 0.07/grasp
  - Disregard electricity and maintenance, and assume every grasp counts

Cost per cell per grasp: ~ ¥ 0.07/grasp
- *4?
- Or *4*2?
- Or *4*?

AncoraSIR.com

Southern University of Science and Technology

Cost Breakdown of Collecting 100k Samples

A *relatively cheap* version in an *ideal* setup

- Robot System (when everything is well-planned):
  - AUBO i5 + Suction Cup + Realsense + Mechanical + Man Hour
  - 90K+1K+1.5K+5K +5K < ¥100K per cell
- Yearly operation time per robot: ~3840 h/year (in motion)
  - 2 shifts of high use or three shifts of moderate use with moderate demands (a portion of the robot's motions with full load, a portion with tool only)
- Cost per robot cell on an hourly basis: ~ ¥ 26 cell/hour
  - Assume 10 sec per grasp
    - 1 hour usage = 60min/h * 60sec/min / 10sec/grasp = 360 grasp/h
- Cost per cell per grasp: ~ ¥ 0.07/grasp
  - Disregard electricity and maintenance, and assume every grasp counts

Probably cost more than ¥ 7K/dataset of 100 K samples per robot cell,
And takes longer than 1mth in practice … ☹☹☹☹

- To generate a dataset of 100K samples
  - 1 robot cell: 100K sample/dataset * ¥ 0.07/grasp = ¥ 7K/dataset
  - Cost of Time/Cell?: 100K sample/dataset / 360 grasp/h ~ 277h ~ 1mth ☹

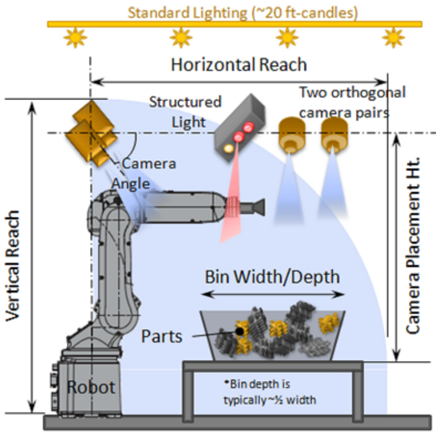
AncoraSIR.com

SUSTech
Southern University of Science and Technology

The Arcade Claw Machine Game

Structured grasping controlled by human player for a reward.

- Insert coins (cost per grasp), countdown starts (~30 sec)
- All planning done by human player
    - Eyes as vision sensor, observe from multiple angles to calibrate
    - Brain as the processing unit, weighing between probability of grasp against value of the prize
    - Hand on joystick to plan claw motion in x-y plane
    - Hand on button to execute the grasp
- Pick execution by the machine, dropping claw in z-axis
- Return the claw to drop coordinate, claws open and toys will drop if grasp success



Scene Segmentation
- What we are going to interact with …

Object Recognition
- Representation & Classification

Pose Estimation
- Object & Picker

Pick Planning
- Picker & Arm

Pick Execution
- MoveIt & PickIt

Advantages of the ACM

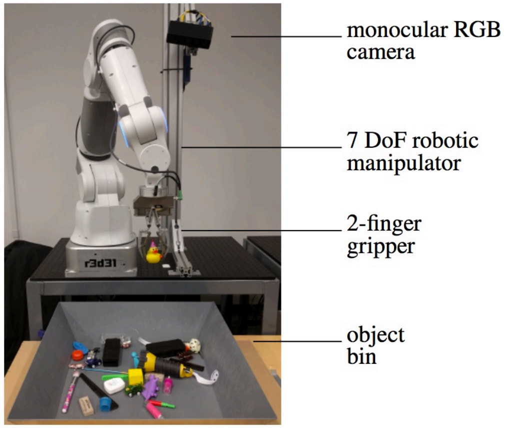
Accessibility and Flexibility towards Shareability and Reproducibility

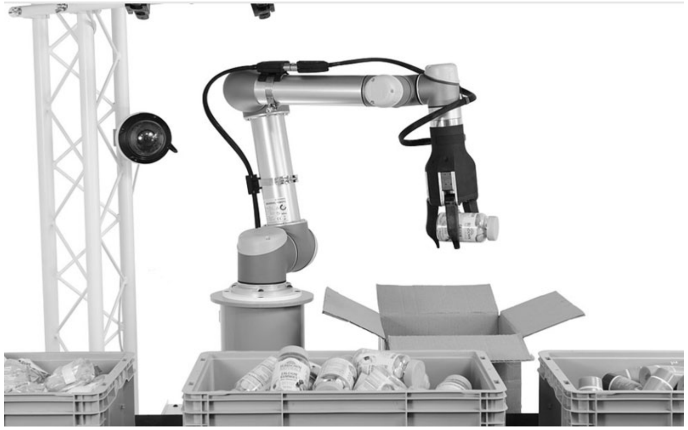
	Arcade Claw	Robotic Cell
Accessibility	Everyone	Professionals
Arm	Cartesian (3D)	Serial (6D)
Gripper	3 Finger (loose)	End-effectors
Vision	Human Eye	Camera
Control	Joystick	Controller
Decision	Human	Algorithms
Object	Almost anything	

- **Object Relocation**
  - Handling object from one location to another
  - Pick & Place | Machine Tending | Packing and Palletizing
- **Material Releasing**
  - Releasing material from the robot to the target location
  - Gluing | Dispensing | Welding | Screwdriving
- **Material Removal**
  - Removing material from the target object using the robot
  - Polishing | Grinding | Deburring
- **Information Gathering**
  - Collecting information using sensors attached to the robot
  - Quality Inspection

Robot Movement
- Fixed
- Patterned
- Continuous
- Changing
- Random

Critical Components
- End-Effector
- Vision System
- I/O Interfacing
- Conveyor Tracking
- Force-Torque Sensor
- External Jig
- Protective Suit





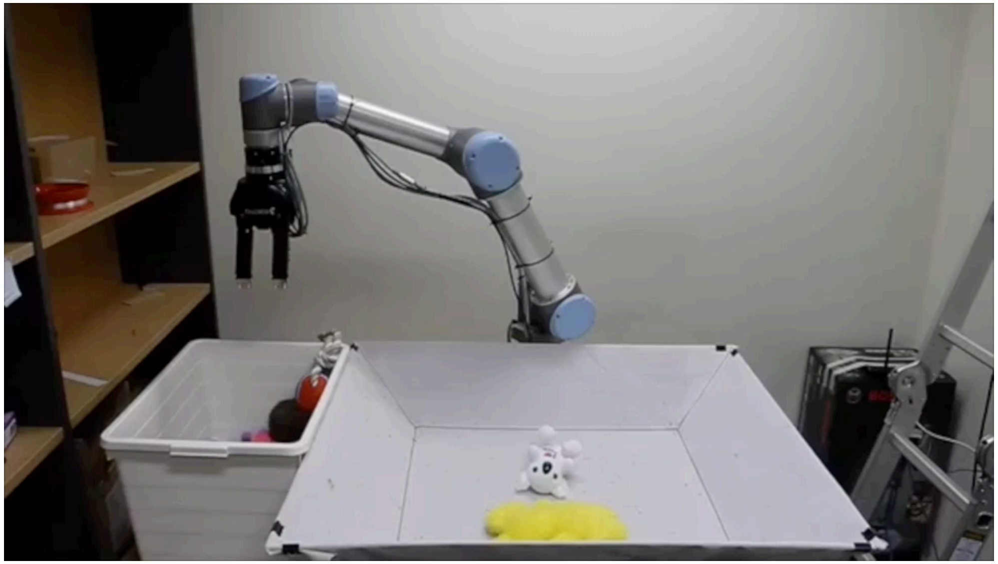
DeepClaw Setup

Task Decomposition

- Start
  - Preliminary Preparation
  - Safety Check
  - From Idle Position (x_idle, y_idle, z_idle=0, theta_idle, open)
- Operation
  - Blind Grips (~5k in 1 week)
  - Blind Learning with all Blind Grips attempted in 1 week
  - Daily Grips (~1k in 1 day)
  - Daily Learning with all Daily Grips attempted in 1 day
  - Repeat Daily Grips Daily Learning for 8 weeks
- End
  - Back to Idle Position (x_idle, y_idle, z_idle=0, theta_idle, open)
  - Learning and Grasping Result Summary

Start Stage

- System Setup
  - World => Pedestal => Arm => FT Sensor => Gripper
  - World => Desk => Tray => Objects (to be picked)
  - World => Desk => Bin => Objects (to be placed)
  - World => Camera Kinect (main rgd-d camera for training, focusing on the tray)
  - World => Camera LifeCam (auxiliary rgd camera for training, focusing on the tray)
  - World => Camera Canon (record the whole experiment, focusing on the whole robot operation)



Hardware

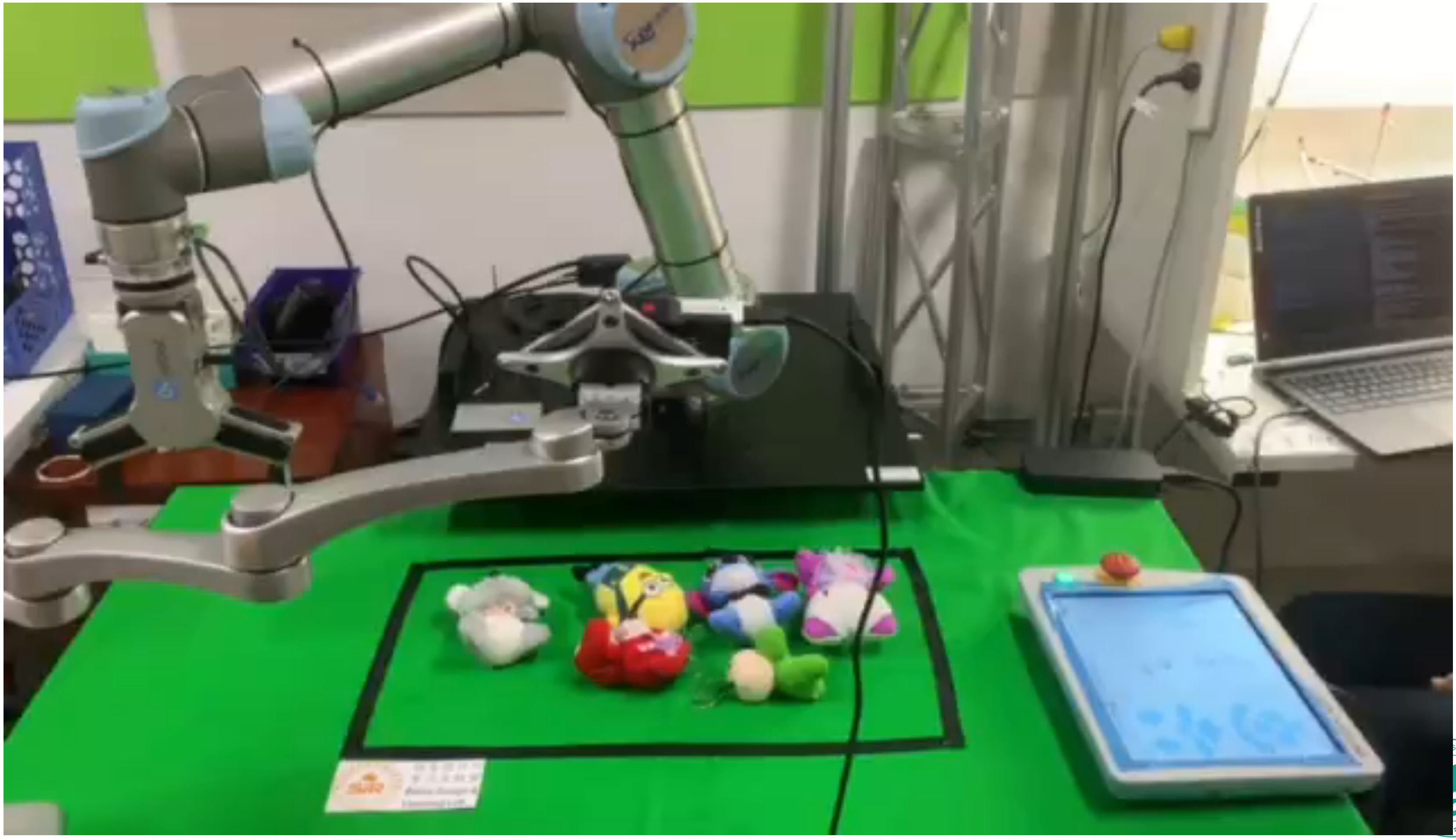
- Robot System (RS)
  - Arm: UR5 from Universal Robot
  - FT Sensor: FT300 from Robotiq
  - Gripper: Adaptive 3 Finger from Robotiq
  - Camera 1: Kinect for X-box One from Microsoft
  - Camera 2: LifeCam from Microsoft
  - Camera 3: Canon EOS M3
- Learning Computer (LC)
  - OS: Ubuntu Trusty 14.04.5
  - CPU: Intel Core i7 6800K Hex-Core 3.4GHz
  - GPU: Single 12GB NVIDIA TITAN X (Pascal)
  - RAM: 32GB Corsair Dominator Platinum 3000MHz (2 X 16GB)
  - SSD: 1TB Samsung 850 Pro Series

AncoraSIR.com

SUSTech
Southern University of Science and Technology

Project 4

https://github.com/ancorasir/BionicDL-CobotLearning-Project4.git



AncoraSIR.com

Hardware

- Robot System
  - UR5/…
  - Realsense
  - OnRobot RG6 gripper/…
- Learning Computer
  - Training: 4 x NVIDIA 1080Ti
  - Inference: can be run on cpu with lower speed
  - Tensorflow-gpu 1.12 + cuda 9.0 + cudnn 7.5
- Toys suitable for the gripper

AncoraSIR.com

SUSTech
Southern University of Science and Technology

The blind grasp dataset

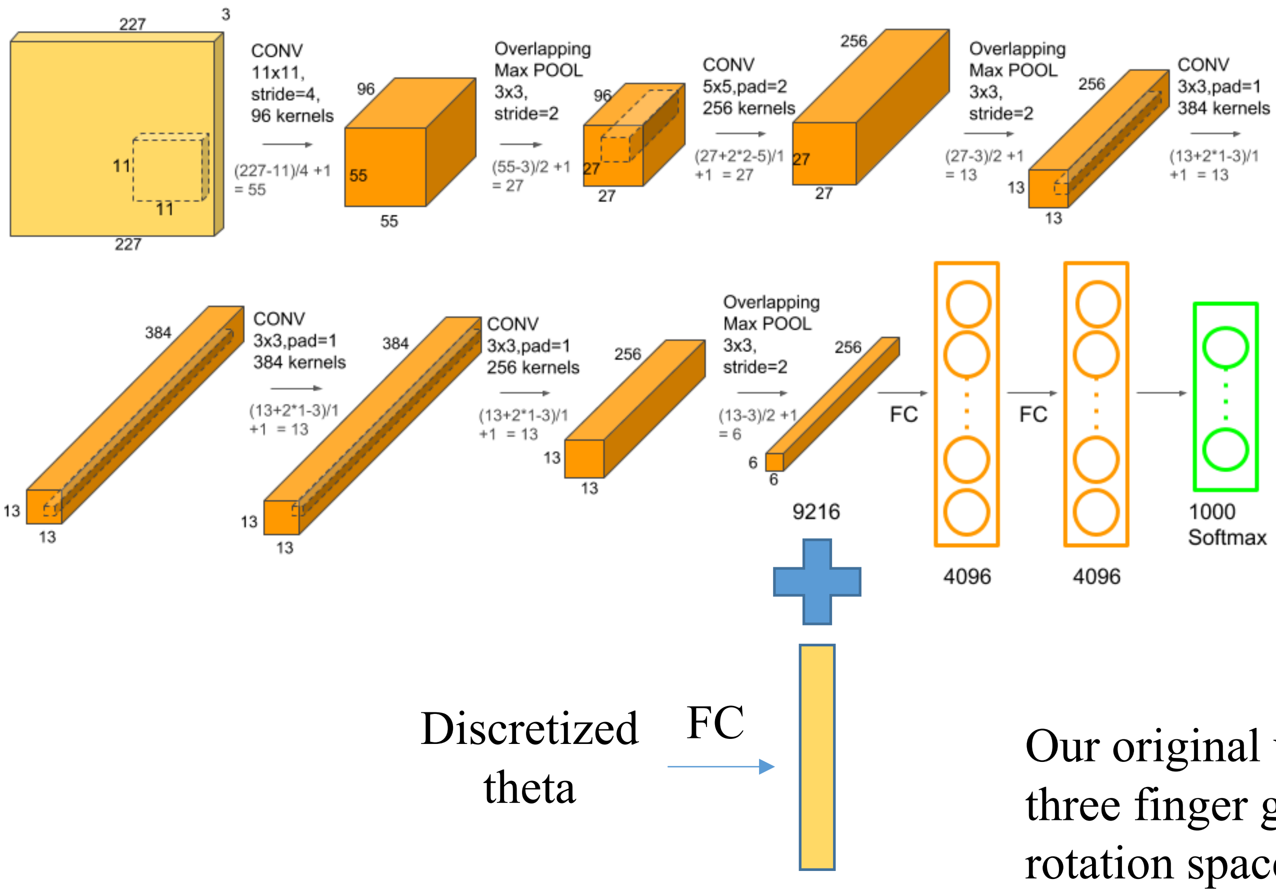
- Grasp pose (x, y, theta) is randomly generated.
  - Image: for each grasp is cropped such that the grasp point is at the center of the cropped image.
  - Rotation angle of gripper is recorded

AncoraSIR

SUSTech
Southern University of Science and Technology

GraspNet

Fine tuning modified Alexnet

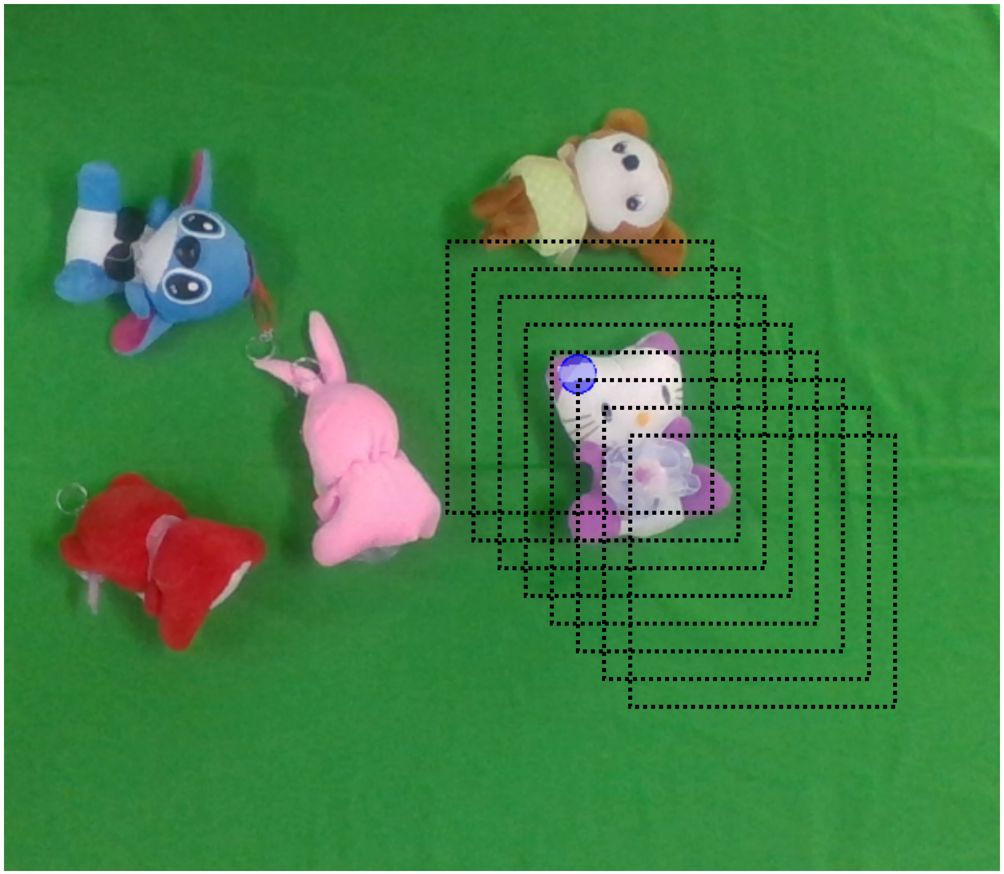


Our original version used a three finger gripper. So the rotation space 0 ~ 2pi is divided into 18 bins.

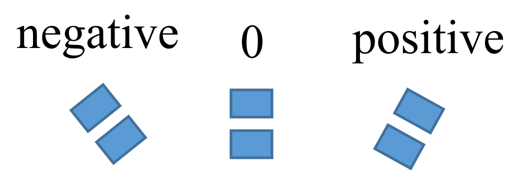
AncoraSIR.com

SUSTech
Southern University of Science and Technology

Predictor



- You may need to crop the original image from realsense if the view area is too big.
- Transform from u, v, theta to robot pose:
  - (u,v) to (x,y) though perspective transformation.
  - Theta to yaw. The theta predicted by the model is



Thank you!

Prof. Song Chaoyang

- Dr. Wan Fang (sophie.fwan@hotmail.com)

AncoraSIR.com