Lecture 13
Robot Learning

Song Chaoyang

Assistant Professor

Department of Mechanical and Energy Engineering

songcy@sustech.edu.cn

Review of data-driven grasping papers

- Dexnet
- Learning hand-eye coordination for robotic grasping with deep learning and large scale data collection
- Yolo

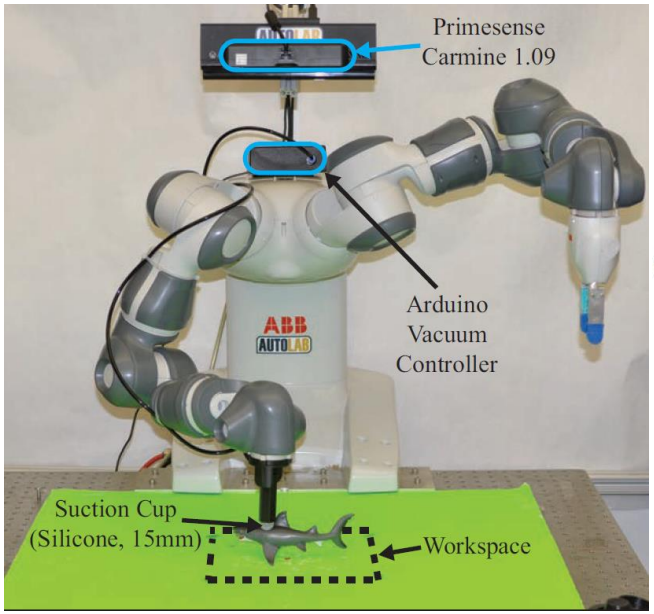
AncoraSIR.com

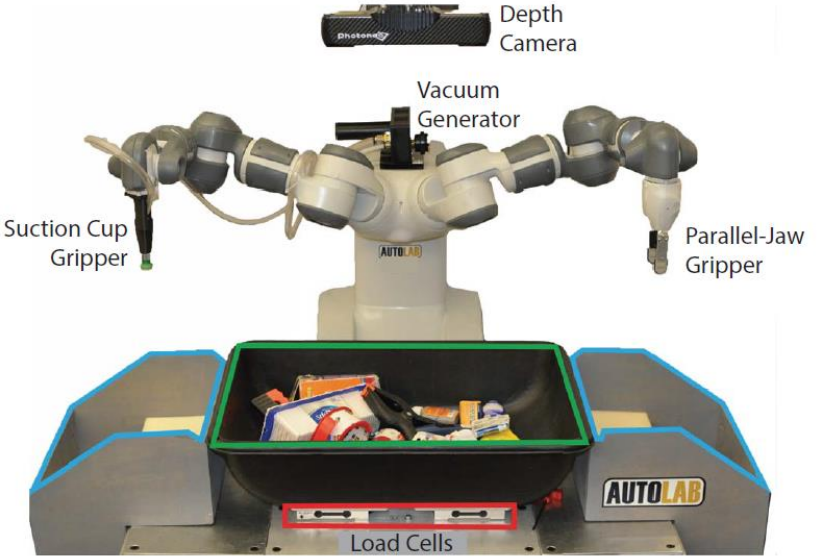
SUSTech
Southern University of Science and Technology

Dexnet

- Dexnet 2.0: Two finger gripper
- Dexnet 3.0: Suction cup end-effector
- Dexnet 4.0: Dual arm with one gripper and one suction cup end-effector

Executed Grasp

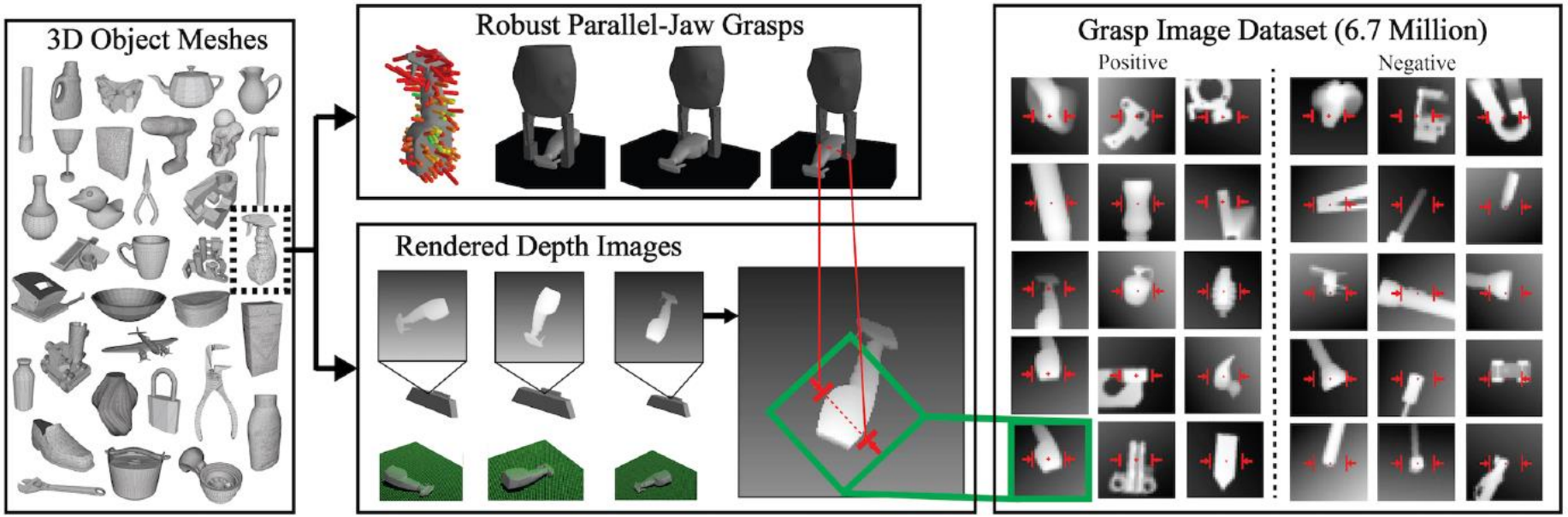




Dexnet 2.0

Dataset

- Dexnet 2.0 is a dataset associating 6.7 million point clouds and analytic grasp quality metrics with parallel-jaw grasps planned using robust quasi-static GWS analysis on a dataset of 1,500 3D object models



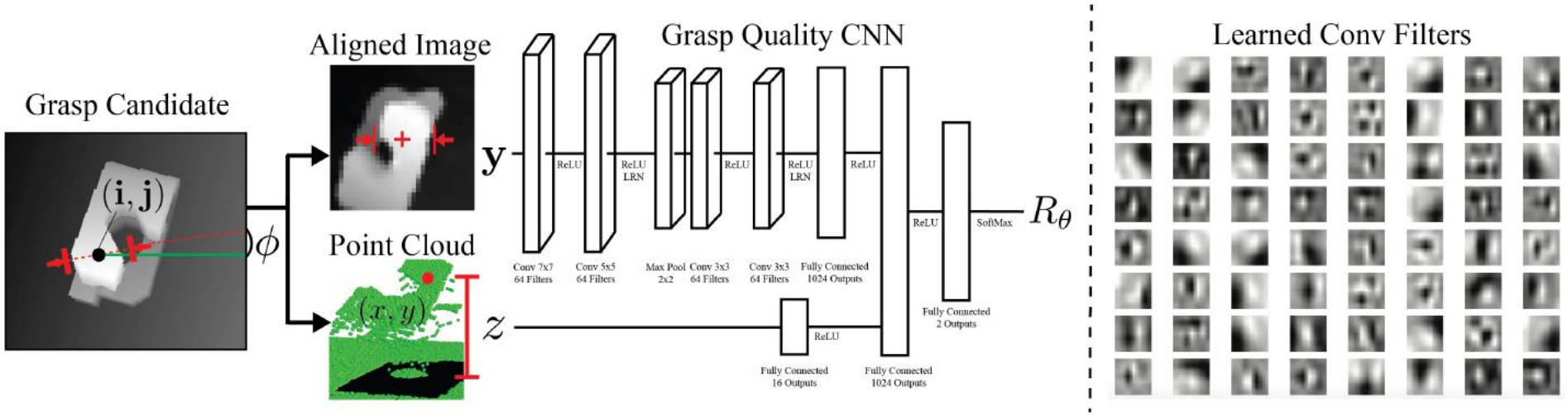
AncoraSIR.com

SUSTech
Southern University of Science and Technology

Dexnet 2.0

Grasp Quality Convolutional Neural Network (GQ-CNN).

- Planar grasp candidates u = (i, j, φ, z) are generated from a depth image and transformed to align the image with the grasp center pixel (i, j) and orientation φ.



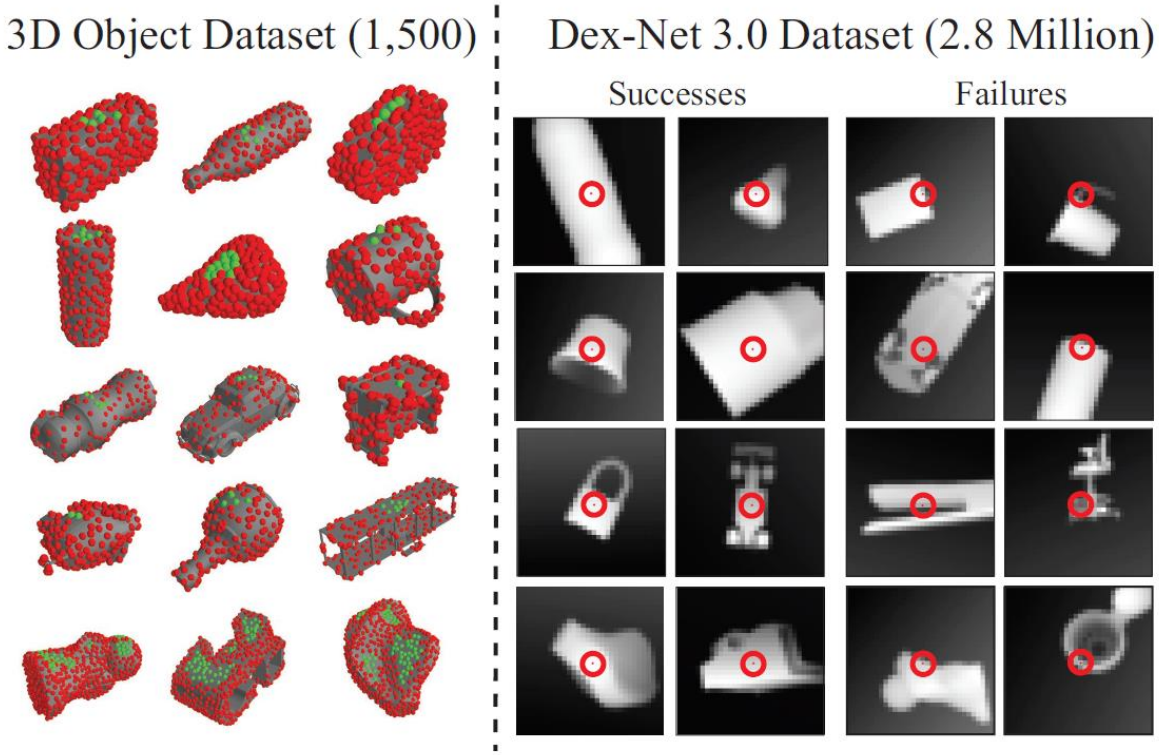
AncoraSIR.com

SUSTech
Southern University of Science and Technology

Dexnet 3.0

Dataset

- The Dex-Net 3.0 point cloud dataset contains approx. 2.8 million tuples of point clouds and suction grasps with robustness labels

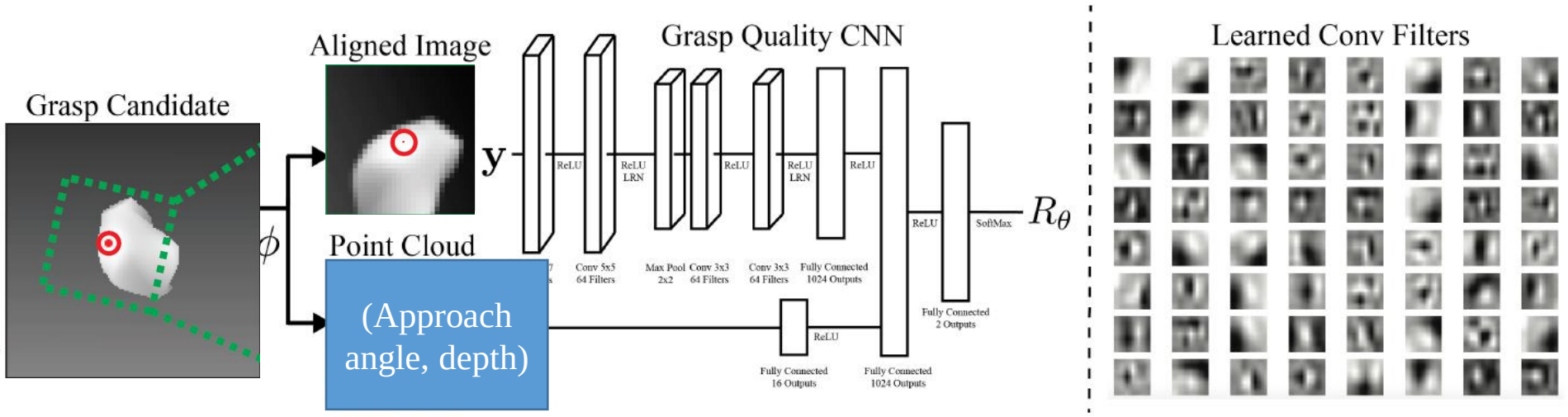


AncoraSIR.com

Dexnet 3.0

policy

- Uniform random sampling of candidate grasp from surface and use CEM to optimize sampling process.
- GQCNN takes the end-effector depth from the camera and orientation as input to a fully connected layer in a separate pose stream.



Dexnet 4.0

- the Dex-Net 4.0 policy consistently clears bins of up to 25 novel objects with reliability greater than 95% at a rate of more than 300 mean picks per hour.

AncoraSIR

Learning hand-eye coordination

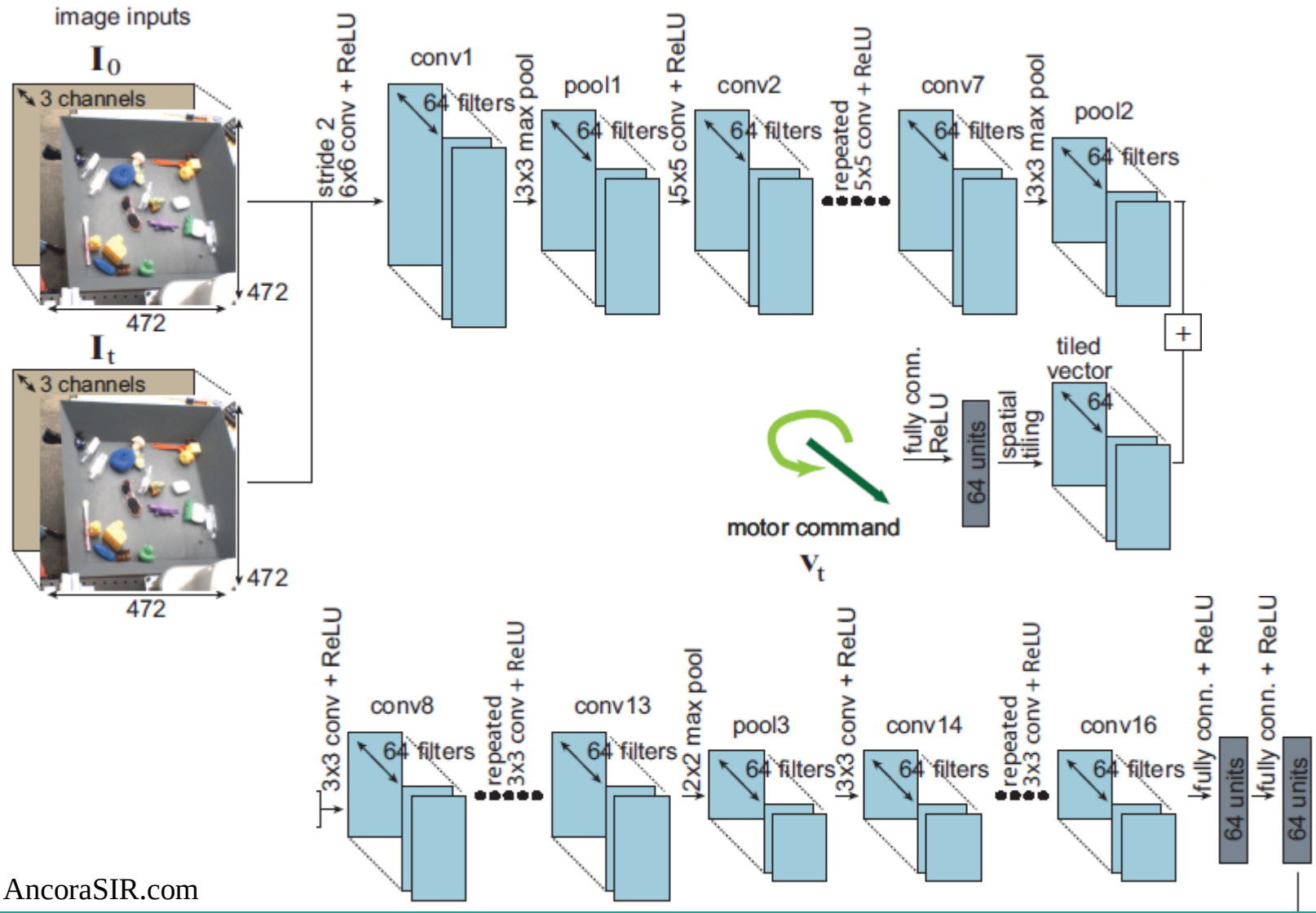
Dataset

- large-scale data collection setup, consisting of 14 robotic manipulators collected over 800,000 grasp attempts to train the CNN grasp prediction model.

AncoraSIR.com

SUSTech
Southern University of Science and Technology

Learning hand-eye coordination

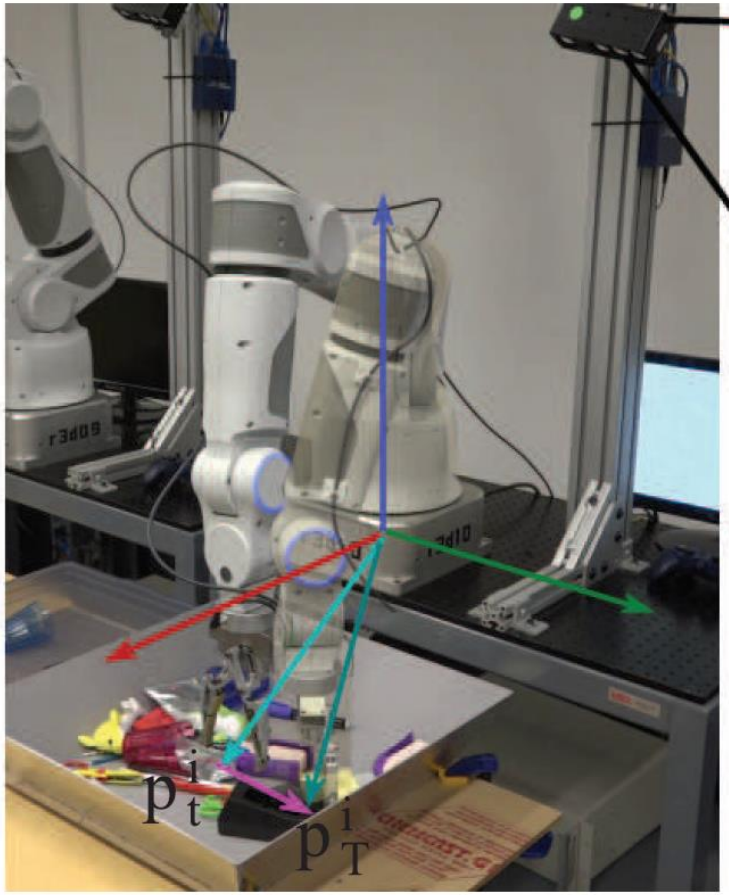


AncoraSIR.com

SUSTech
Southern University of Science and Technology

Learning hand-eye coordination

- Model predict motor command directly
- No hand eye calibration needed



I_t^i

samples $(\mathbf{I}_t^i, \mathbf{p}_T^i - \mathbf{p}_t^i, \ell_i)$

$\rightarrow \mathbf{p}_t^i$

$\rightarrow \mathbf{p}_T^i$

$\rightarrow \mathbf{v}_t^i$

AncoraSIR.com

SUSTech
Southern University of Science and Technology

Learning hand-eye coordination

Transferability

- Failure rates with Kuka IIWA robots

Training data	Average failure rate	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Random policy	89.79 ± 5.8%	88.24%	94.12%	91.18%	79.90%	88.73%	96.57%
2.7m non-Kuka images	76.80 ± 12.3%	87.25%	89.22%	75.49%	84.31%	60.29%	64.22%
3k Kuka images	84.80 ± 3.6%	82.35%	91.18%	82.84%	85.29%	81.37%	85.78%
30k Kuka images	74.75 ± 6.5%	75.49%	87.25%	73.04%	69.12%	71.08%	72.55%
300k Kuka images	41.50 ± 7.3%	50.00%	49.51%	42.65%	39.78%	32.35%	35.29%
3m Kuka images	34.31 ± 9.5%	46.08%	42.16%	31.86%	30.88%	19.12%	35.78 %
8m Kuka images	32.84 ± 13.8%	52.94%	38.24%	24.02%	30.88%	12.75%	38.24%
2.7m non-Kuka images and 3k Kuka images (finetuned)	70.92 ± 10.8%	76.96%	74.51%	76.47%	73.53%	49.02%	75.00%
2.7m non-Kuka images and 30k Kuka images (finetuned)	62.34 ± 8.4%	75.49%	60.29%	62.25%	66.18%	50.00%	59.80%
2.7m non-Kuka images and 300k Kuka images (finetuned)	35.62 ± 14.6%	43.63%	34.31%	36.27%	70.49%	29.90%	39.22%
2.7m non-Kuka images and 3m Kuka images (finetuned)	32.19 ± 10.3%	50.98%	24.02%	31.37%	25.00%	25.49%	36.27%
2.7m non-Kuka images and 8m Kuka images (finetuned)	25.65 ± 10.5%	46.57%	**25.00%**	**19.12** %	22.55%	18.14%	22.55%
2.7m non-Kuka images and 3k Kuka images (joint)	70.26 ± 13.4%	67.16%	60.78%	76.96%	76.96%	50.98%	88.73%
2.7m non-Kuka images and 30k Kuka images (joint)	48.45 ± 12.2%	58.82%	55.88%	59.80%	47.55%	28.92%	39.71%
2.7m non-Kuka images and 300k Kuka images (joint)	35.62 ± 16.6%	60.29%	48.04%	37.75%	29.41%	15.69%	22.55%
2.7m non-Kuka images and 3m Kuka images (joint)	27.61 ± 6.6%	30.39%	35.78%	31.37%	27.94%	23.04%	17.16%
2.7m non-Kuka images and 8m Kuka images (joint)	**22.82 ± 5.3%**	**30.39%**	27.12%	24.18%	**19.61%**	**17.65%**	**17.97%**

AncoraSIR.com

SUSTech
Southern University of Science and Technology

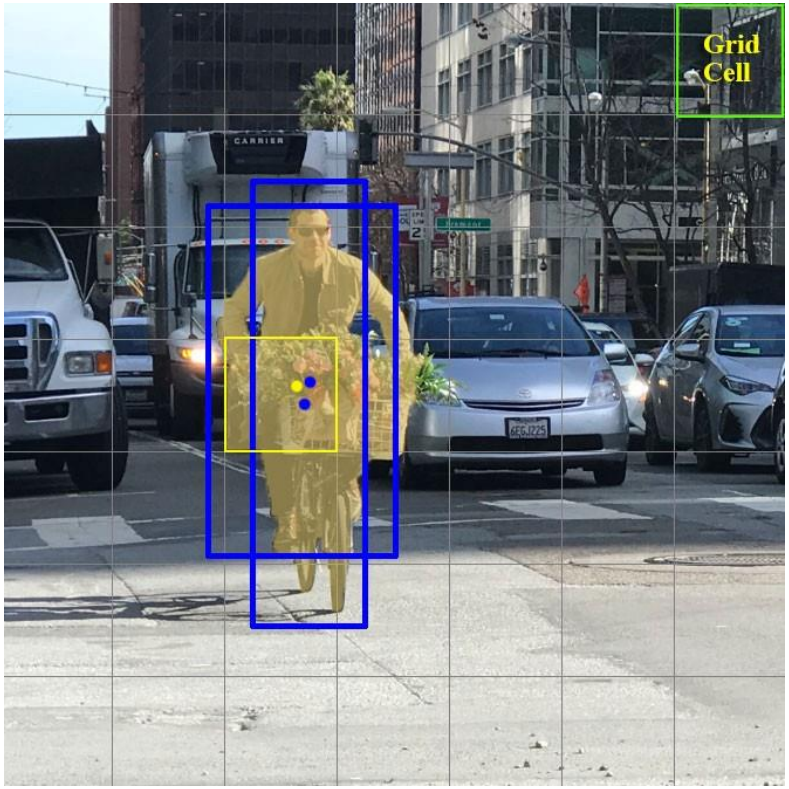
YOLO v3

You only look once

YOLO

Bounding Box Prediction

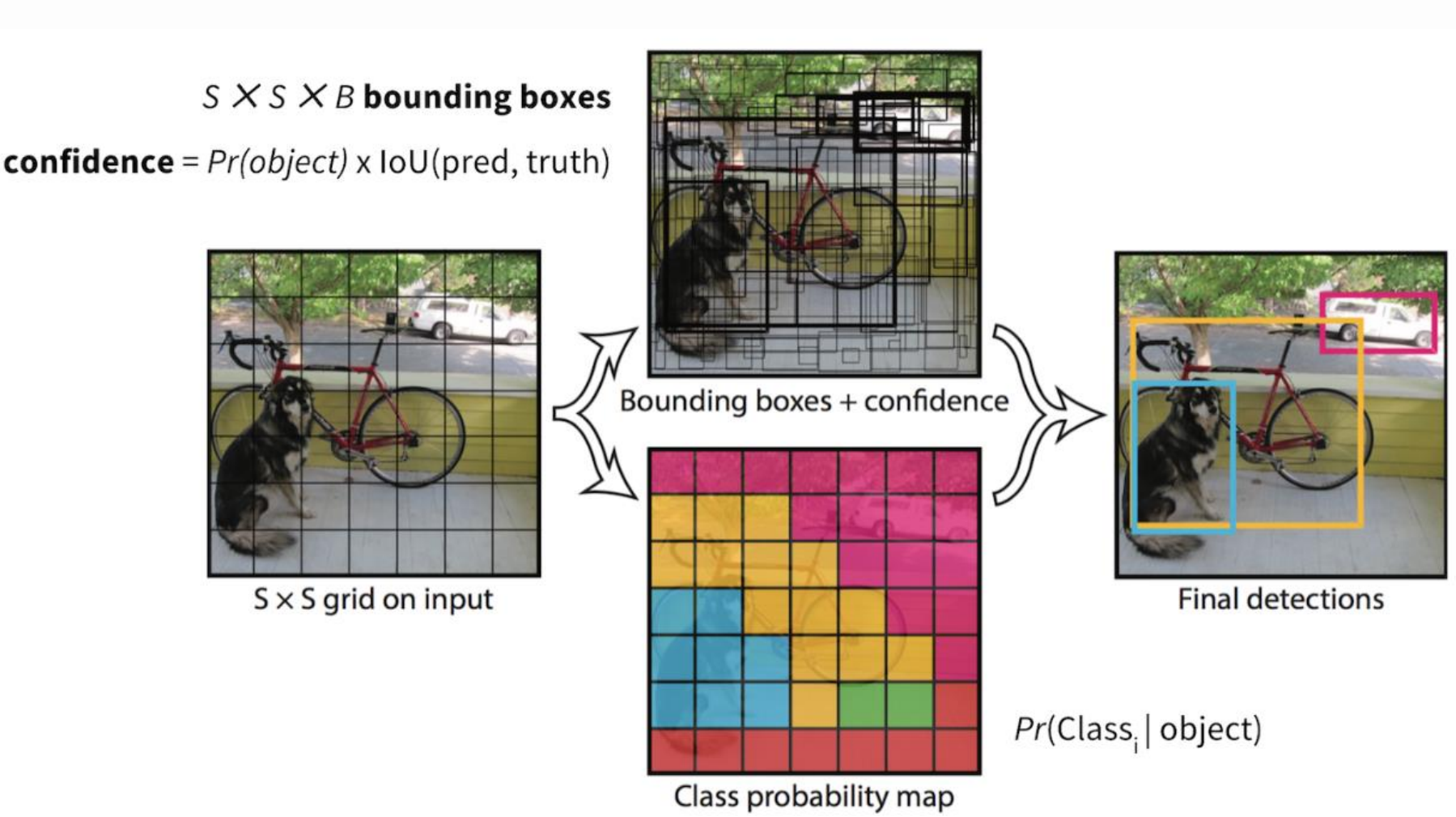
- Split an image into S×S cells. Each cell predicts only one object and
  - the coordinates of B bounding boxes (center x-coord, center y-coord, width, height) — (x,y,w,h)
  - a confidence score indicates the likelihood that the cell contains an object
  - a probability of object class conditioned on the existence of an object in the bounding box
  - The total prediction values for one image is S×S×(5B+C),



AncoraSIR.com

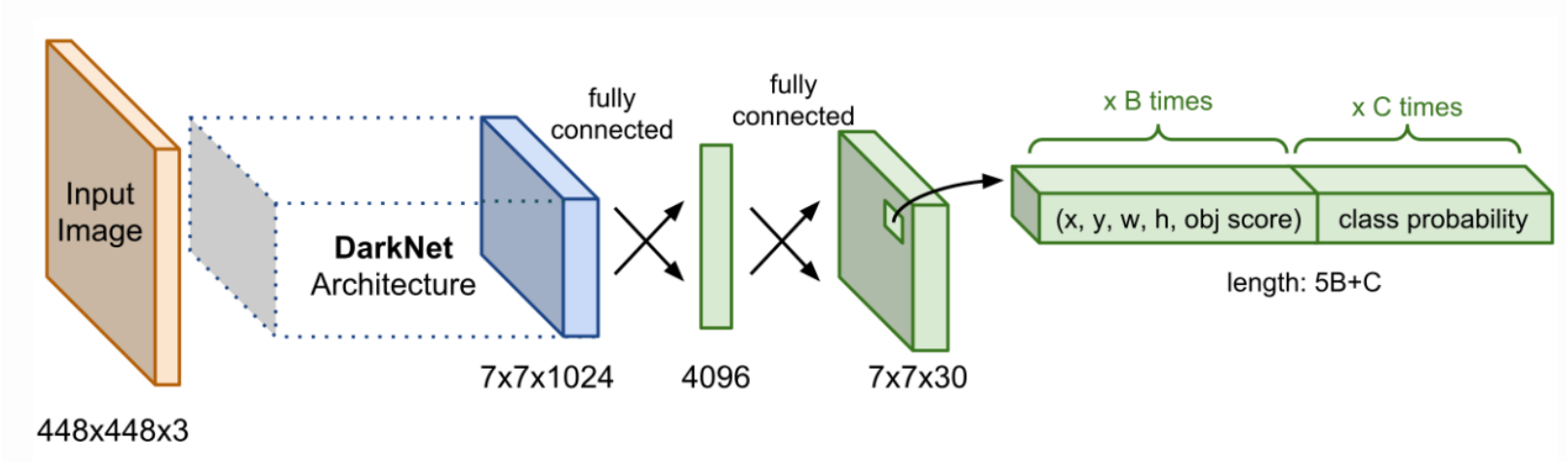
YOLO

Bounding Box Prediction



YOLO

Network Architecture



Website: https://github.com/YunYang1994/tensorflow-yolov3
https://github.com/srp-31/Data-Augmentation-for-Object-Detection-YOLO-

AncoraSIR.com