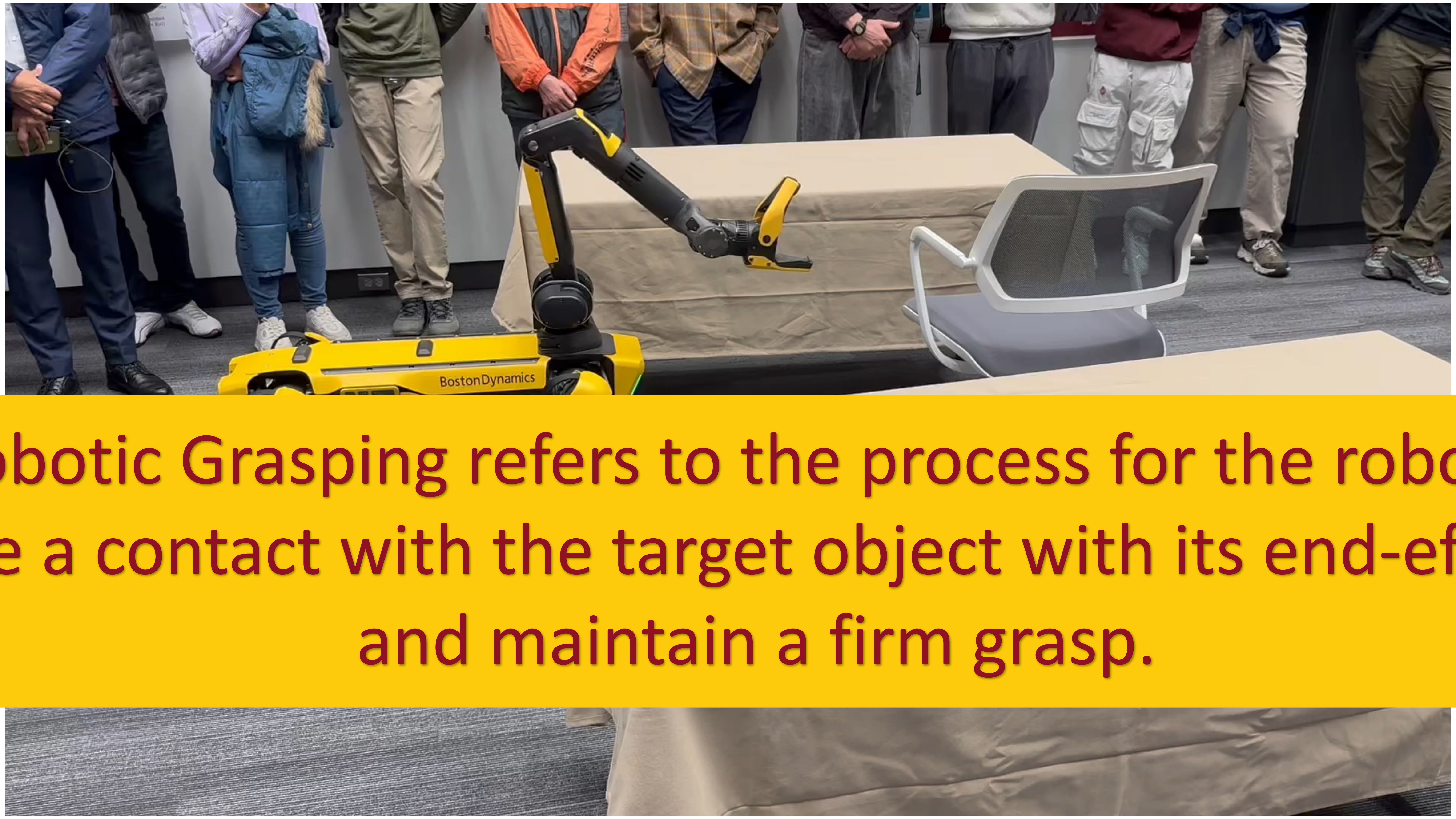
DeepRob

Lecture 13
Robot Grasp Learning
University of Minnesota

Slides created by Xun Tu and Karthik Desingh, Fall 2024

Robotic Grasping

- What is robotic grasping?

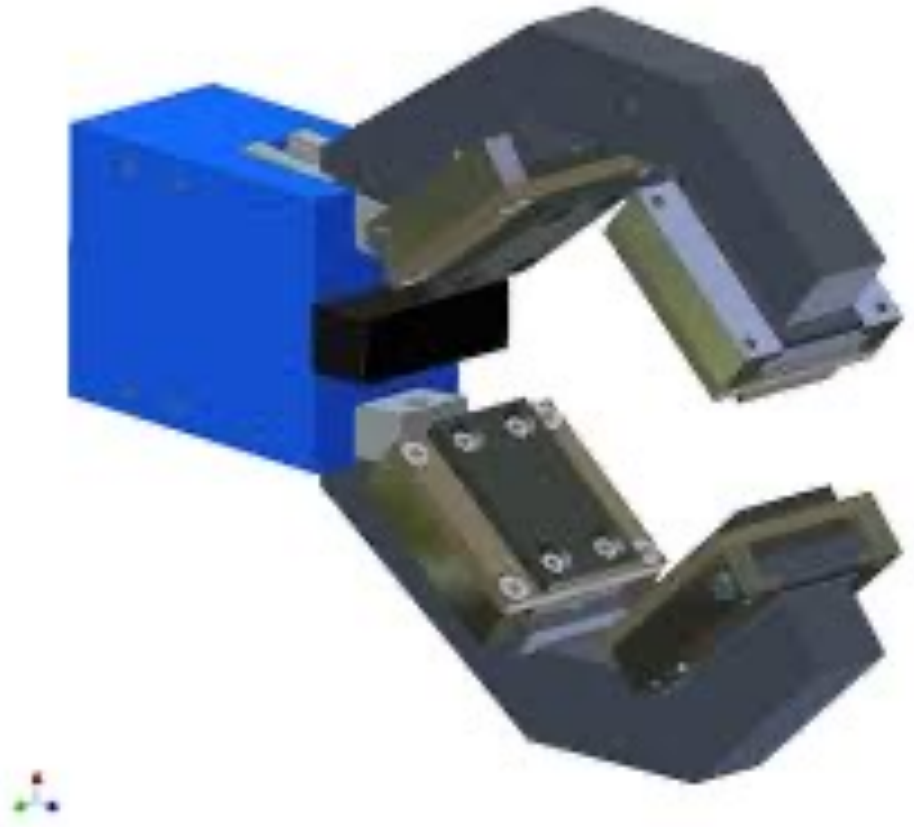


Robotic Grasping refers to the process for the robot to make a contact with the target object with its end-effector, and maintain a firm grasp.

Robotic Grasping – End-effector/Gripper

Parallel Gripper

https://onrobot.com/en/products/2fg7



Jaw Gripper

https://www.agi-automation.com/design-guidelines-for-pneumatic-gripper/

Dexterous Hand Gripper

https://www.shadowrobot.com/

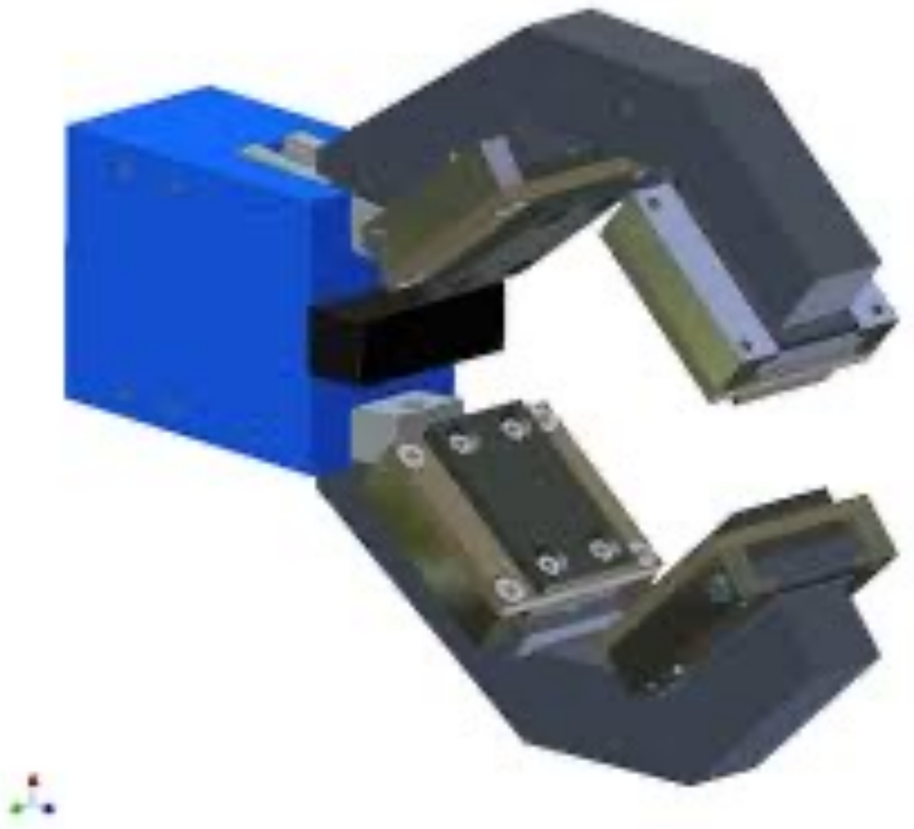
Suction Gripper

https://test.tm-robot.com/en/product/robotiq-vacuum-gripper-epick/

Robotic Grasping – End-effector/Gripper

Parallel Gripper

https://onrobot.com/en/products/2fg7



Jaw Gripper

https://www.agi-automation.com/design-guidelines-for-pneumatic-gripper/

Dexterous Hand Gripper

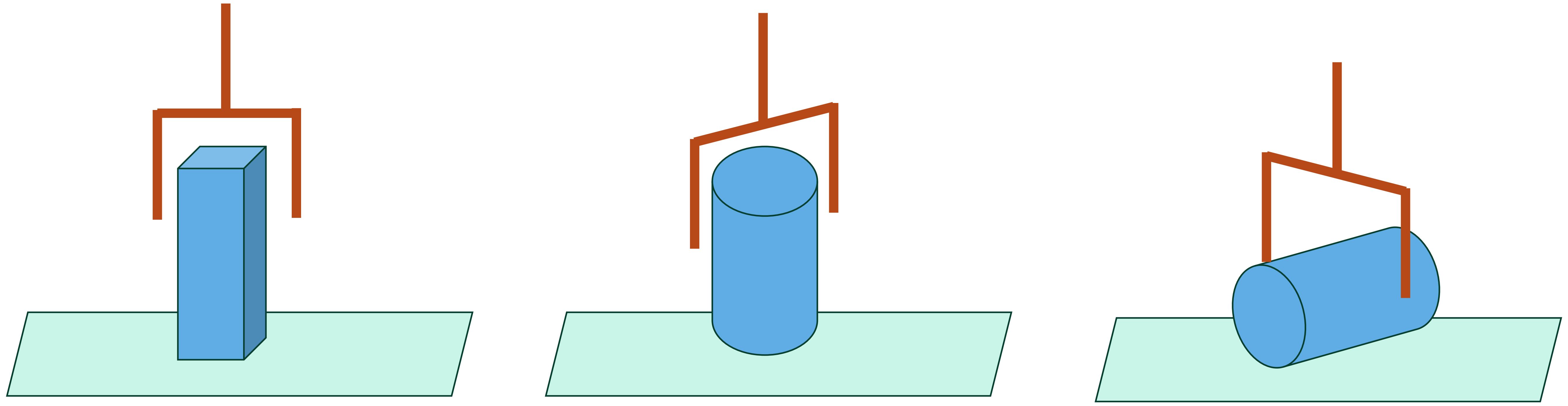
https://www.shadowrobot.com/

Suction Gripper

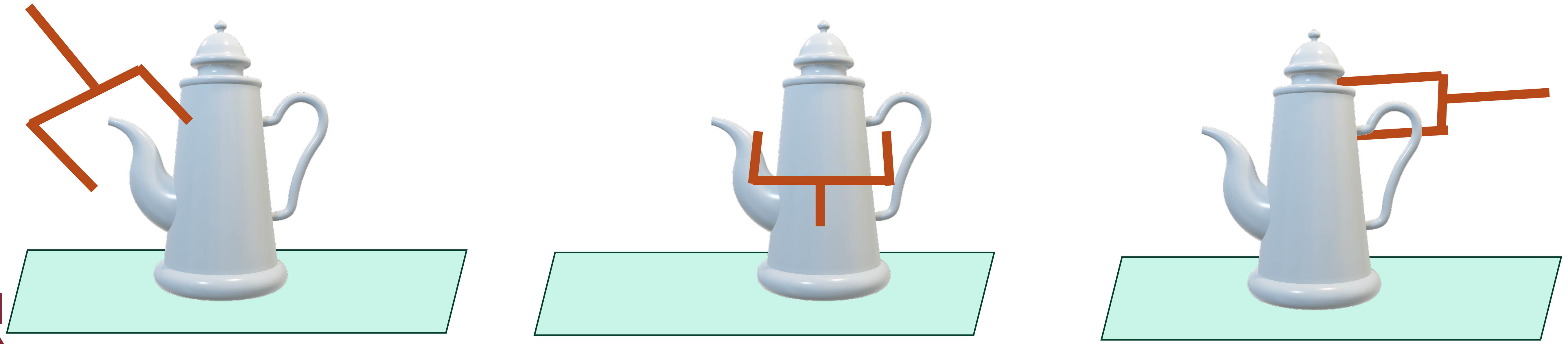
https://test.tm-robot.com/en/product/robotiq-vacuum-gripper-epick/

Robotic Grasping - Grasp Pose

- Grasping in SE(2) pose



- Grasping in SE(3) pose



Grasping in $SE(2)$ pose

Grasping in SE(2) pose

Action Direction: Top-down

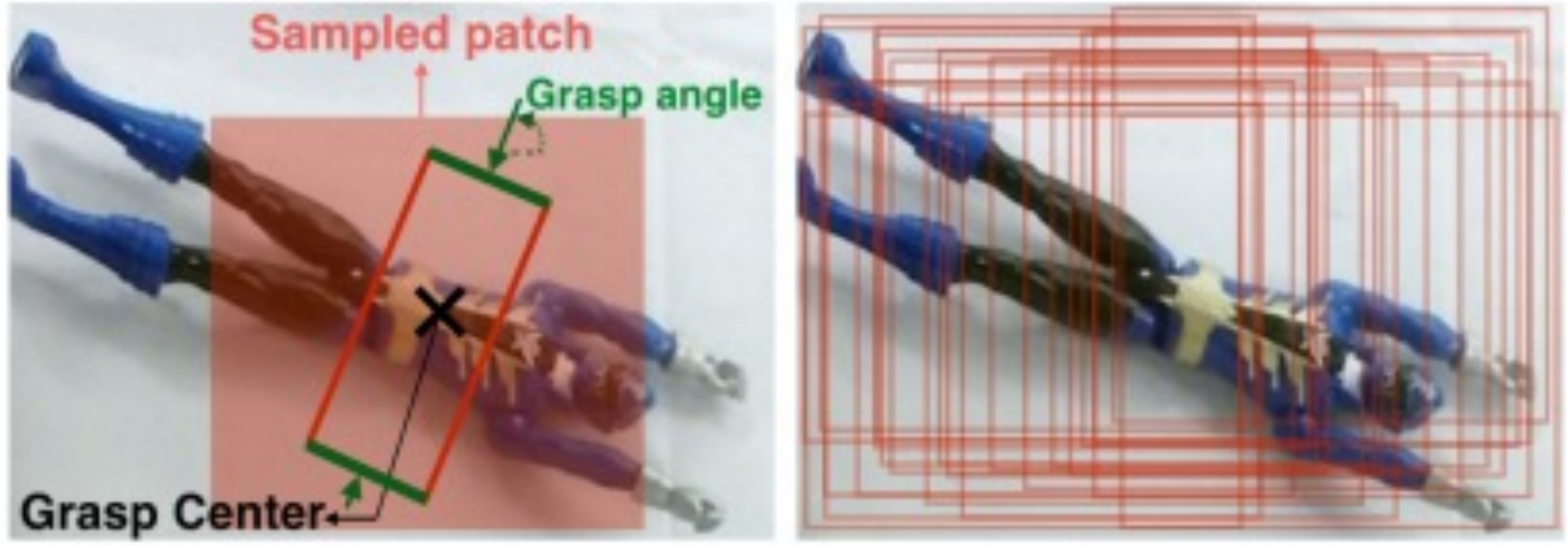
Input: RGB-D images, point cloud

Output: $(\underbrace{x, y}_{\text{Location}}, \underbrace{\theta}_{\text{Rotation Angle}})$

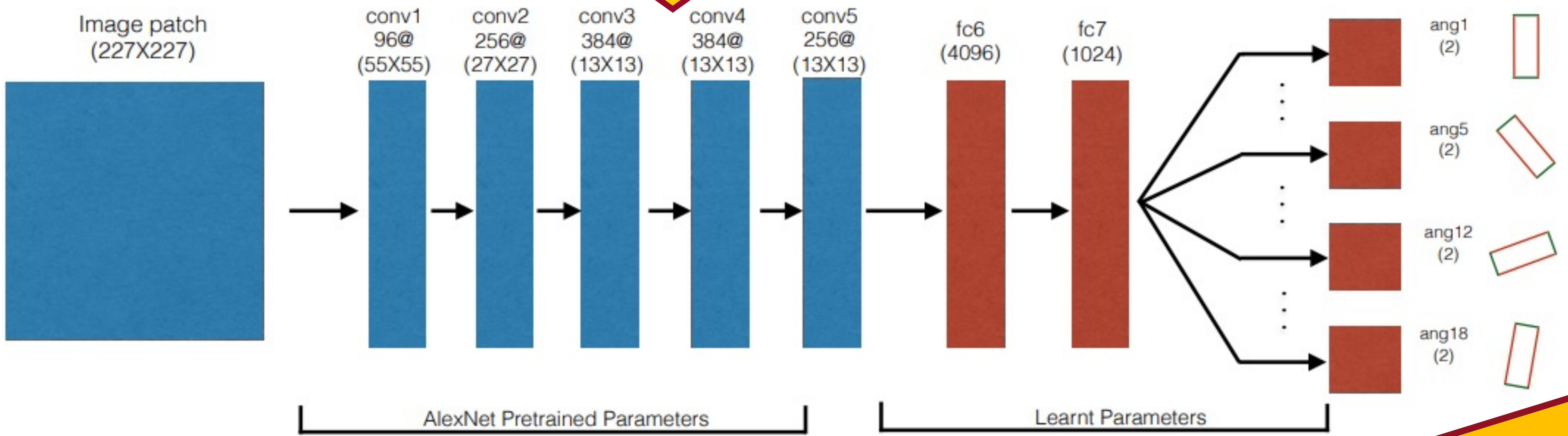
Grasping in SE(2) pose

- Supersizing Self-supervision: Learning to Grasp from 50K Tries and 700 Robot Hours
- Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics
- Sample Efficient Grasp Learning Using Equivariant Models

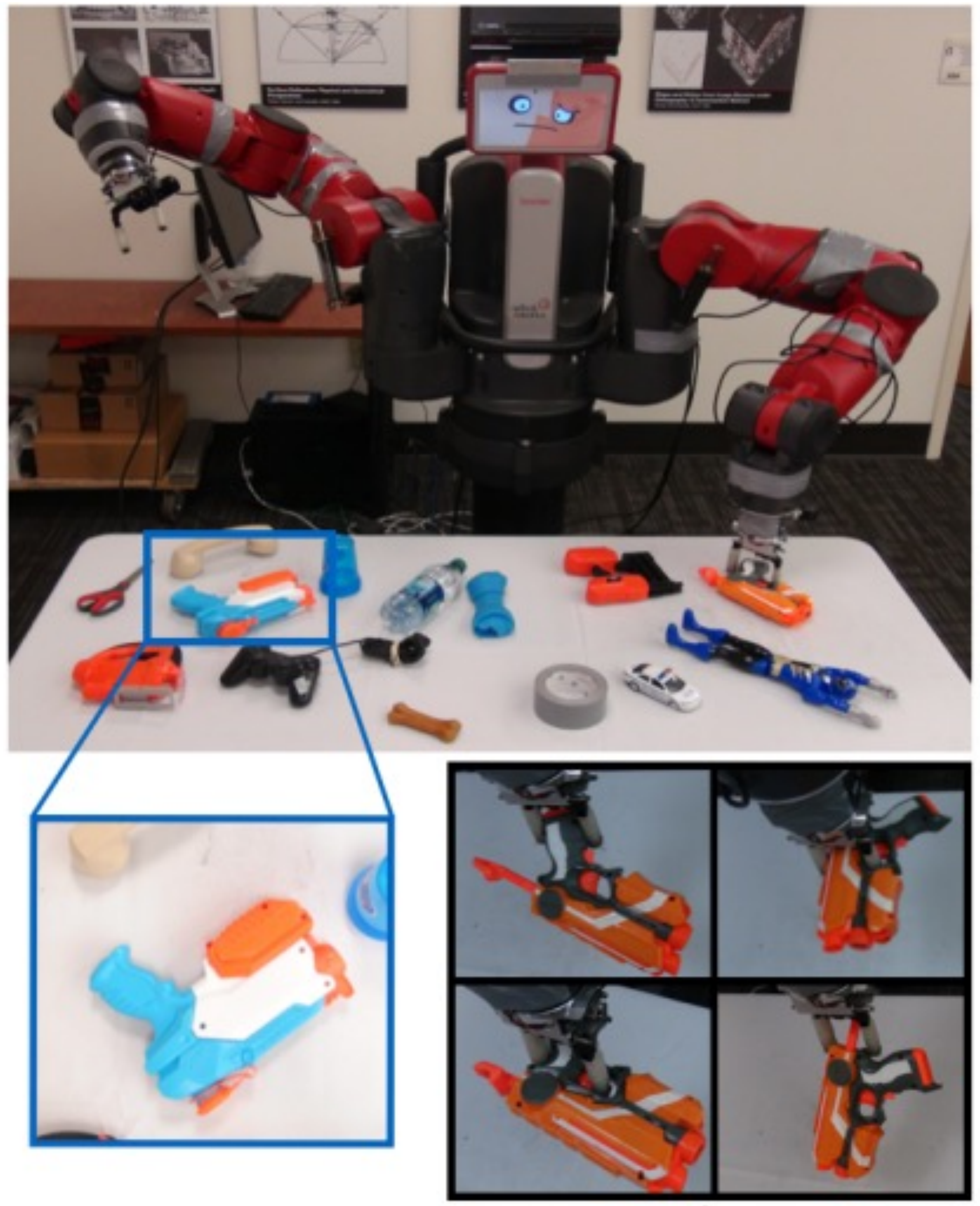
Supervising Self-supervision: Learning to Grasp from 50K Tries and 700 Robot Hours



Sample a patch on the input image



Resize the patch to 227x227 and predict a grasp score for each of the rotation angle $(10^o, 20^o, \cdots)$



Execute the grasp pose by a real-robot

L. Pinto and A. Gupta, 'Supersizing self-supervision: Learning to grasp from 50K tries and 700 robot hours', in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 3406–3413.

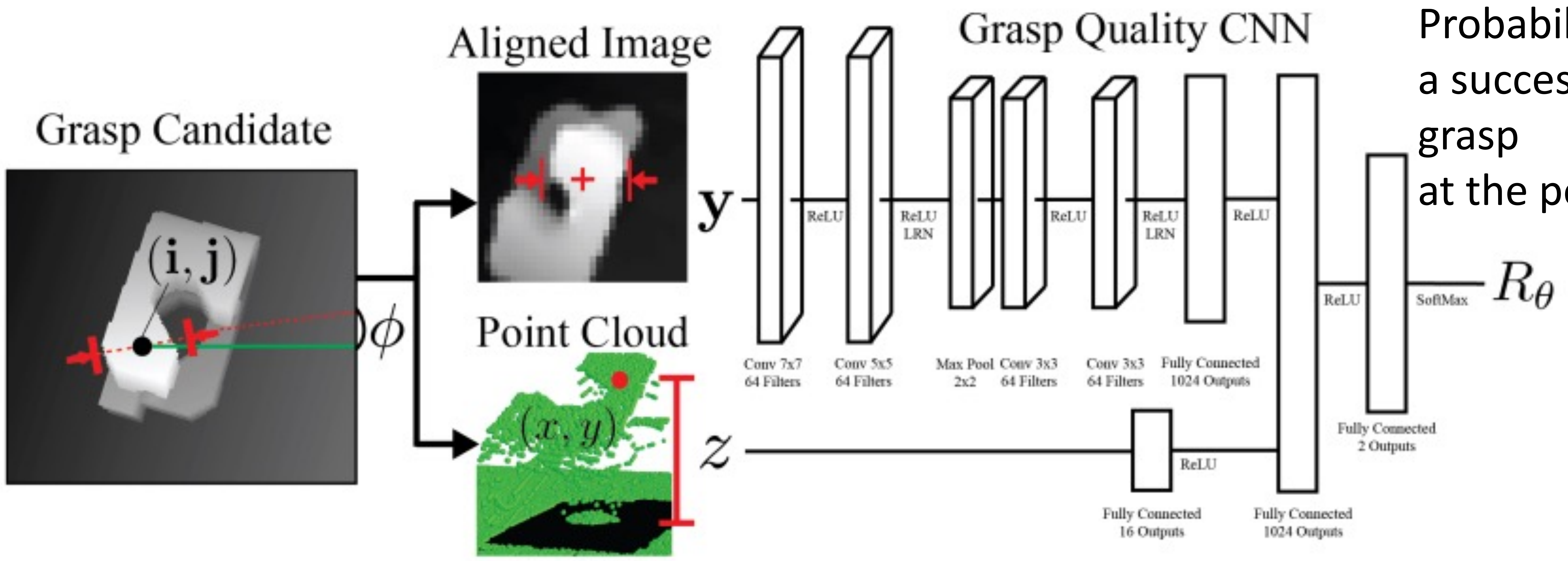
Dex-Net 2.0

J. Mahler et al., "Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics," Robotics: Science and Systems (RSS), 2017.

Dex-Net 2.0

Architecture

Generate grasp candidates by sampling antipodal contact points on the point cloud

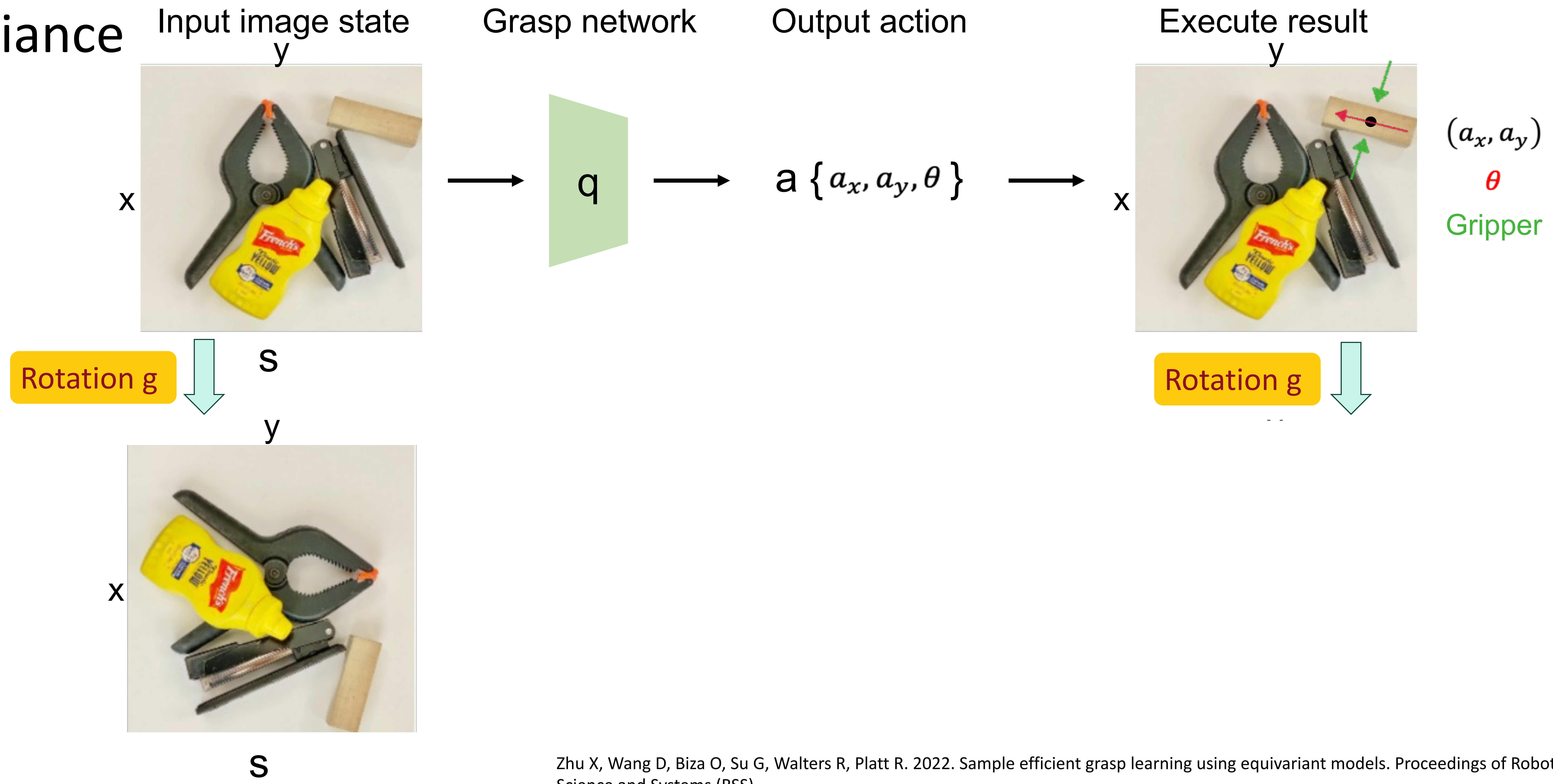


Probability of a successful grasp at the pose

J. Mahler et al., "Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics," Robotics: Science and Systems (RSS), 2017.

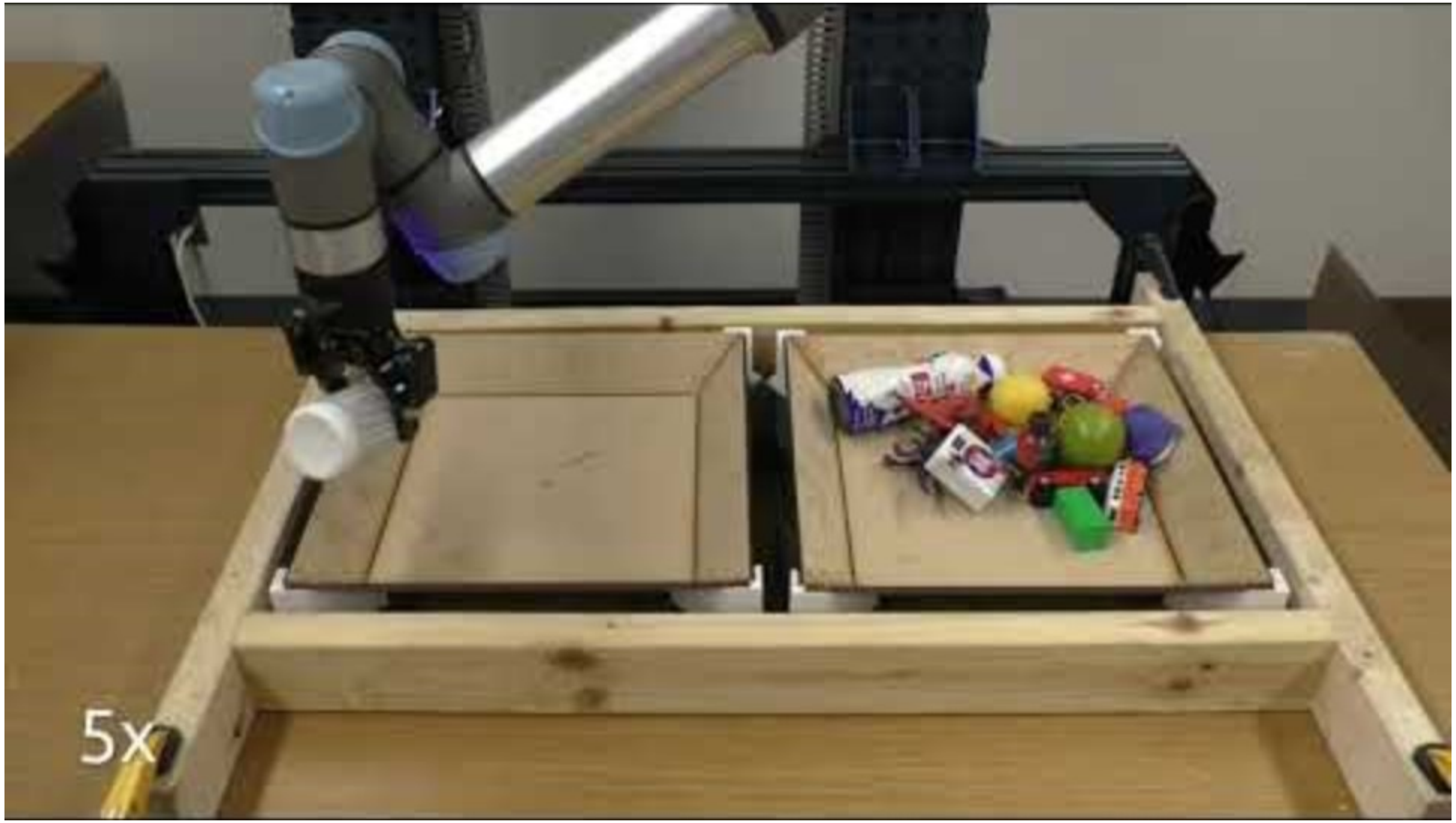
Sample Efficient Grasp Learning Using Equivariant Models

- Equivariance



Zhu X, Wang D, Biza O, Su G, Walters R, Platt R. 2022. Sample efficient grasp learning using equivariant models. Proceedings of Robotics: Science and Systems (RSS)

Sample Efficient Grasp Learning Using Equivariant Models



Zhu X, Wang D, Biza O, Su G, Walters R, Platt R. 2022. Sample efficient grasp learning using equivariant models. Proceedings of Robotics: Science and Systems (RSS)

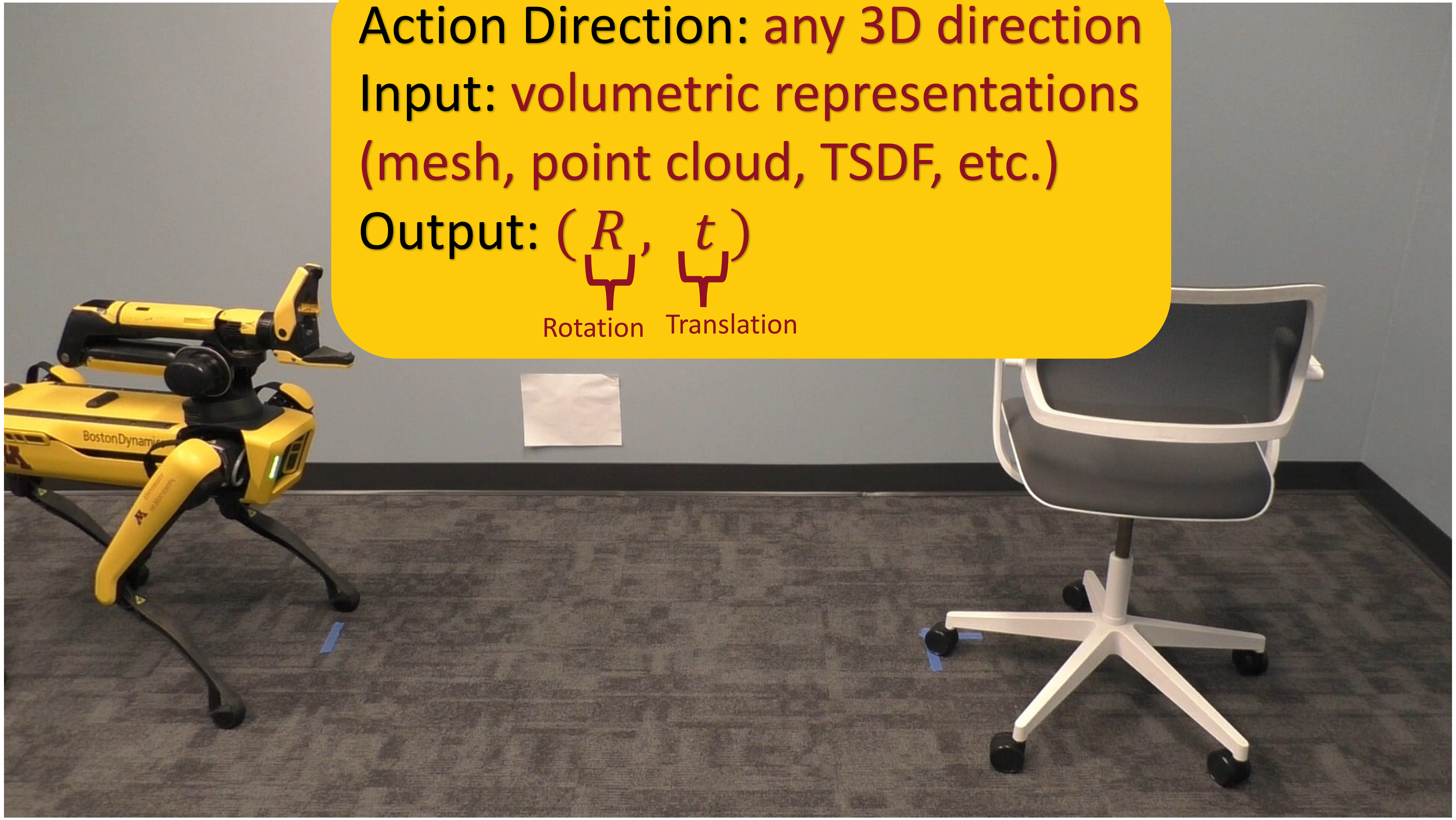
Grasping in SE(3) pose

Grasping in SE(3) pose

Action Direction: any 3D direction

Input: volumetric representations (mesh, point cloud, TSDF, etc.)

Output: $(\underbrace{R}_{\text{Rotation}}, \underbrace{t}_{\text{Translation}})$

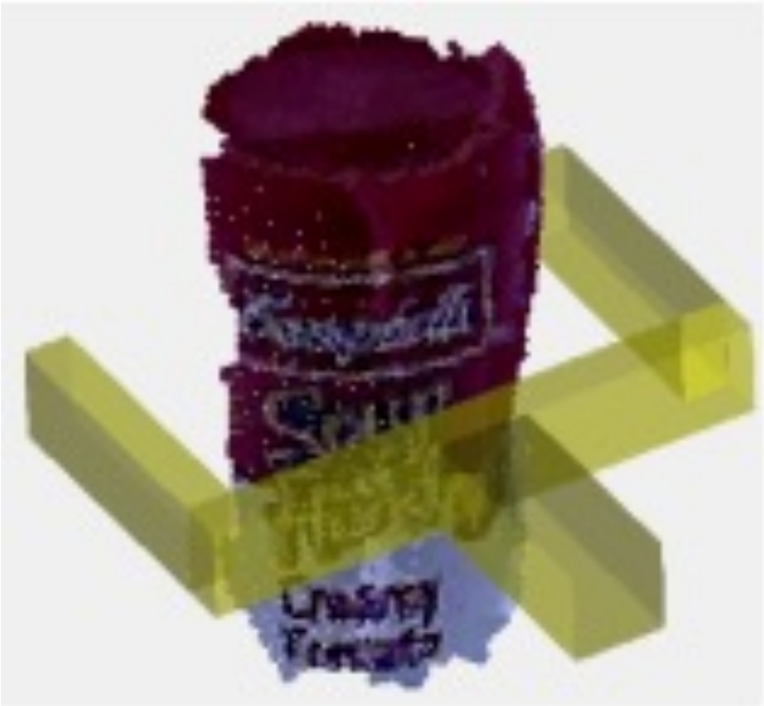


Grasping in SE(3) pose

- High precision grasp pose detection in dense clutter
- GraspNet-1Billion
- Contact-GraspNet
- GraspNeRF

High precision grasp pose detection in dense clutter

- Step 1: Sample grasp candidates (using Darboux Frame)



Example of one grasp candidate obtained from **Darboux Frame**

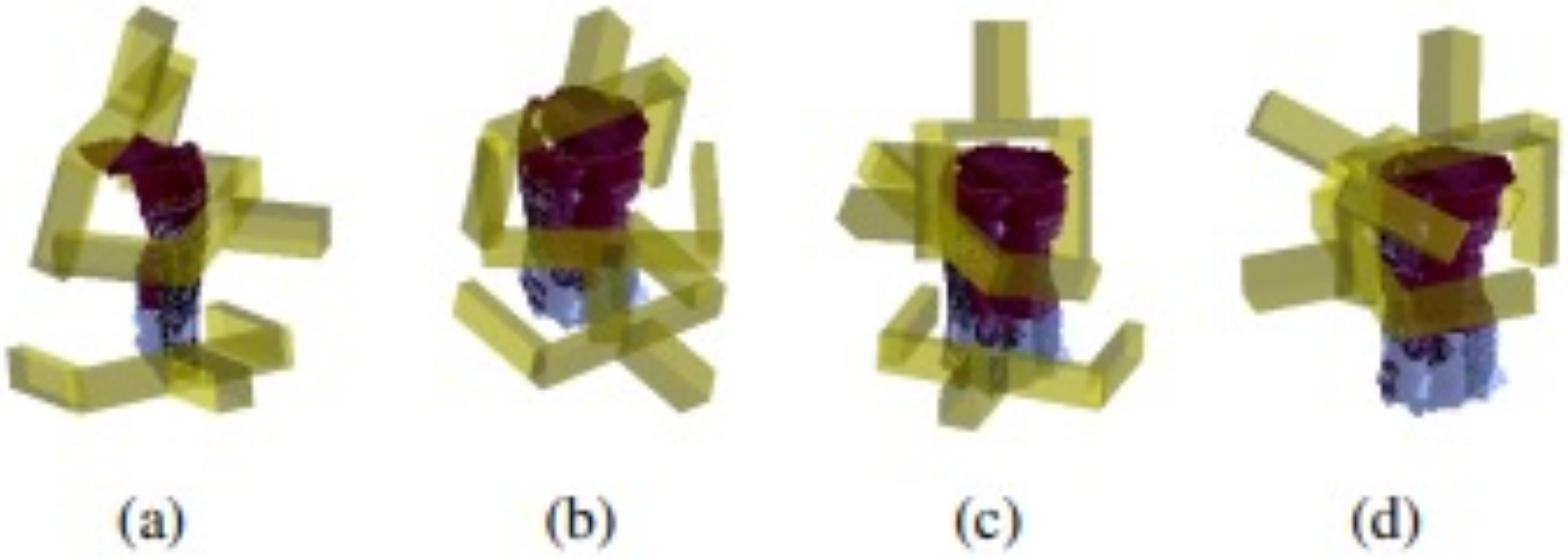


Fig. 1: Illustrations of grasp candidates found using our algorithm. Each image shows three examples of a gripper placed at randomly sampled grasp candidate configurations.

Gualtieri M, Ten Pas A, Saenko K, Platt R. 2016. High precision grasp pose detection in dense clutter. In 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 598–605. IEEE

High precision grasp pose detection in dense clutter

- Step 2: Classify the grasp candidates

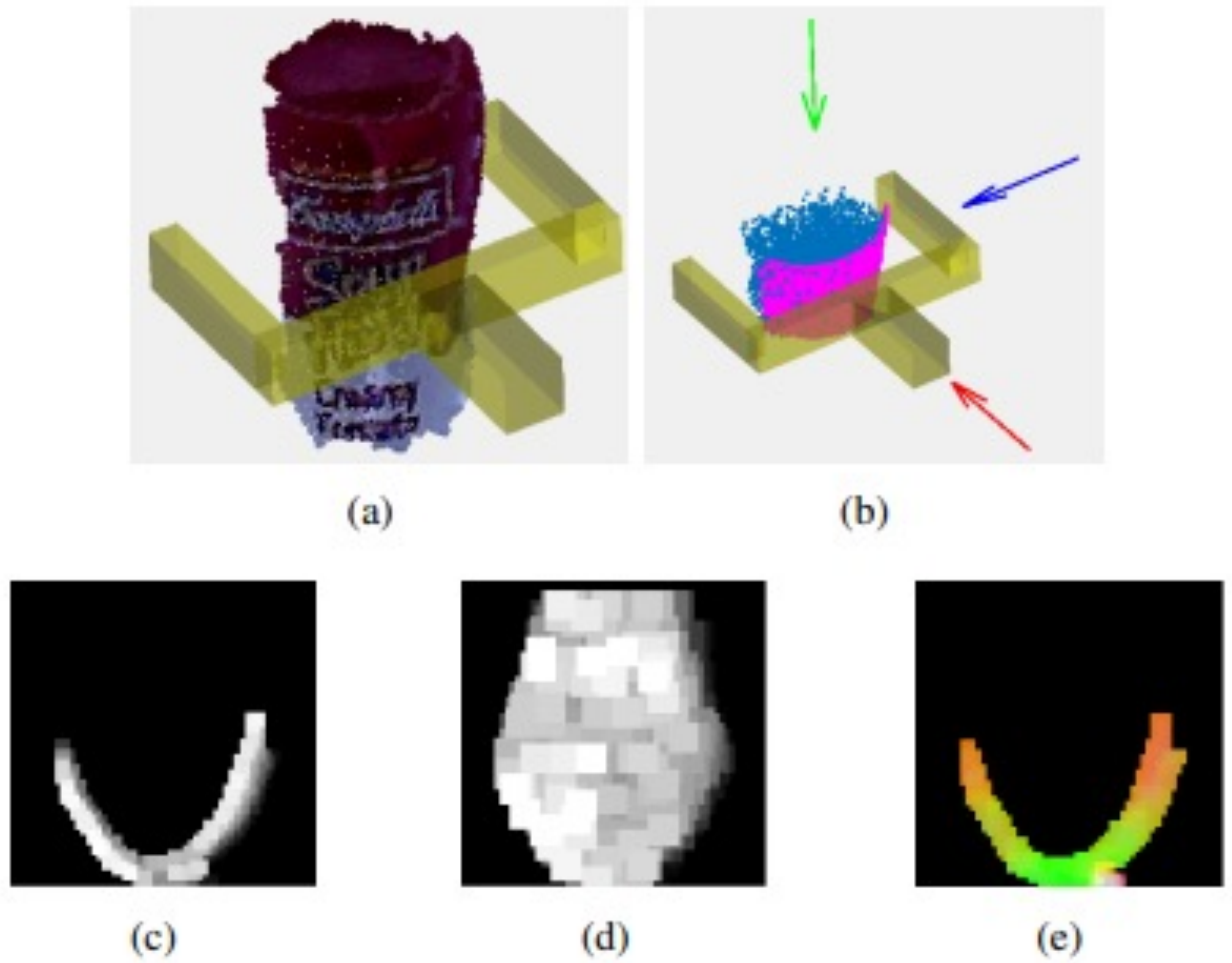


Fig. 2: Grasp representation. (a) A grasp candidate generated from partial point cloud data. (b) Local voxel grid frame. (c-e) Examples of grasp images used as input to the classifier.

LeNet

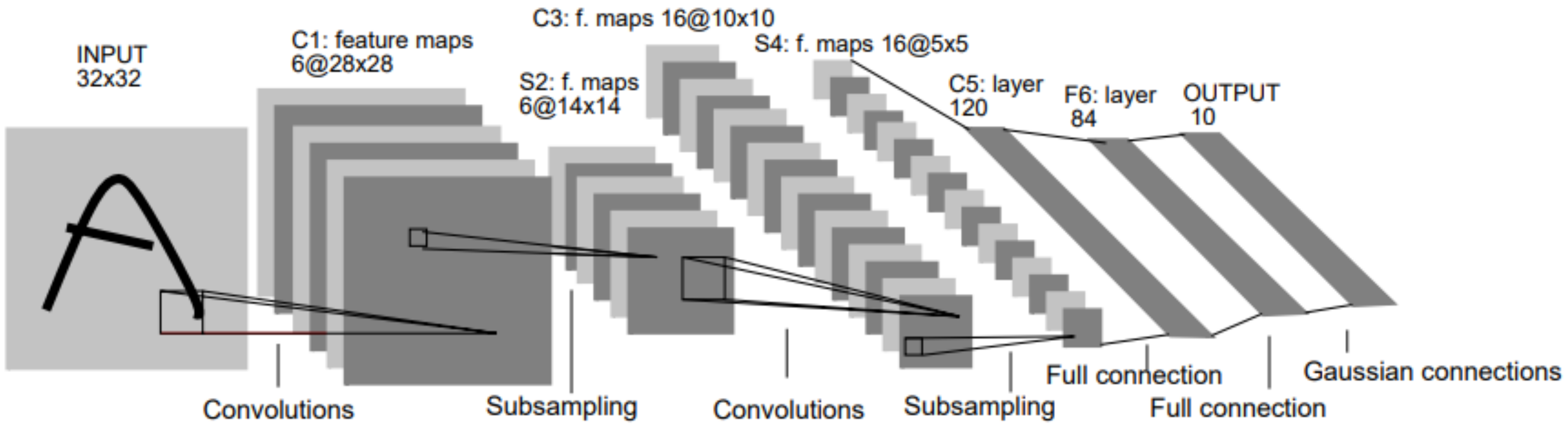
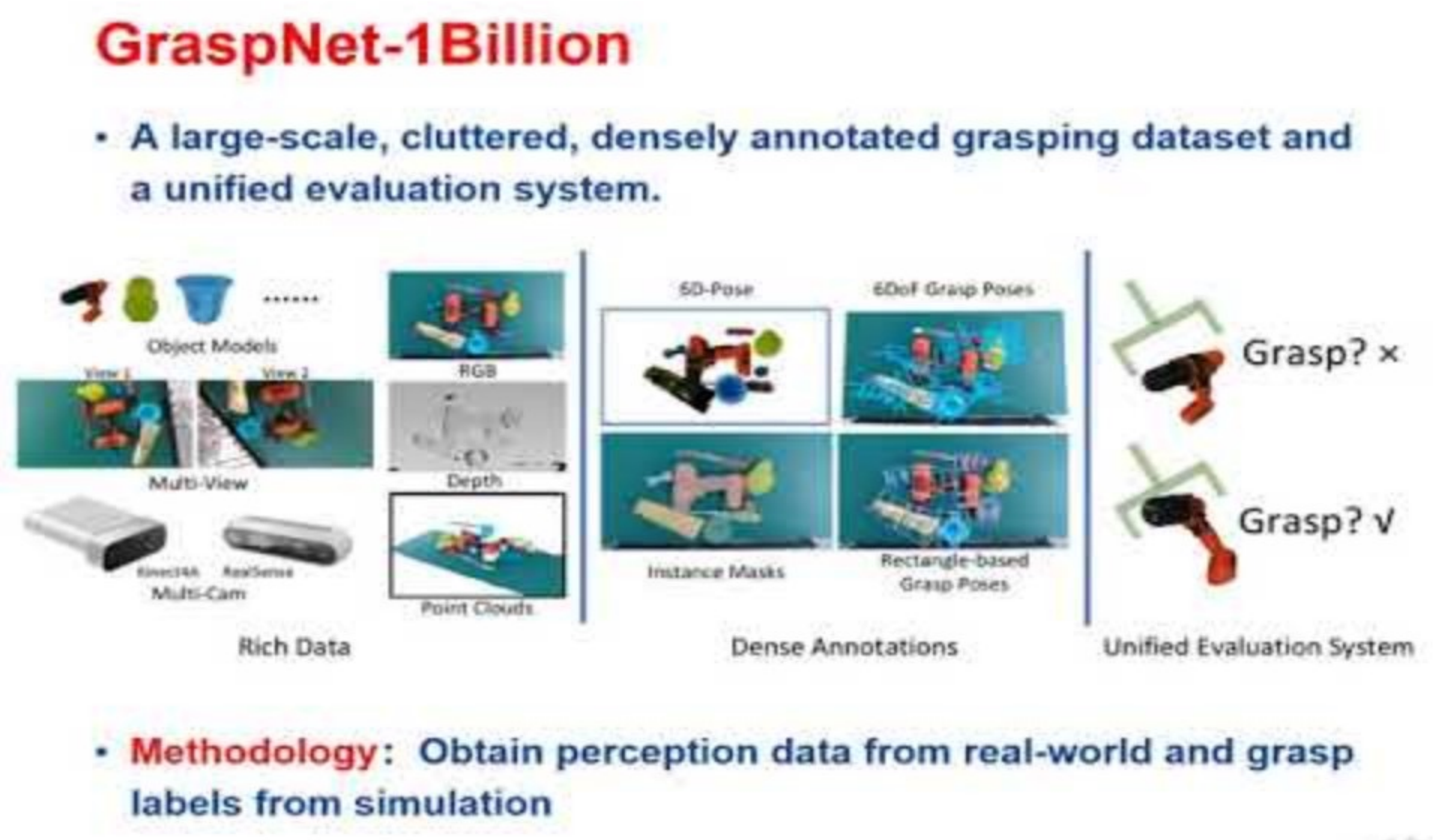


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

https://yann.lecun.com/exdb/publis/pdf/lecun-98.pdf

Gualtieri M, Ten Pas A, Saenko K, Platt R. 2016. High precision grasp pose detection in dense clutter. In 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 598–605. IEEE

GraspNet-1Billion



Fang HS, Wang C, Gou M, Lu C. 2020. GraspNet-1Billion: A Large-Scale Benchmark for General Object Grasping. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11444–11453

GraspNet-1Billion

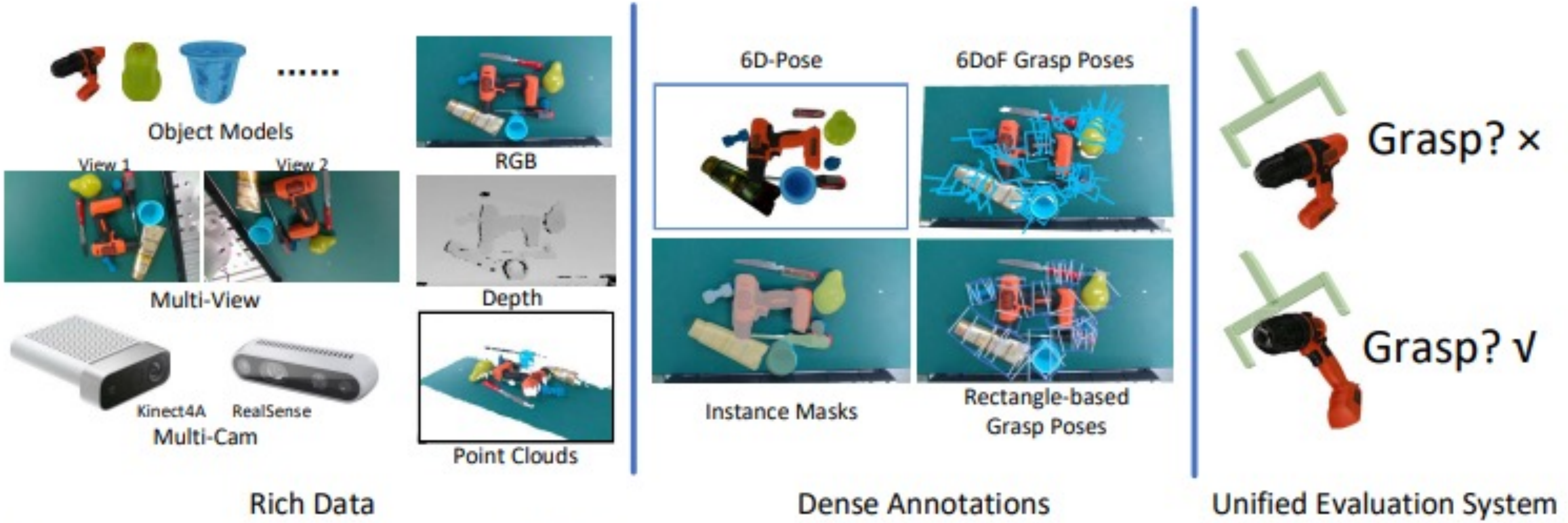


Figure 2. The key components of our dataset. RGB-D images are taken using both RealSense camera and Kinect camera from different views. The 6D pose of each object, the grasp poses, the rectangle grasp poses and the instance masks are annotated. A unified evaluation system is also provided.

Fang HS, Wang C, Gou M, Lu C. 2020. GraspNet-1Billion: A Large-Scale Benchmark for General Object Grasping. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11444–11453

GraspNet-1Billion

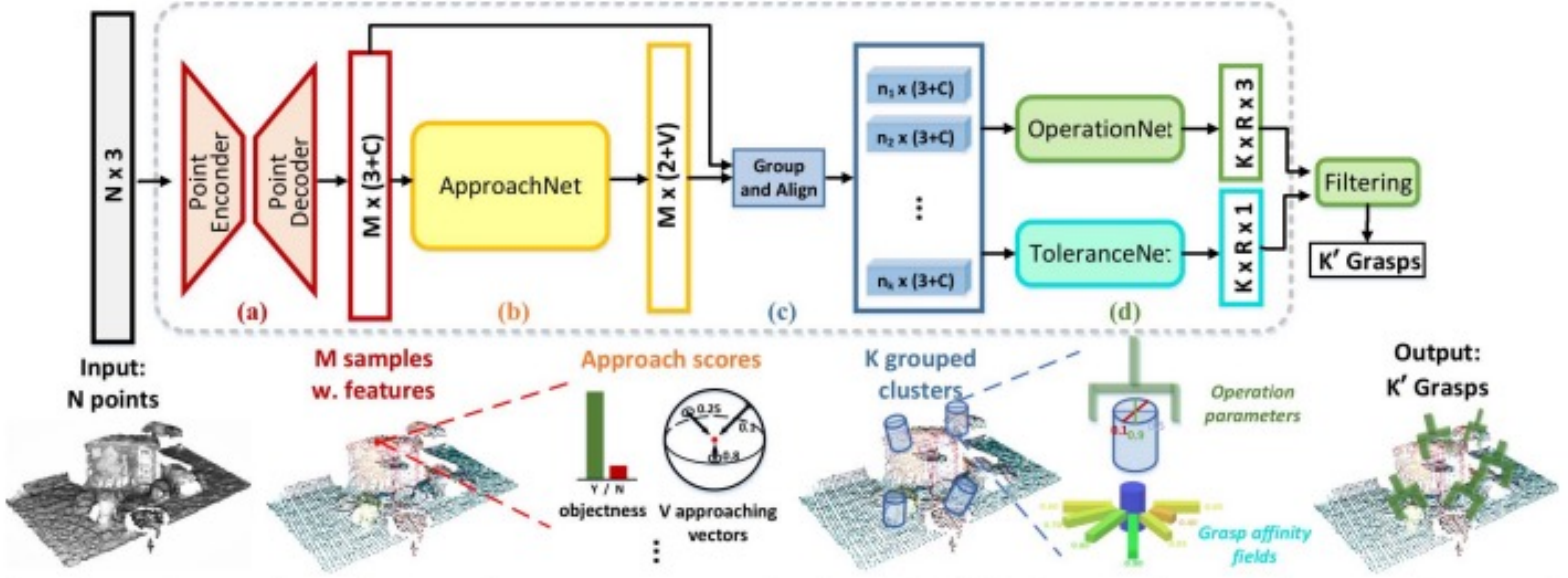
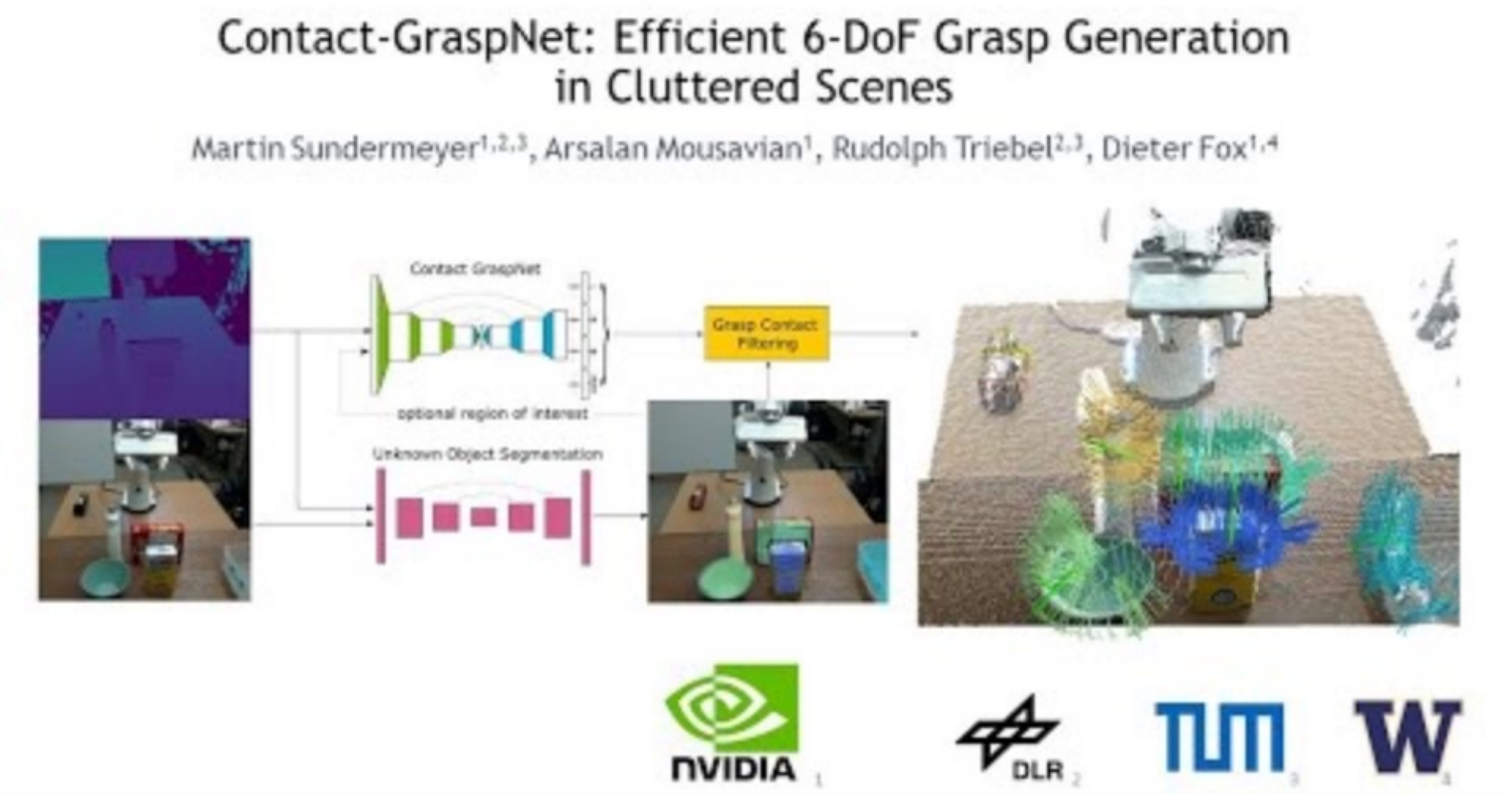


Figure 5. Overview of our end-to-end network. (a) For a scene point cloud with N point coordinates as input, a point encoder-decoder extracts cloud features and samples M points with C-dim features. (b) Approaching vectors are predicted by ApproachNet and are used to (c) grouped points in cylinder regions. (d) OperationNet predicts the operation parameters and ToleranceNet predicts the grasp robustness. See text for more details.

Fang HS, Wang C, Gou M, Lu C. 2020. GraspNet-1Billion: A Large-Scale Benchmark for General Object Grasping. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11444–11453

Contact-GraspNet



Sundermeyer M, Mousavian A, Triebel R, Fox D. 2021. Contact-graspnet: Efficient 6-dof grasp generation in cluttered scenes. In 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 13438–13444. IEEE

Contact-GraspNet

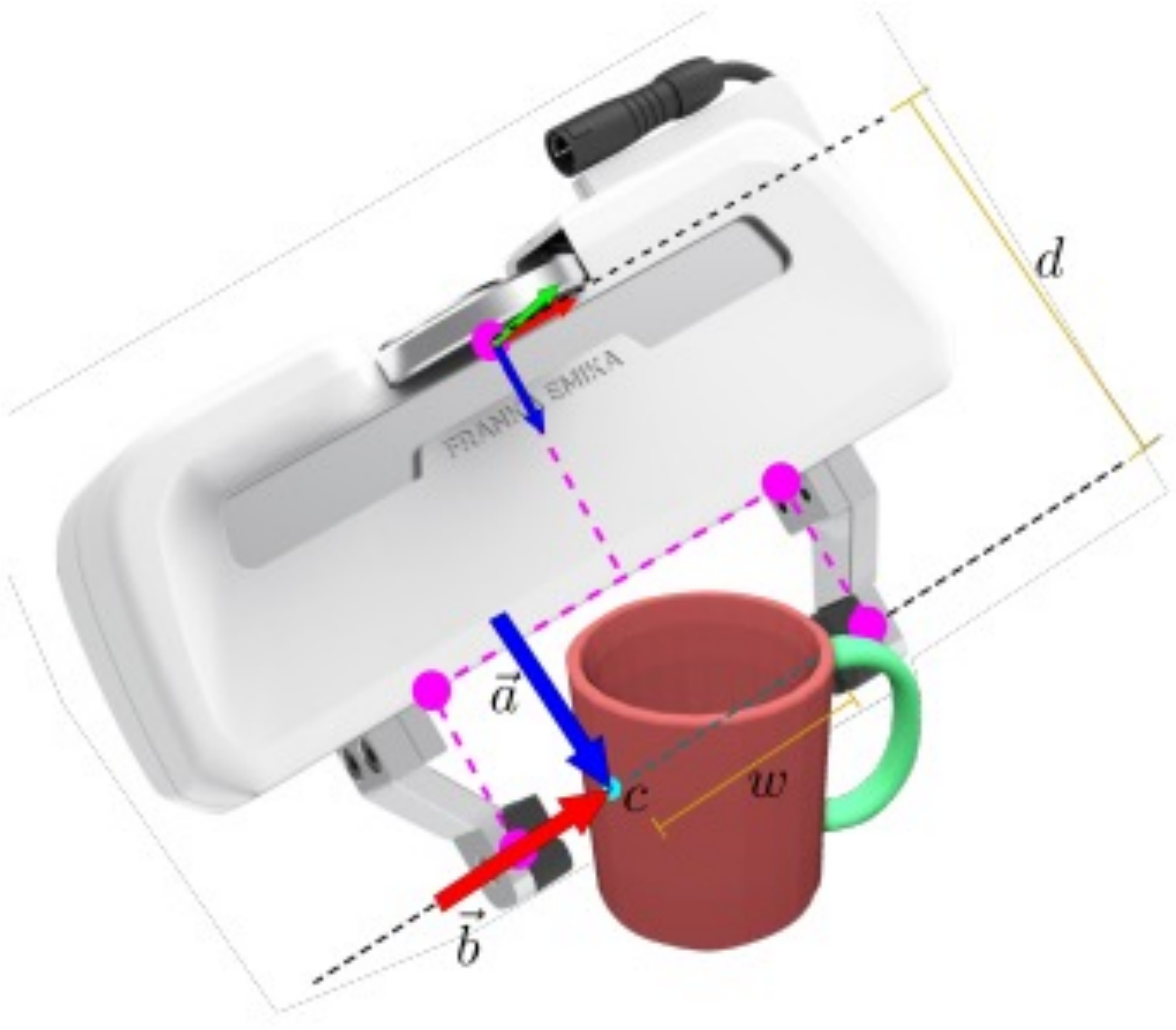
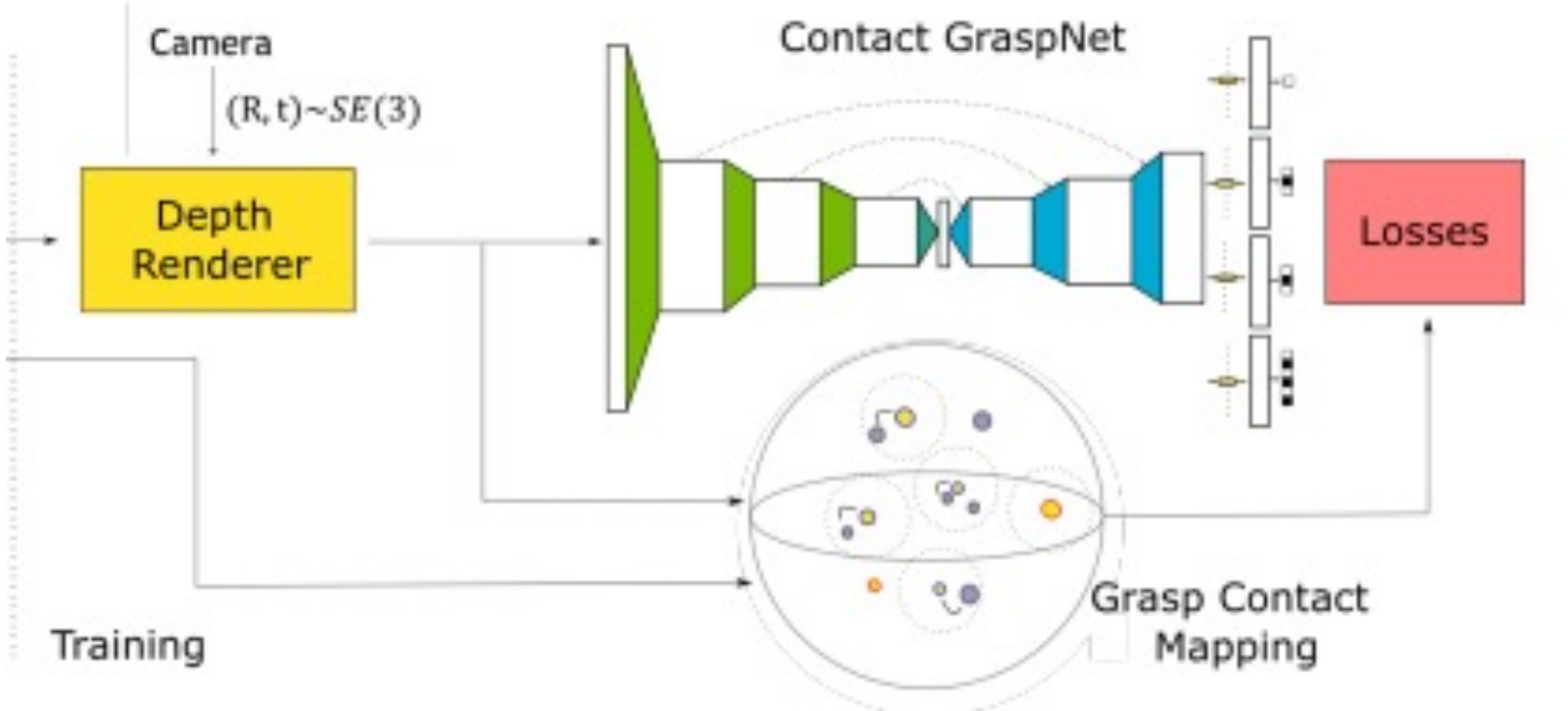


Fig. 3. Our grasp representation: c depicts an observed contact point. $\mathbf{a}$ and $\mathbf{b}$ constitute the 3-DoF rotation, w is the predicted grasp width, d the distance from baseline to base frame. In pink we show the five gripper points $\mathbf{v}$ that we used in the l_{add-s} loss.



Sundermeyer M, Mousavian A, Triebel R, Fox D. 2021. Contact-graspnet: Efficient 6-dof grasp generation in cluttered scenes. In 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 13438–13444. IEEE

Contact-GraspNet

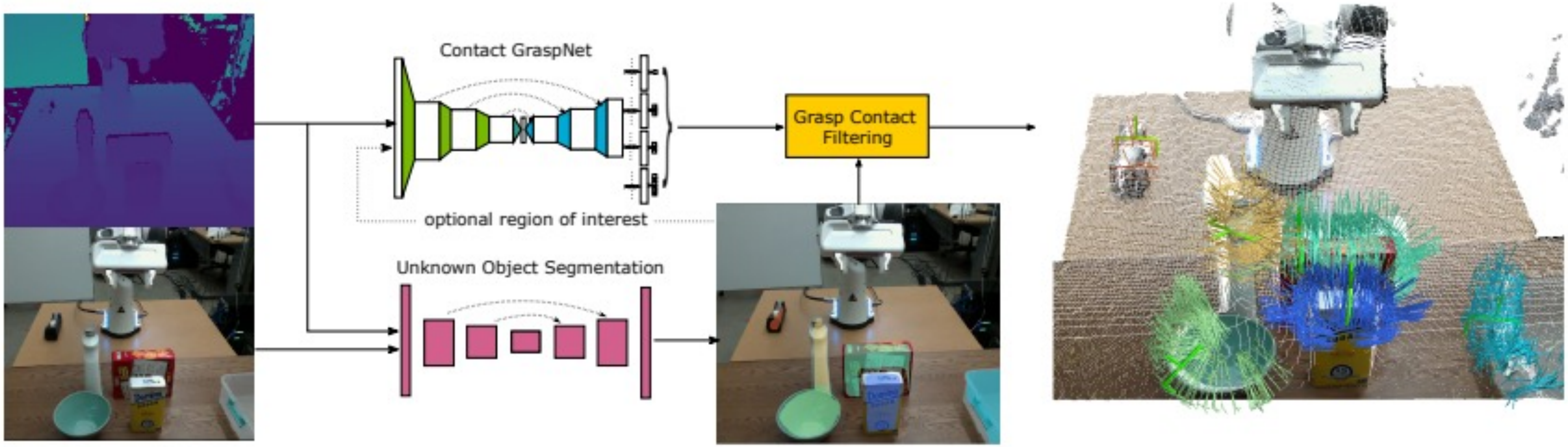
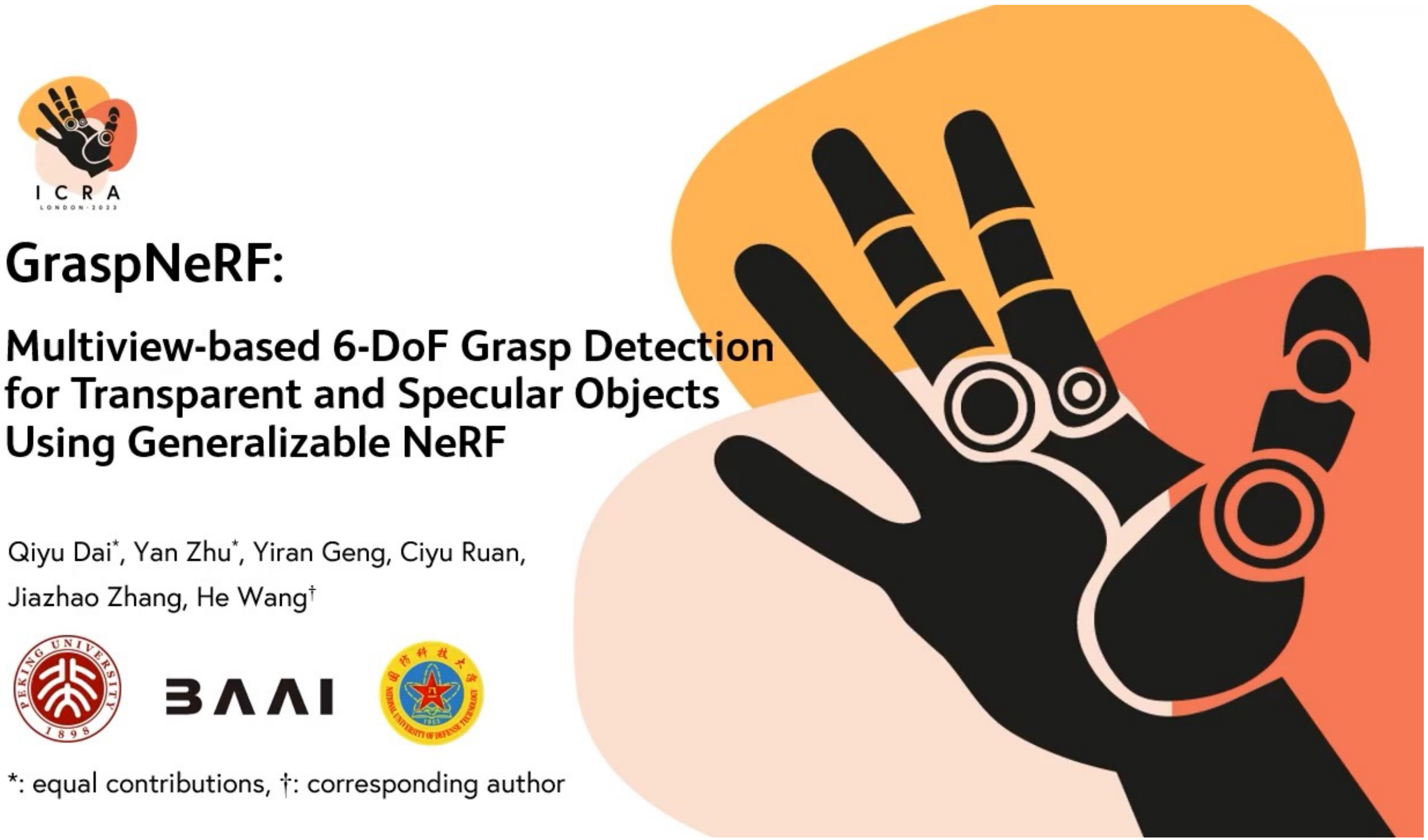


Fig. 4. Full Inference Pipeline: We segment unknown objects from an RGB-D image using [15]. Our Contact-GraspNet processes the full scene point cloud or a local region of interest around a target object. Predicted 6-DoF grasps are then associated to object segments by filtering their contact points. On the right we show the predicted 6-DoF grasp distribution and, in bold, the most confident grasp per segment.

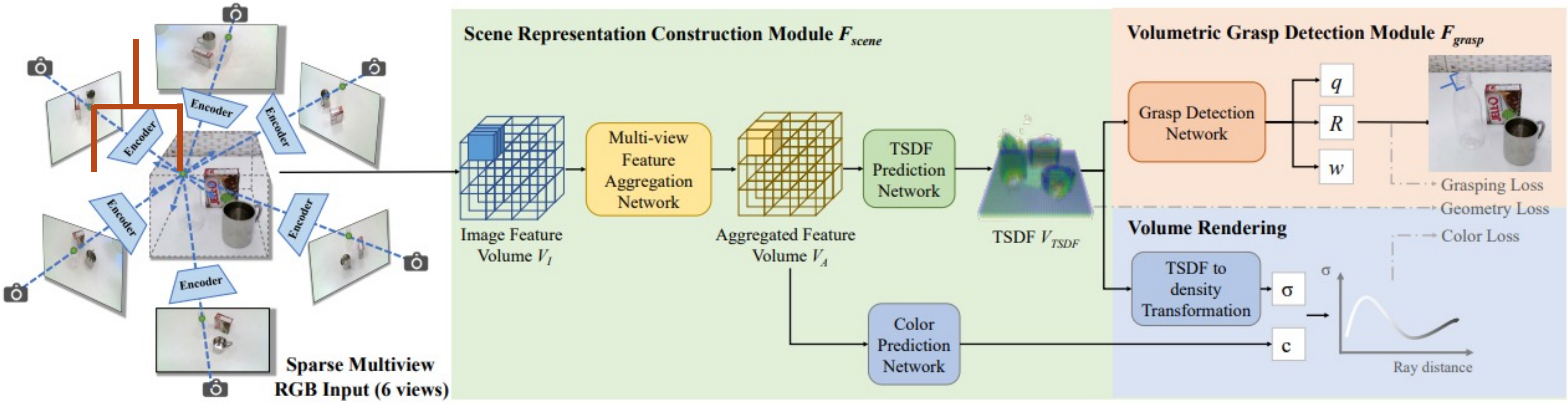
Sundermeyer M, Mousavian A, Triebel R, Fox D. 2021. Contact-graspnet: Efficient 6-dof grasp generation in cluttered scenes. In 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 13438–13444. IEEE

GraspNeRF



Q. Dai, Y. Zhu, Y. Geng, C. Ruan, J. Zhang, and H. Wang, "GraspNeRF: Multiview-based 6-DoF Grasp Detection for Transparent and Specular Objects Using Generalizable NeRF," IEEE International Conference on Robotics and Automation (ICRA), 2023.

GraspNeRF



Collect images at different poses

Construct the 3D scene and convert it into TSDF (Truncated Signed Distance Function)

Predict grasp poses from TSDF

Q. Dai, Y. Zhu, Y. Geng, C. Ruan, J. Zhang, and H. Wang, “GraspNeRF: Multiview-based 6-DoF Grasp Detection for Transparent and Specular Objects Using Generalizable NeRF,” IEEE International Conference on Robotics and Automation (ICRA), 2023.

Let’s look at other object categories in the context of robotic grasping

Transparent Objects

- Evo-NeRF:

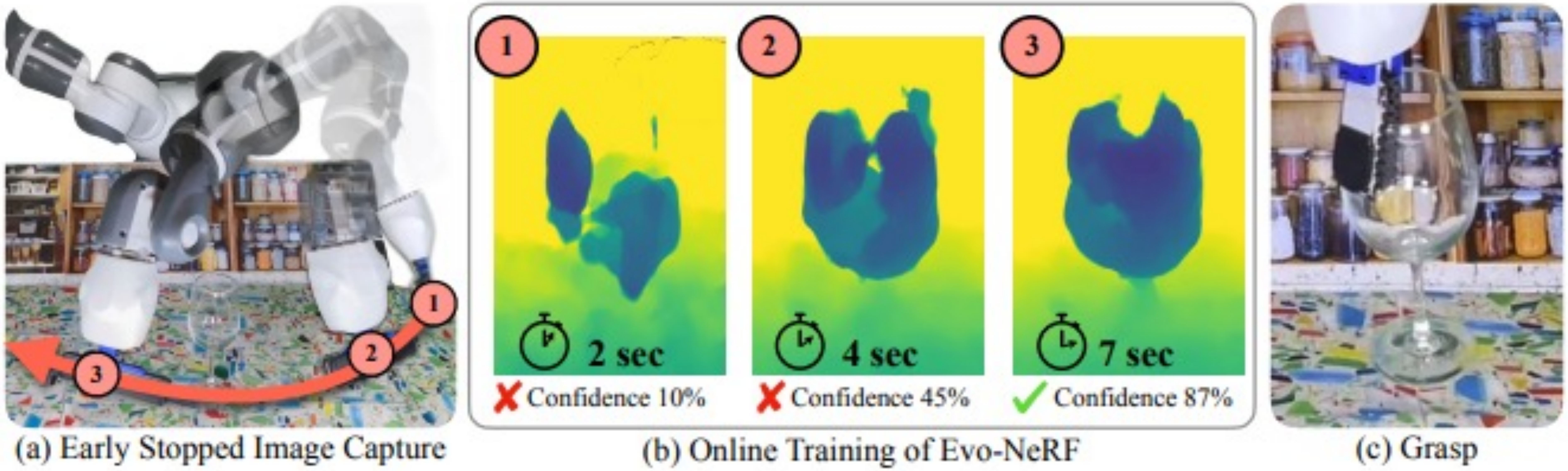


Figure 2: **Evo-NeRF for rapid grasping: (a)** The robot begins capturing images along a hemisphere trajectory (red arrow) **(b)** Evo-NeRF trains a NeRF during arm motion, building graspable geometry of the wineglass. Grasp confidence from RAG-Net builds as NeRF learns geometry, reaching the stopping threshold at (3). **(c)** The robot executes the grasp predicted by RAG-Net at the early stop point.

J. Kerr *et al.*, 'Evo-nerf: Evolving nerf for sequential robot grasping of transparent objects', in *6th Annual Conference on Robot Learning*, 2022.

Deformable Objects

- Active garment recognition and target grasping point detection using deep learning

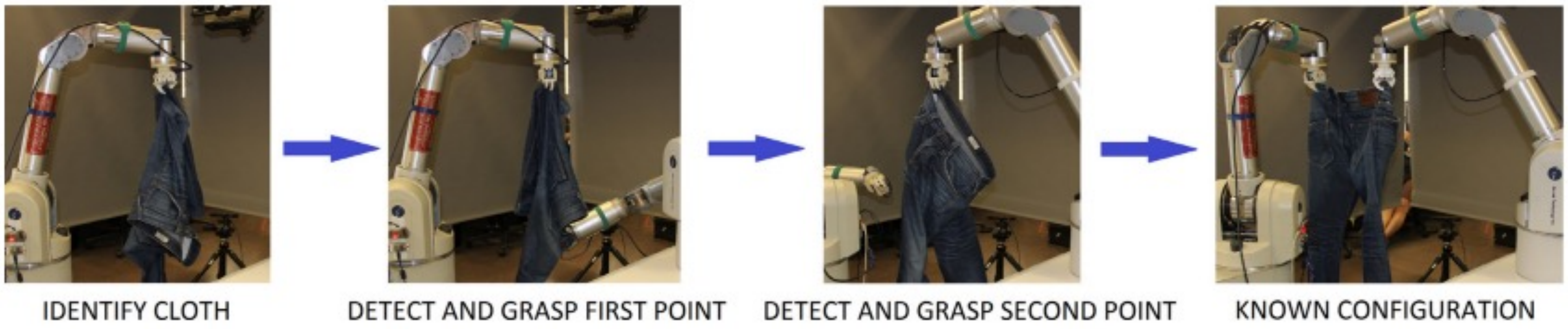


Fig. 1. Process to move from holding an unidentified grasped garment to holding it from the predefined points. At each detection step the robot rotates the garment until the grasping points are visible.

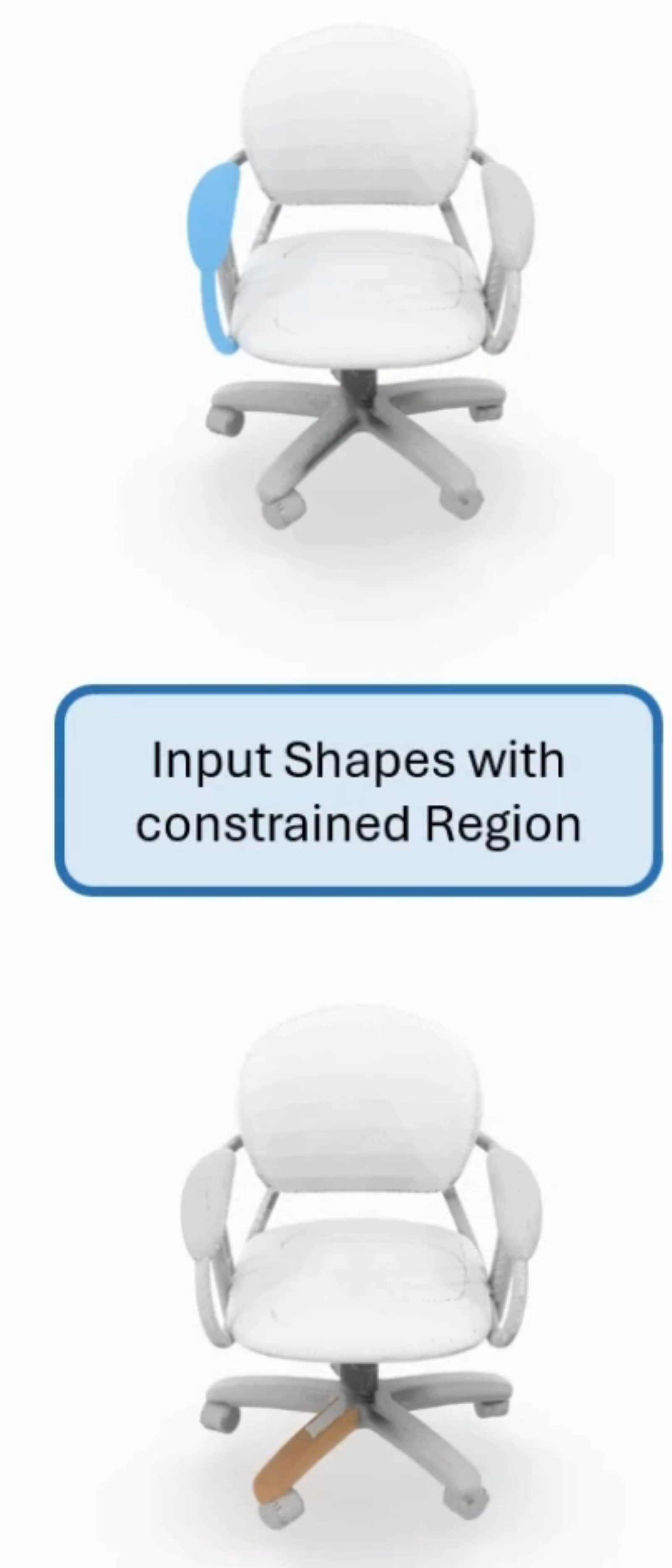
Classify the type of the garment from an image using CNN

Predict the visibility and Cartesian location of the possible grasp points from depth images using CNN

E. Corona, G. Alenyà, A. Gabas, and C. Torras, "Active garment recognition and target grasping point detection using deep learning," Pattern Recognition, vol. 74, pp. 629–641, 2018, doi: https://doi.org/10.1016/j.patcog.2017.09.042.

Large Objects

- CGDF:



G. Singh *et al.*, 'Constrained 6-DoF Grasp Generation on Complex Shapes for Improved Dual-Arm Manipulation', *arXiv [cs.RO]*. 2024.

Next Lecture:
Imitation Learning

Project 3 released

- Instructions available on the website
  - Here: https://rpm-lab.github.io/CSCI5980-F24-DeepRob/projects/project3/
  - Uses PROPS Detection dataset
- Implement CNN for classification and Faster R-CNN for detection
- Autograder will be available soon!
- **Due Monday, October 28th 11:59 PM CT**

DR

Final Project Proposal—Due Today

DeepRob: What is a Project Proposal?

- Create a 3 page proposal on a google-doc/overleaf 🔥 Please keep the proposal excluding the references to not more than 3 pages.
- Should contain a title that describes the project (keep it simple)
- Should contain full name(s), email addresses of the team.
- Should contain the following sections.
  - **Objective** - What capability does this project aim to give a robot? For example you should be able to say - *"This project aims to impart the capability of .... to the robot. Given a observation in the form of ...., the robot will be able to do ....."*
  - **Input-Output during Inference time** - What are the input and output variables of the system you are building? For example you should be able to say - *"The robot/model takes in RGBD observation I of size $H \times W \times 3$, gripper pose $G \in SE(3)$ and produces action $A \in SE(3)$ ...."*
  - **Method** - What is the algorithm, pipeline, or neural network architecture you are proposing to develop the capability? If it has an algorithm, please describe it. If it is a neural network architecture, describe it. If it is a learning method, what is the training objective, what are the loss functions you will experiment on.
    - **Illustrative figure** - can help quickly understand the method being proposed and the big idea.
  - **Data collection** - Assuming that all the projects in this course is data-driven, where does the data for your project come from (existing datasets, or simulation env) ? Are you going to collect new data?
  - **Evaluation** - How will you evaluate if your method worked? What will you compare with? What is the measure of success?
  - **Resources** - What will be the resources you will use for this project? Is this your desktop or laptop? MSI? Are you using a real-robot setup? If yes, describe the setup. Are you using simulation environment? If yes, describe the setup.
  - **Timeline** - Please plan a weekly schedule and things to accomplish on a weekly basis to successfully finish the project. Do you due diligence to consider other commitments in your semester while creating this timeline for everyone in the group. Discuss this timeline in detail with other members of the group to ensure success. You can tabulate this.
    - Week 10/21-10/25 - Task [member1] - Task[member2] ....
    - Week 10/28-11/01 - Task [member1] - Task[member2] ....
    - ...
  - **Deliverables** - What do you plan to deliver at the end of the project time? Real-world demo? Code for others to use? Make this a technical paper?
  - **Summary of 3 papers** - Please read 3 papers as a group and summarize them with relation to your project. How will you use the techniques from this paper in your work?
  - **References** - Please include any reference material (papers, code, datasets) that you found online that is relevant to your project. This includes all the images you use requires a source citation.
- Please check the grammar or spelling mistakes.
- An upload link will be made available for the submission.

DeepRob

Lecture 13
Robot Grasp Learning
University of Minnesota

Slides created by Xun Tu and Karthik Desingh, Fall 2024