OCTOBER 1 - 5, 2023

IEEE/RSJ International Conference on Intelligent Robots and Systems

武漢大學 工业科学研究院

The Institute of Technological Sciences, Wuhan University

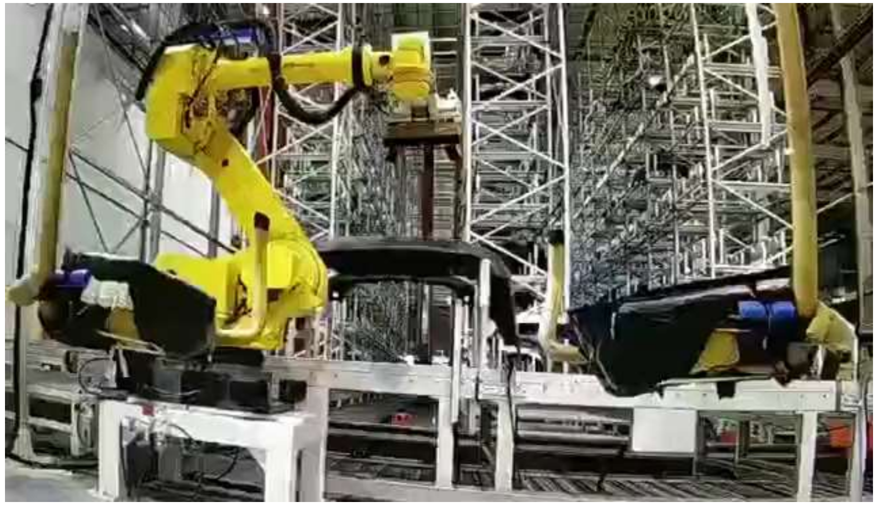
Robust Robotic Grasping
- From Hundreds to Millions

Miao Li

Wuhan University

2023.10.1

Grasping is very useful for many industry applications

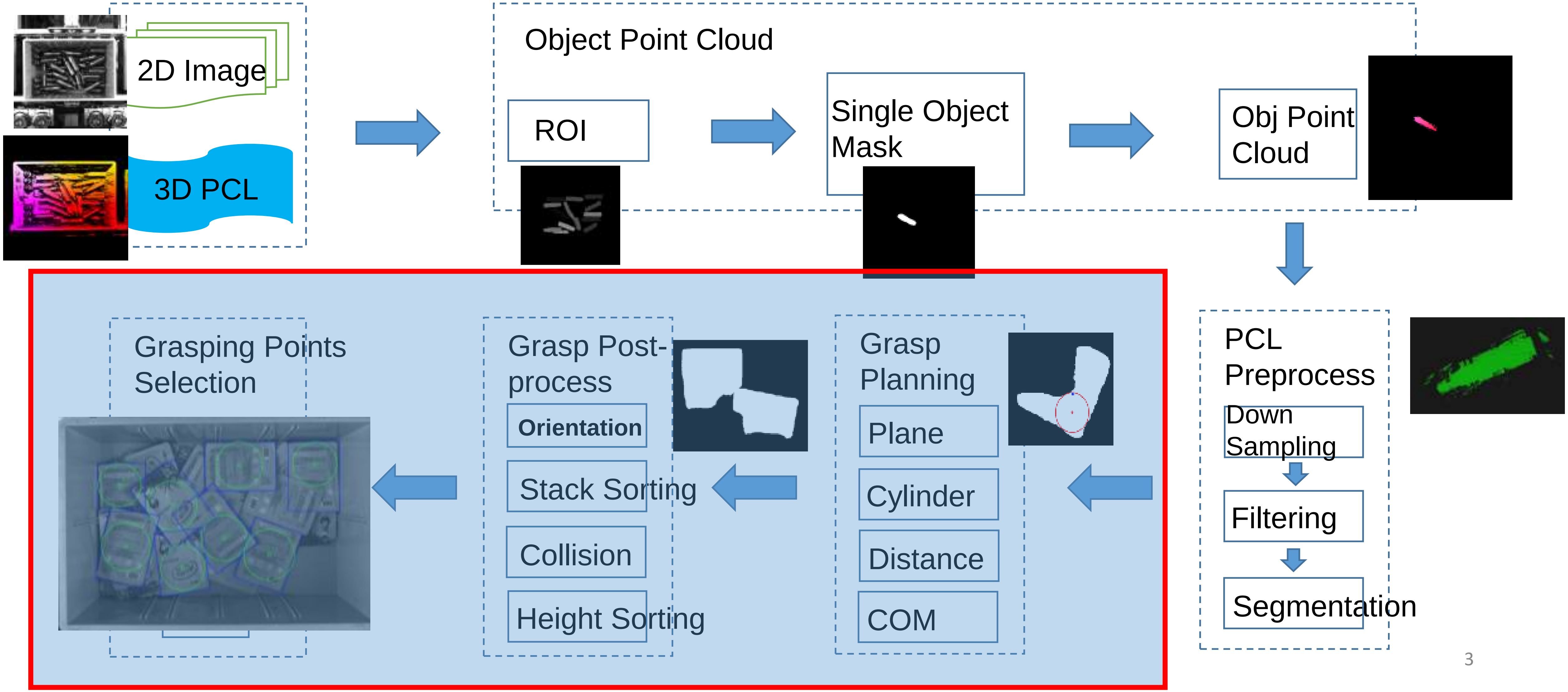


Automotive

Warehouse

Machining

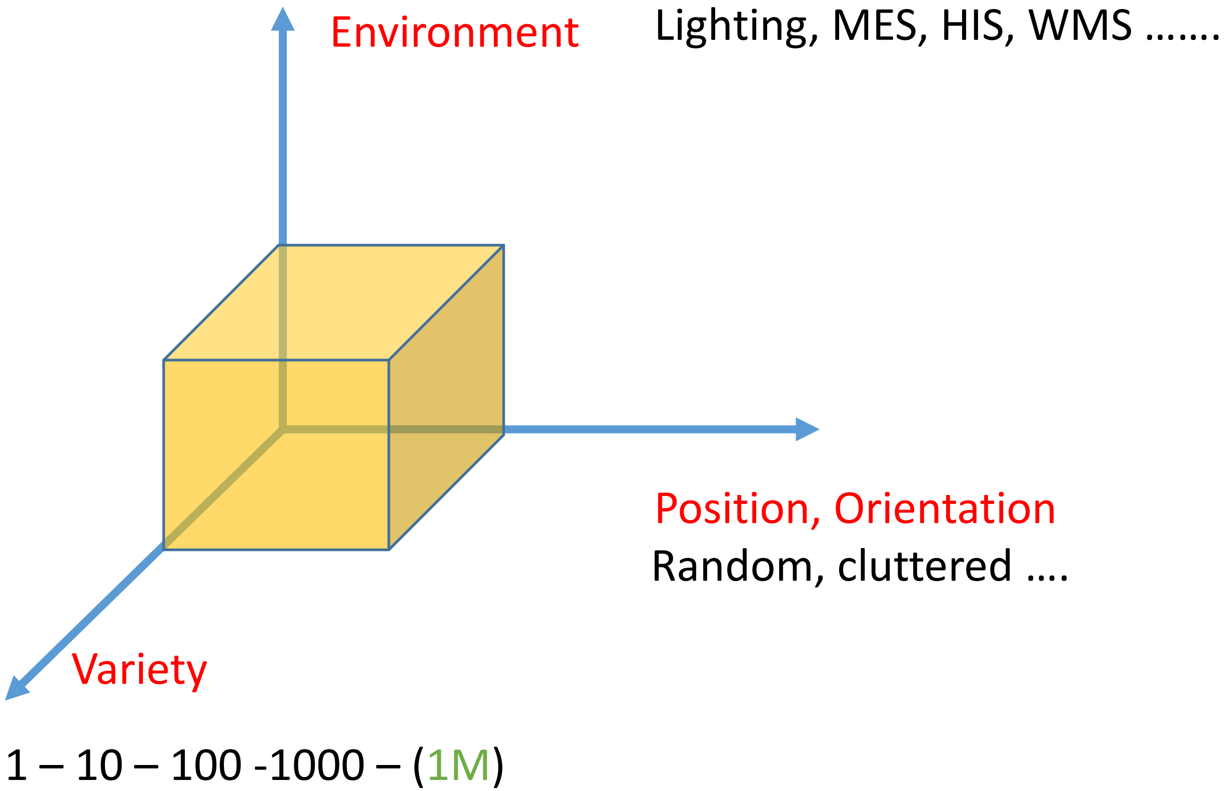
A general pipeline of robotic grasping



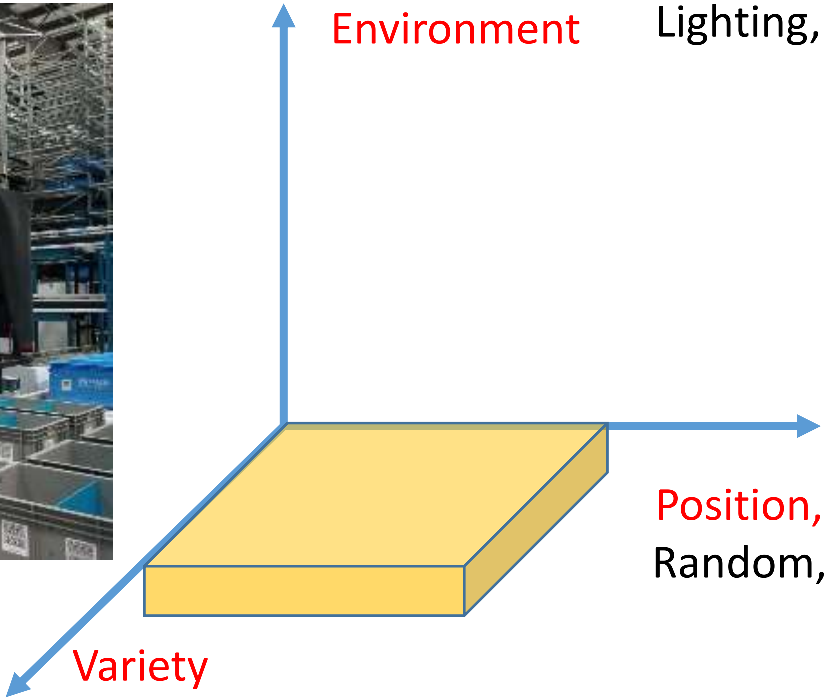
What is missing for grasping?

- **How about Millions of Objects?**

What is the new challenge?



What is the new challenge?

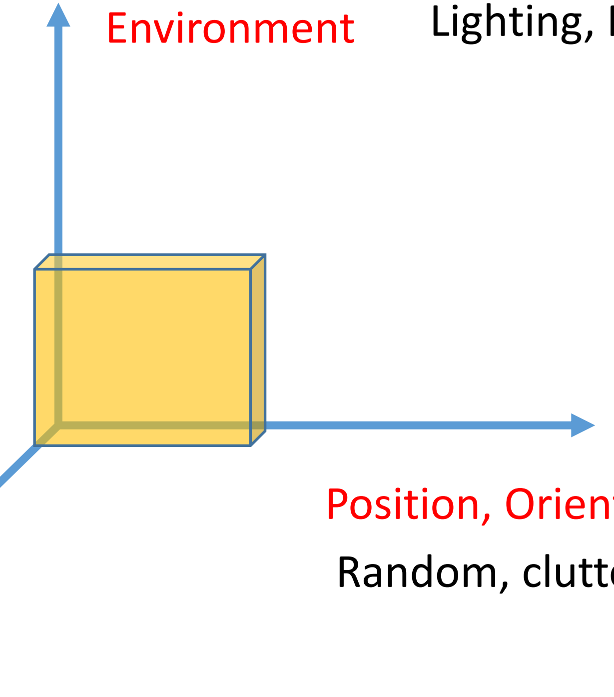


Lighting, MES, HIS, WMS …….

Random, cluttered …. (object can be mixed)

1 – 10 – 100 -1000 – (1M)

What is the new challenge?



Lighting, MES, HIS, WMS …….

Random, cluttered ….

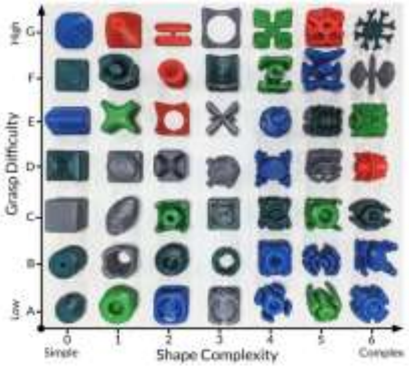
1 – 10 – 100 -1000 – (1M)

Pipeline for the new challenge

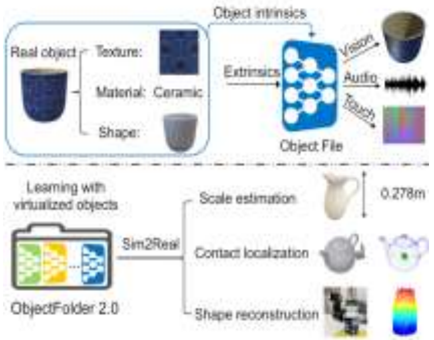
Dataset

Grasping Datasets

Cornell



EGAD

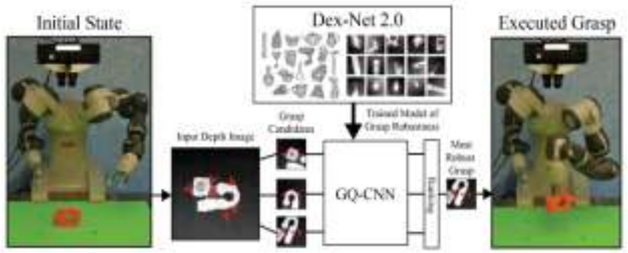


ObjectFolder

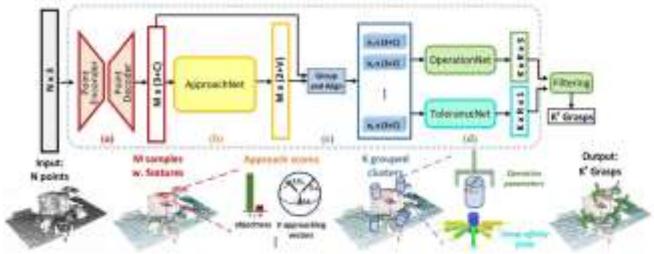
YCB

Data-Driven Grasp Planning

Neural Network

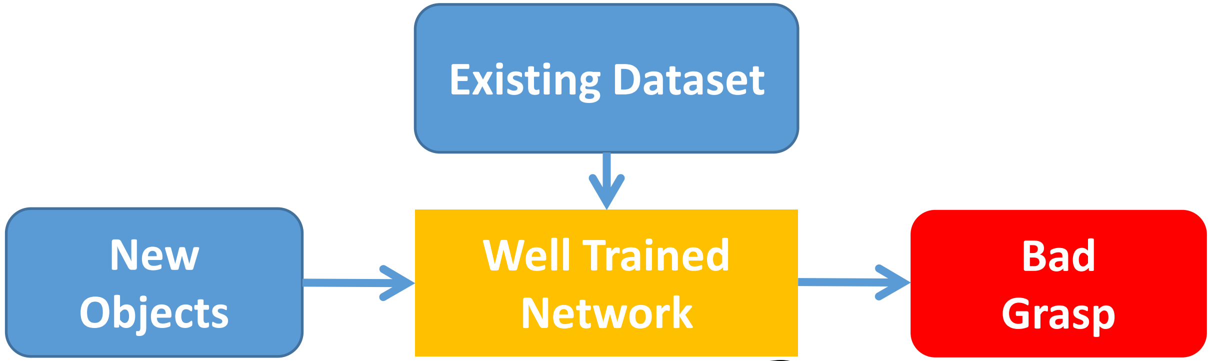


DexNet



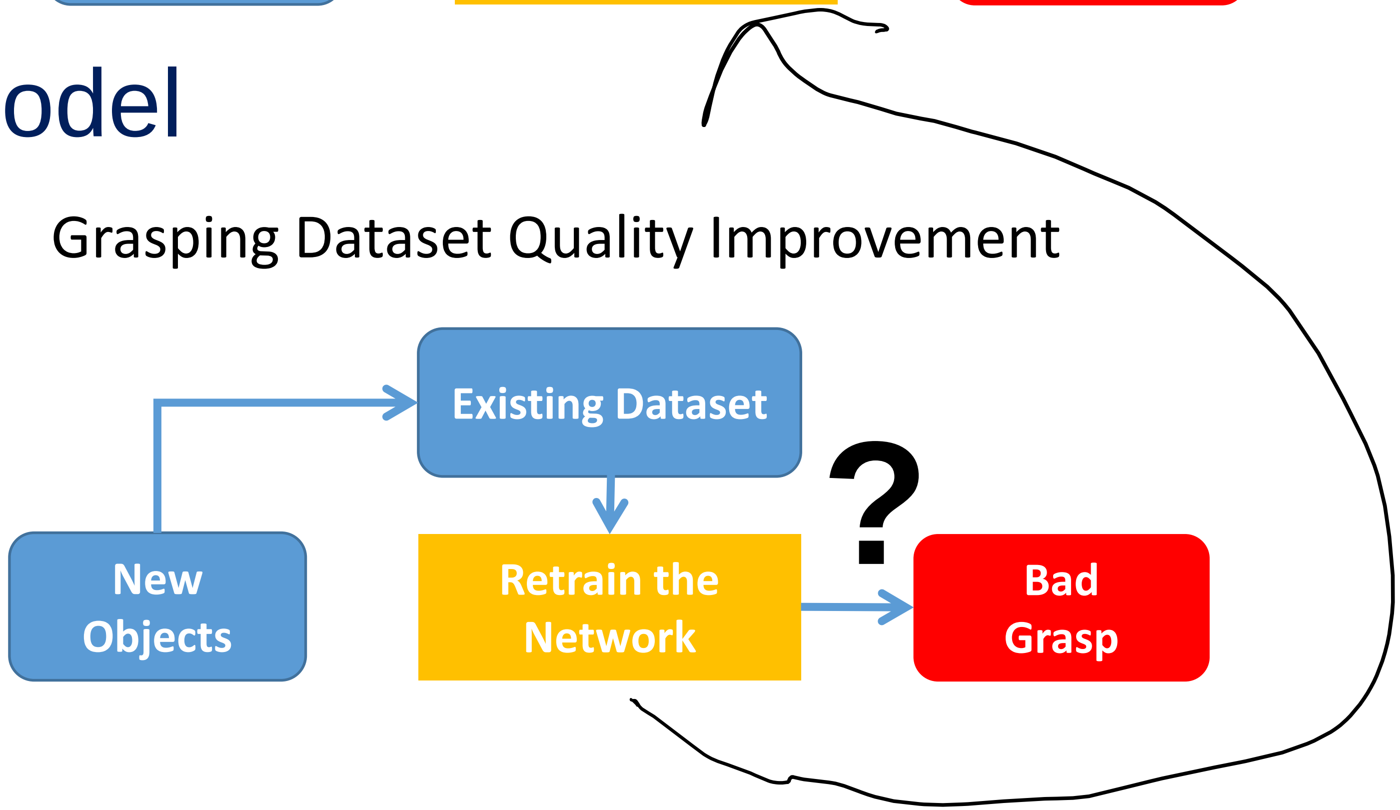
GraspNet-1Billion

What will happen if a grasp fails?



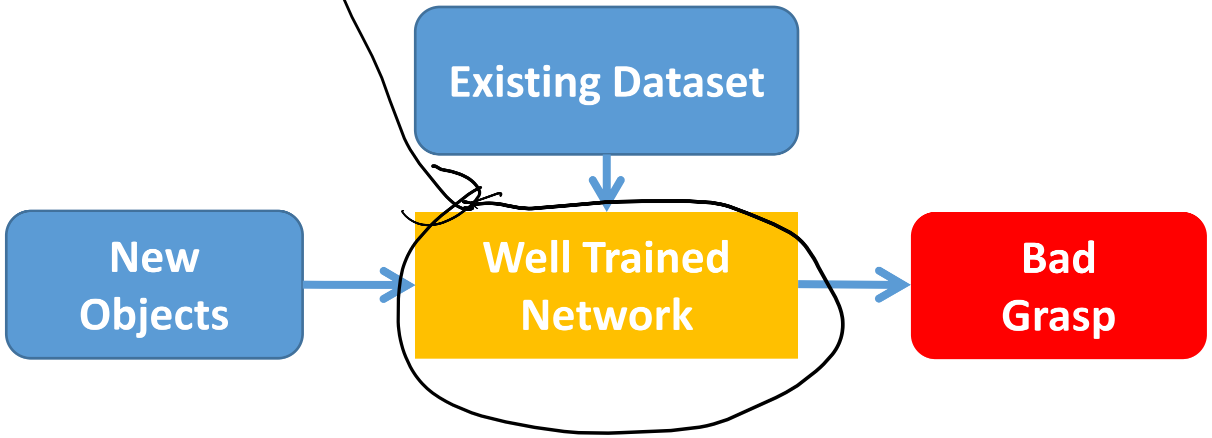
Retrain the model

Grasping Dataset Quality Improvement



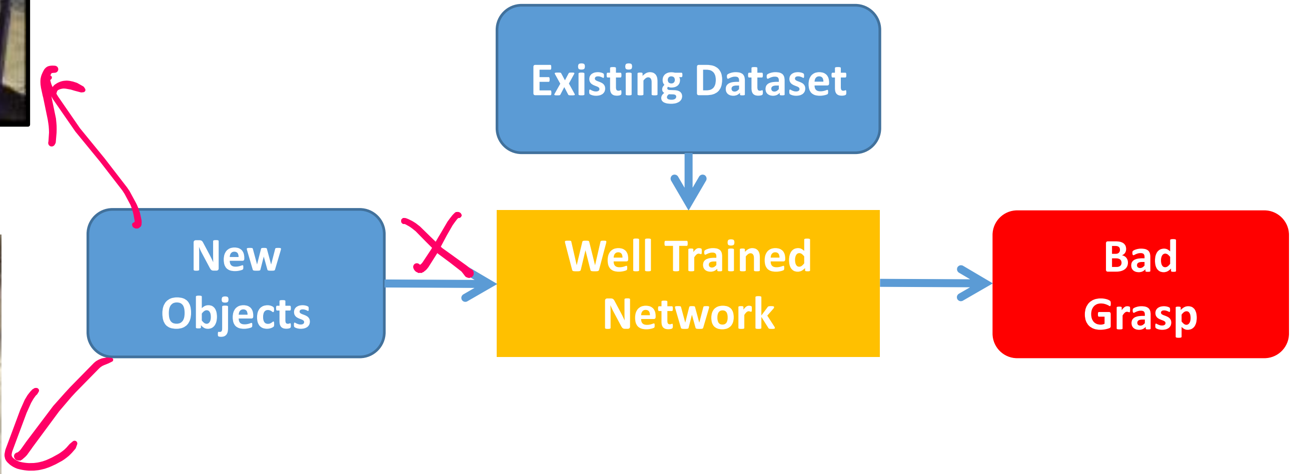
What will happen if a grasp fails?

Design new algorithm



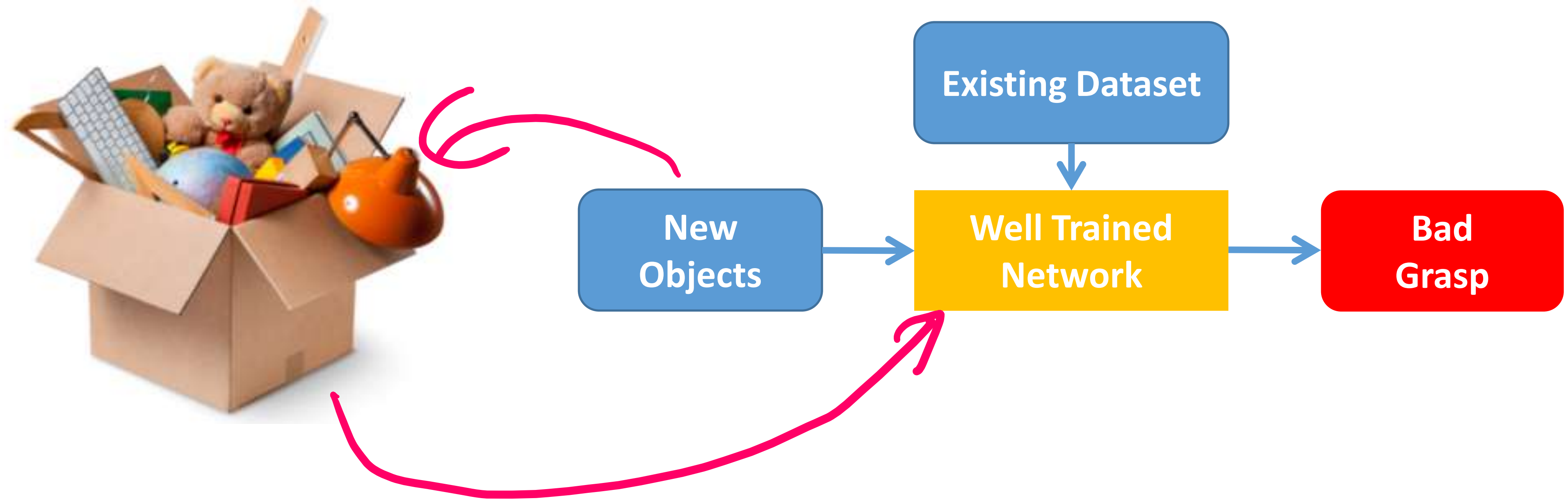
What will happen if a grasp fails?

Human in the loop

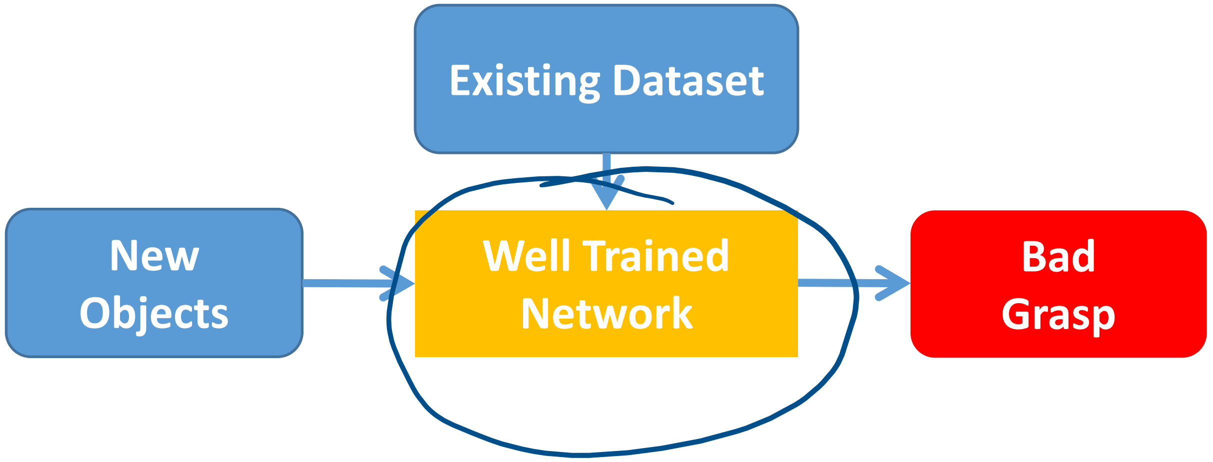


What will happen if a grasp fails?

Re-design the objects

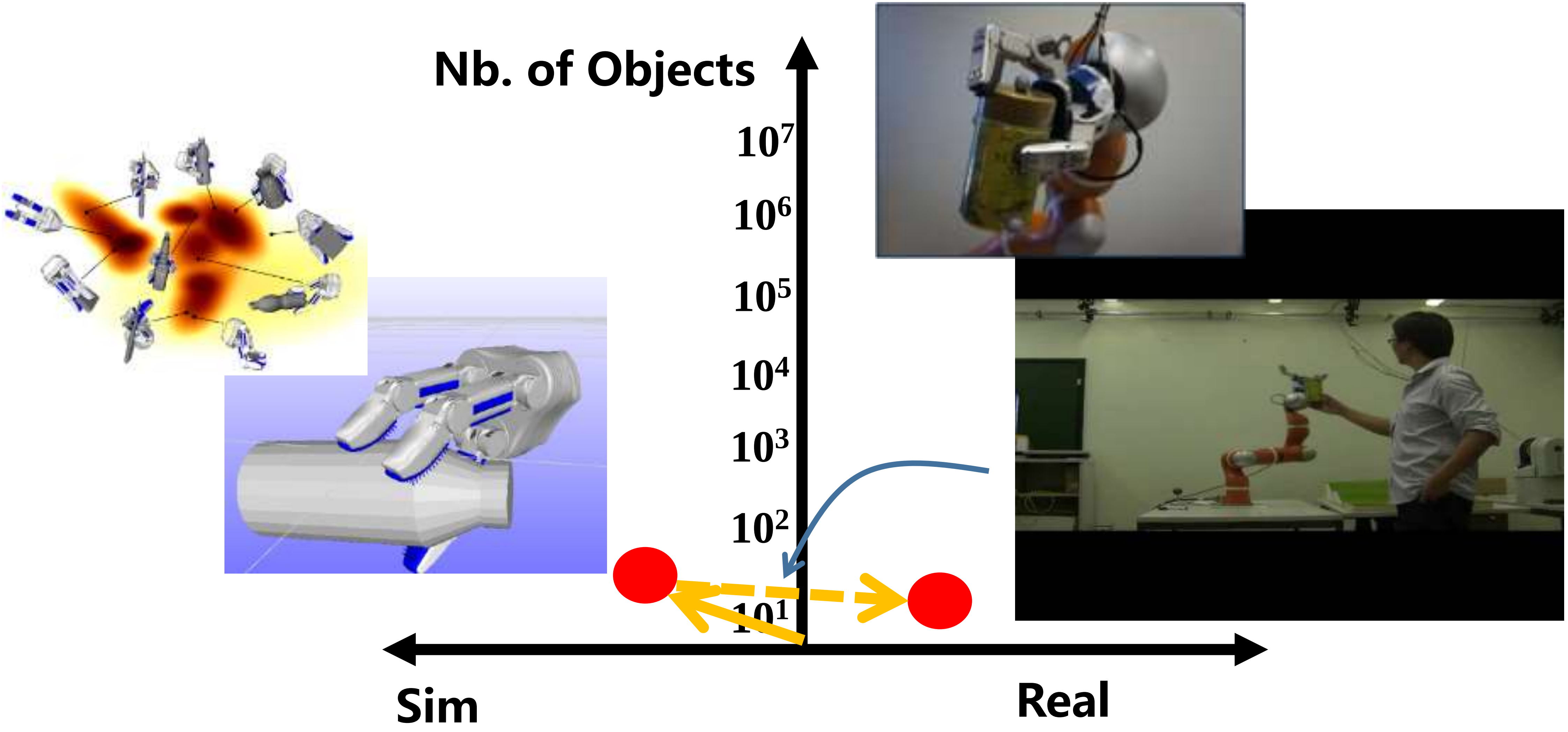


What will happen if a grasp fails?



Can we get a robust (and better) pre-trained planning algorithm before failures?

Recap of the grasp planning (1)



Model-based+ small data learning

Recap of the grasp planning (2)

ROBOTICS

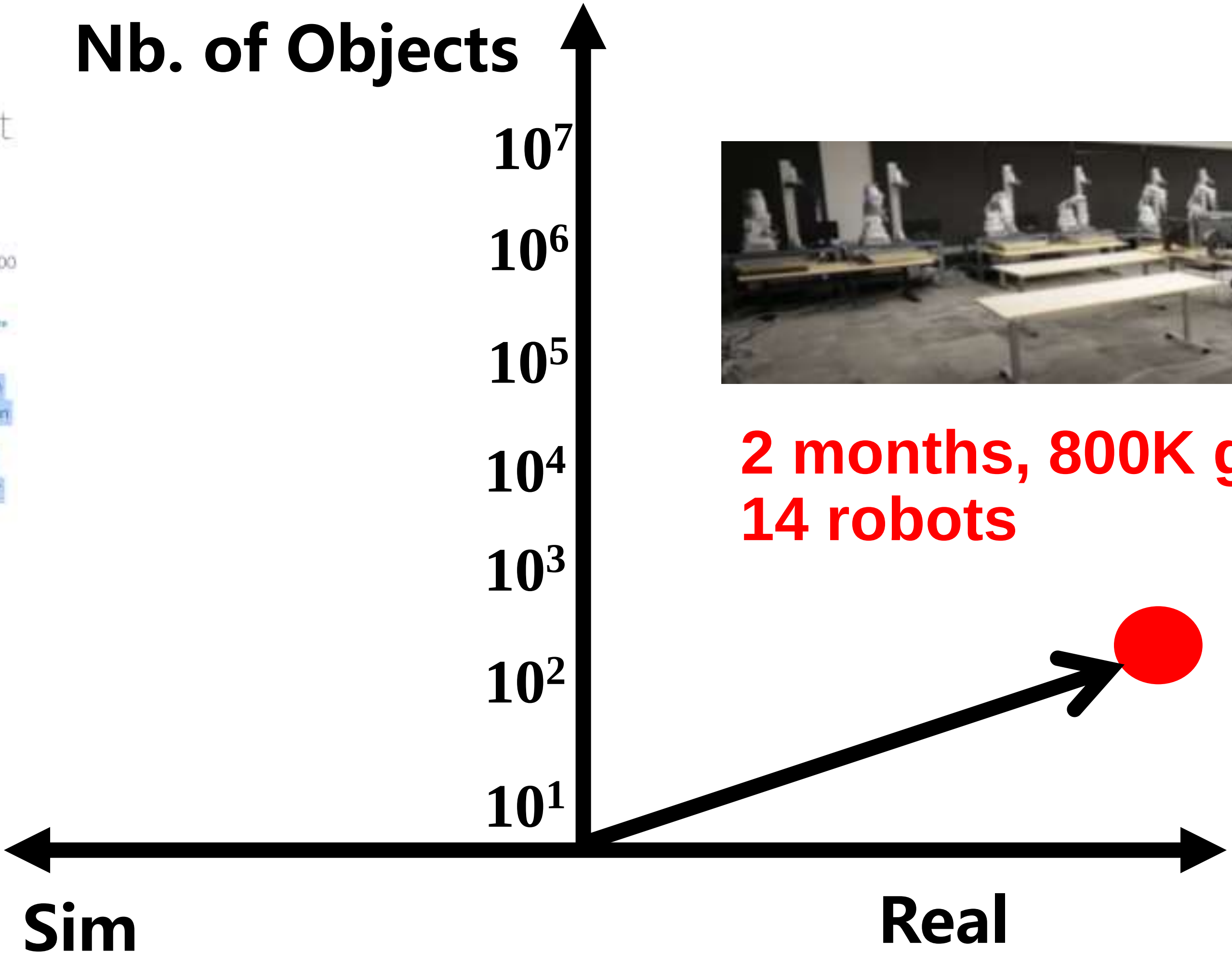
Amazon releases largest dataset for training "pick and place" robots

Dataset of images collected in an industrial setting features more than 190,000 objects, orders of magnitude more than previous datasets.

Share

In an effort to improve the performance of robots that pick, sort, and pack products in warehouses, Amazon has publicly released the largest dataset of images captured in an industrial product-sorting setting. Where the largest previous dataset of industrial images featured on the order of 100 objects, the Amazon dataset, called ARMBench, features more than 190,000 objects. As such, it could be used to train "pick and place" robots that are better able to generalize to new products and contexts.

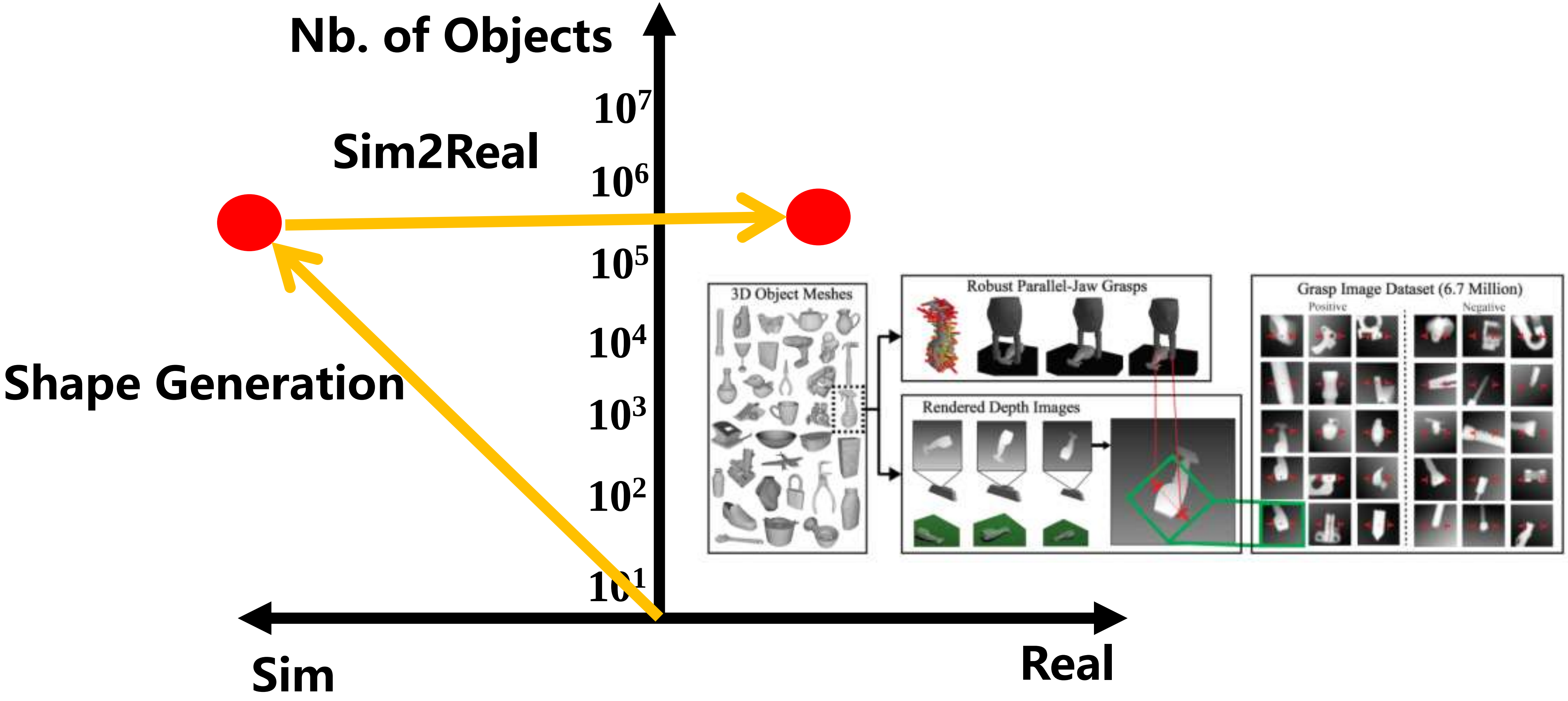
190000 objects



2 months, 800K grasps
14 robots

Big data in real world

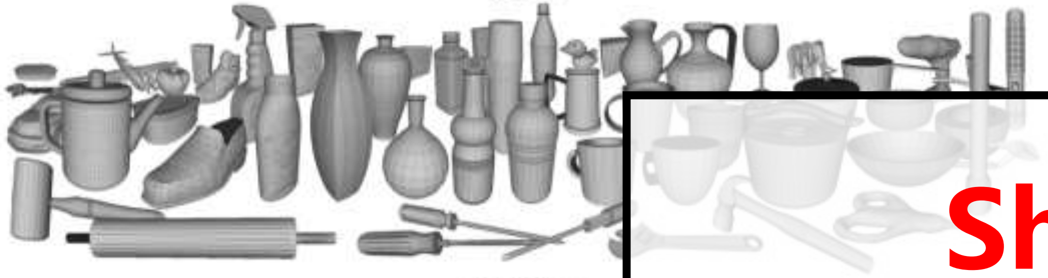
Recap of the grasp planning (3)



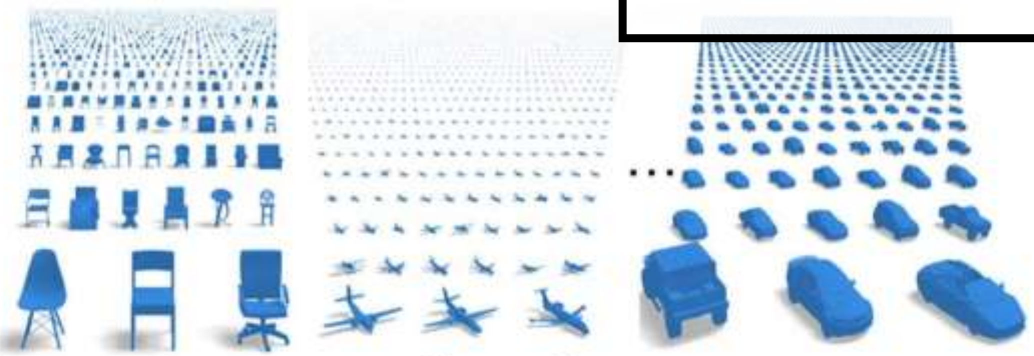
Big data in simulation + Sim2Real

Shape generation

YCB

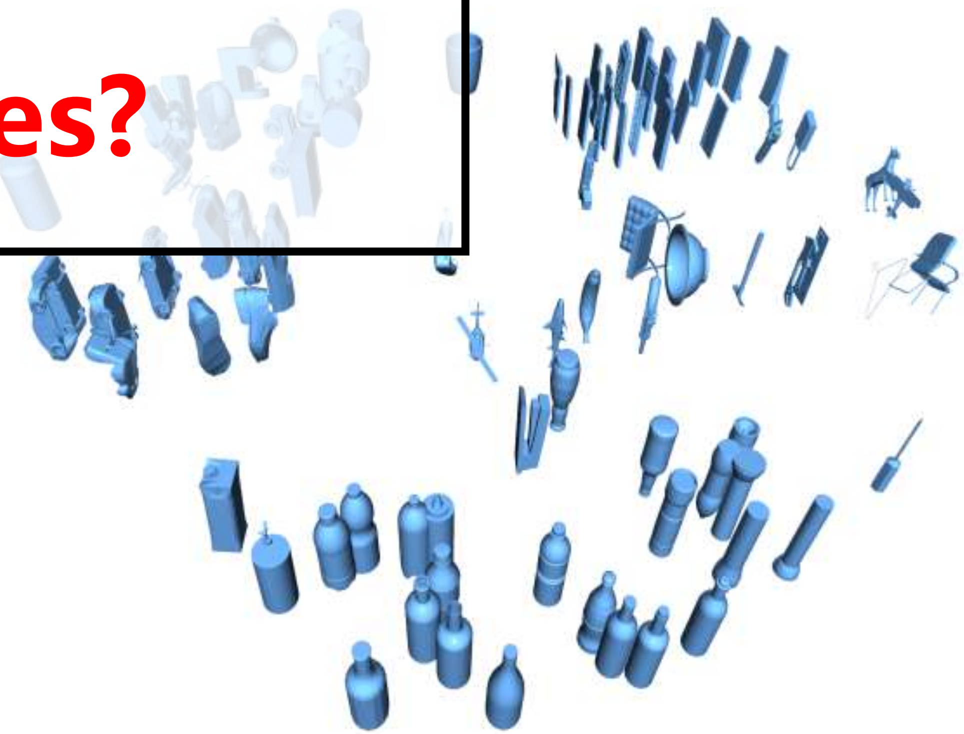


3DNet

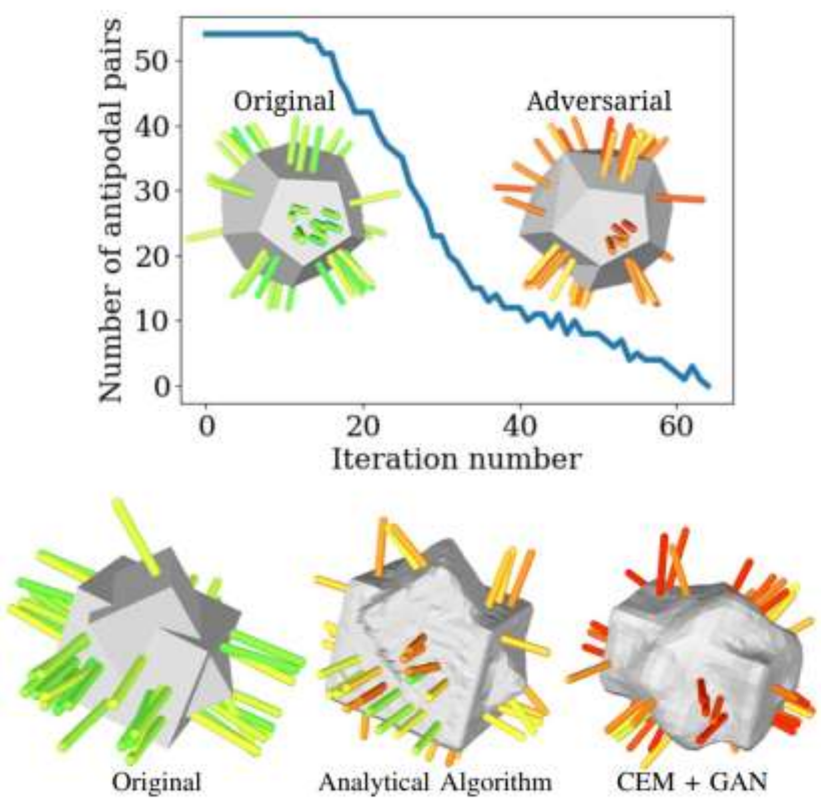


ShapeNet

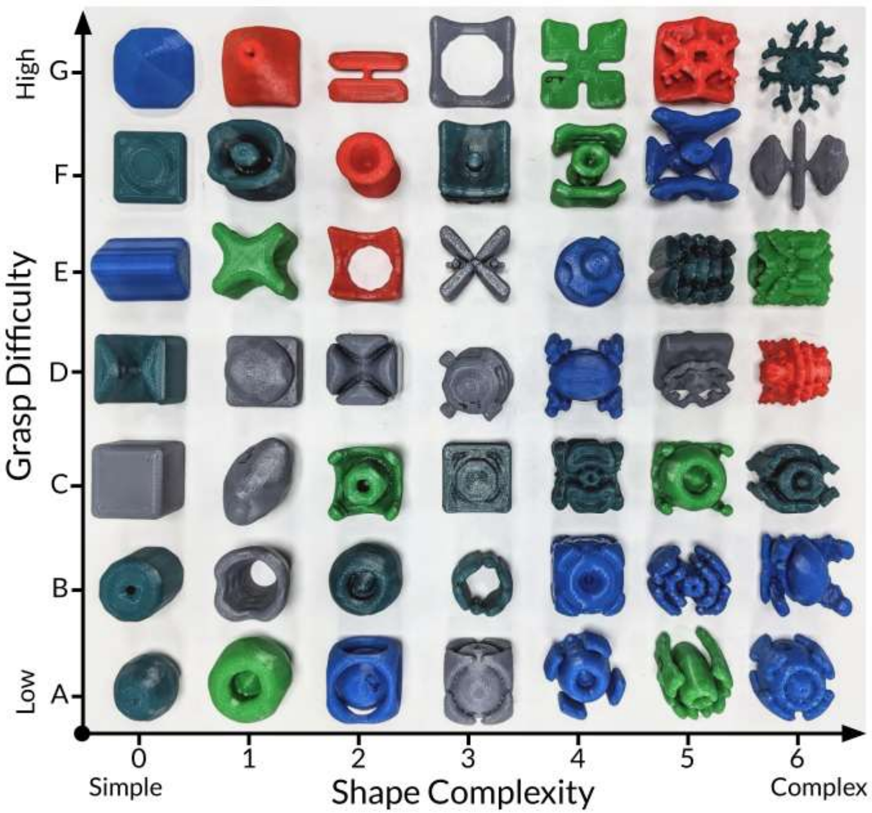
Shapes?



Shape generation- Previous works



Adversarial Grasp Object
(Wang et. Al 2019)



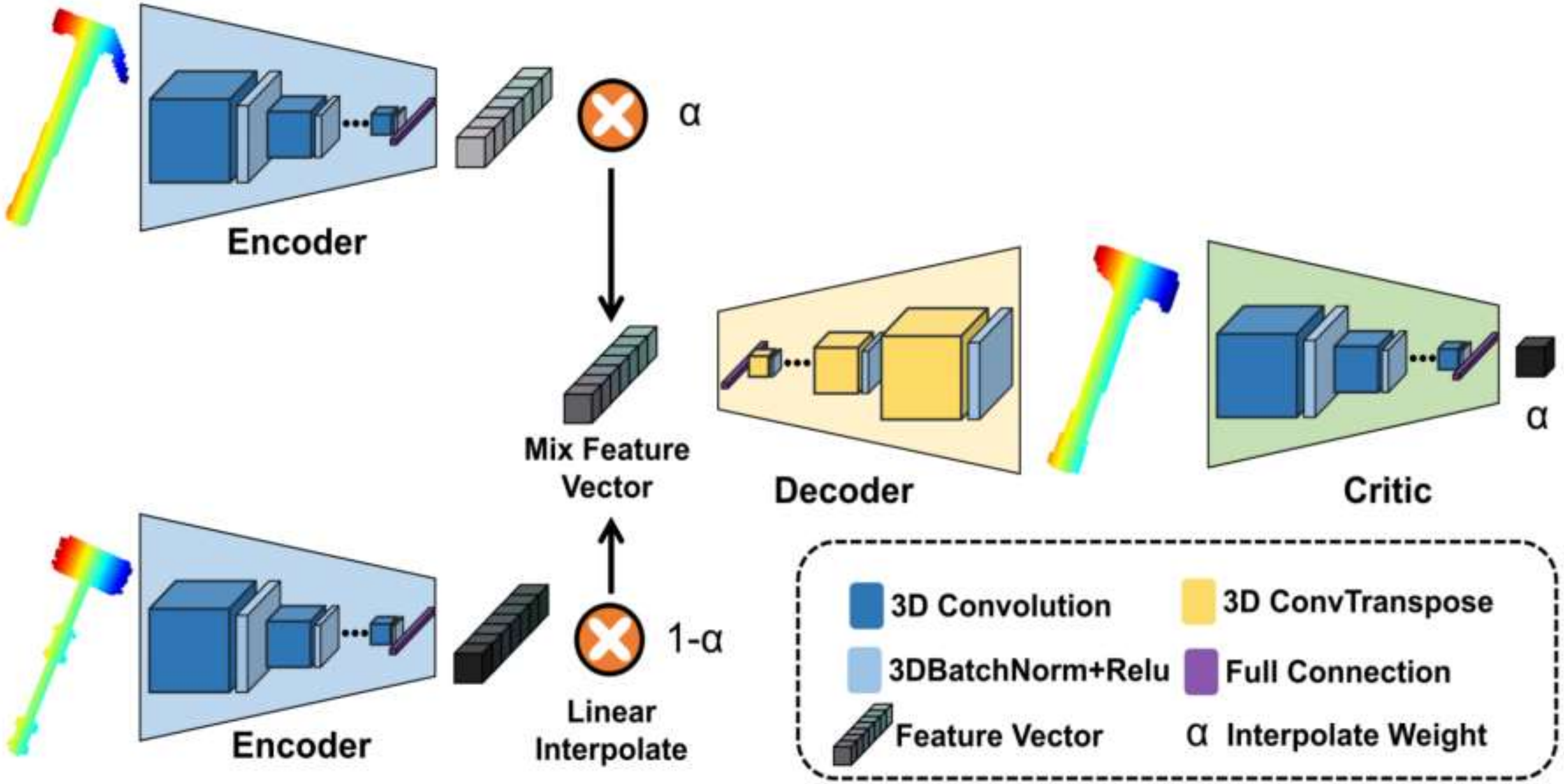
EGAD (Morrison et. Al 2020)

A class of “adversarial grasp objects that are physically similar to a given object but significantly less ”graspable” in terms of a specified robot grasping policy.

The objects in EGAD are geometrically diverse, filling a space ranging from simple to complex shapes and from easy to difficult to grasp

Shape generation

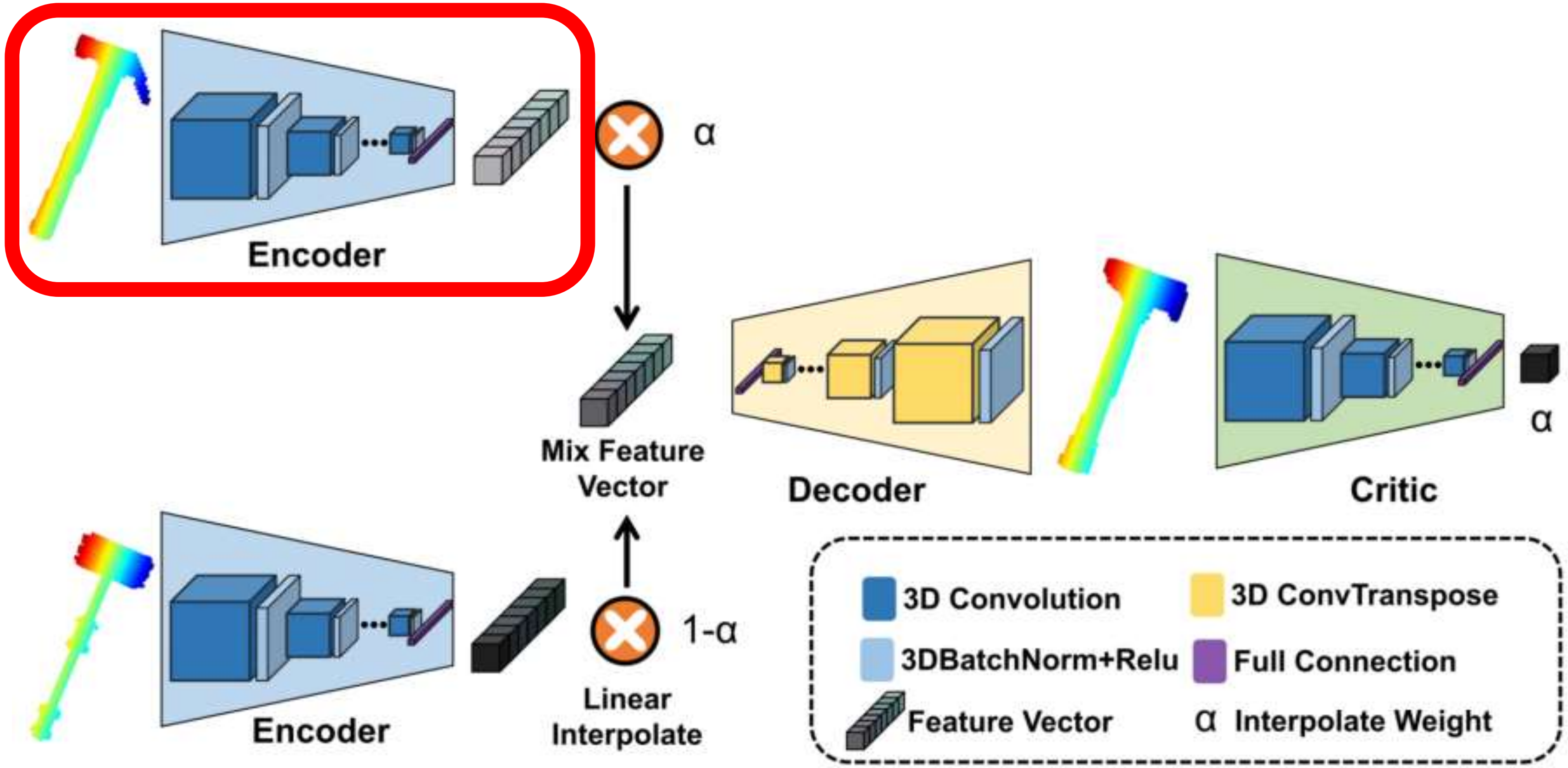
AE-Critic Network



- Deep shape generation for robust grasping， Sci-China, 2023 (under review)
- Improving robotic grasping ability through deep shape generation, 2022

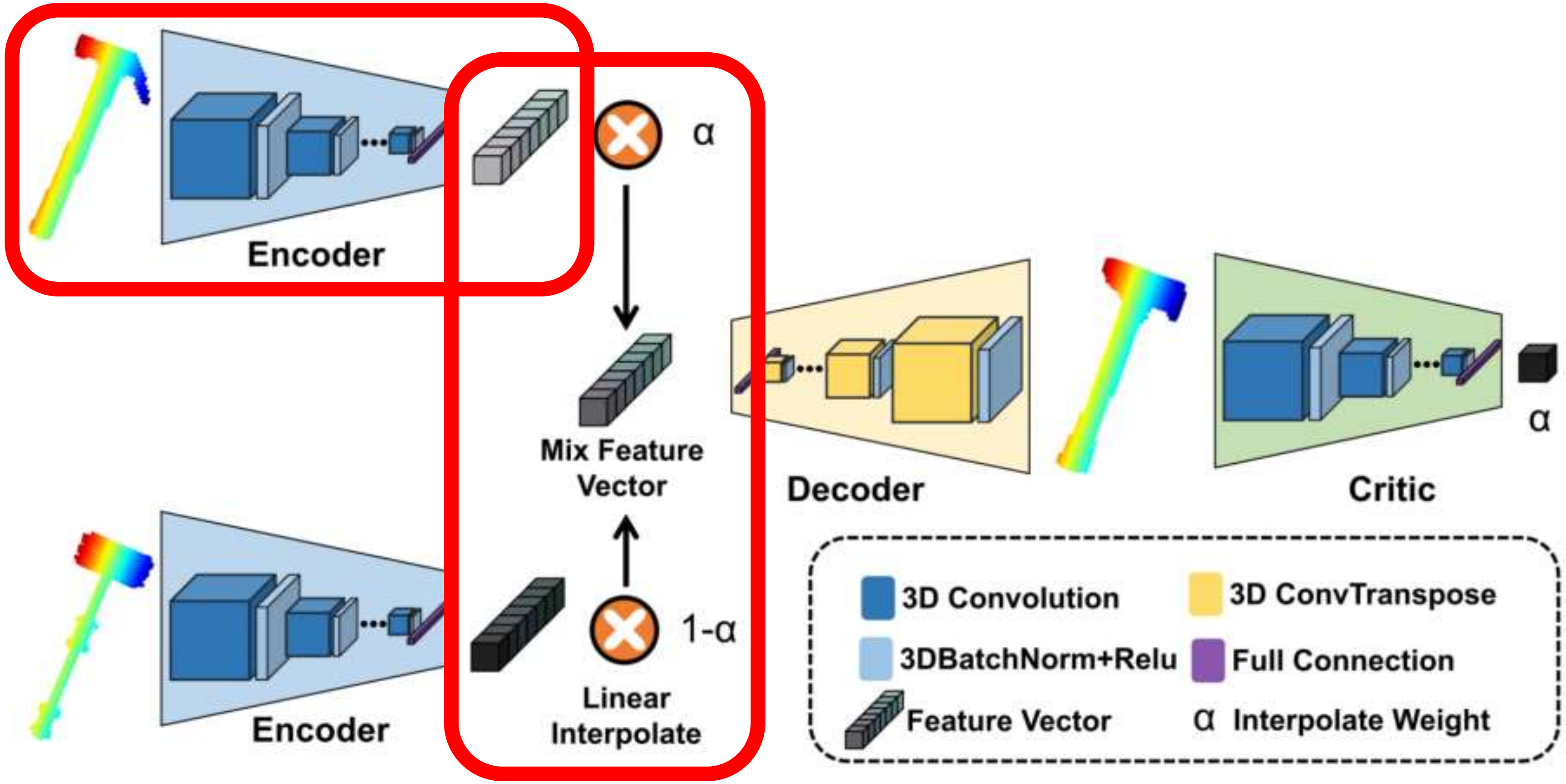
Object shape encoding

- **Encode shapes into feature vectors**



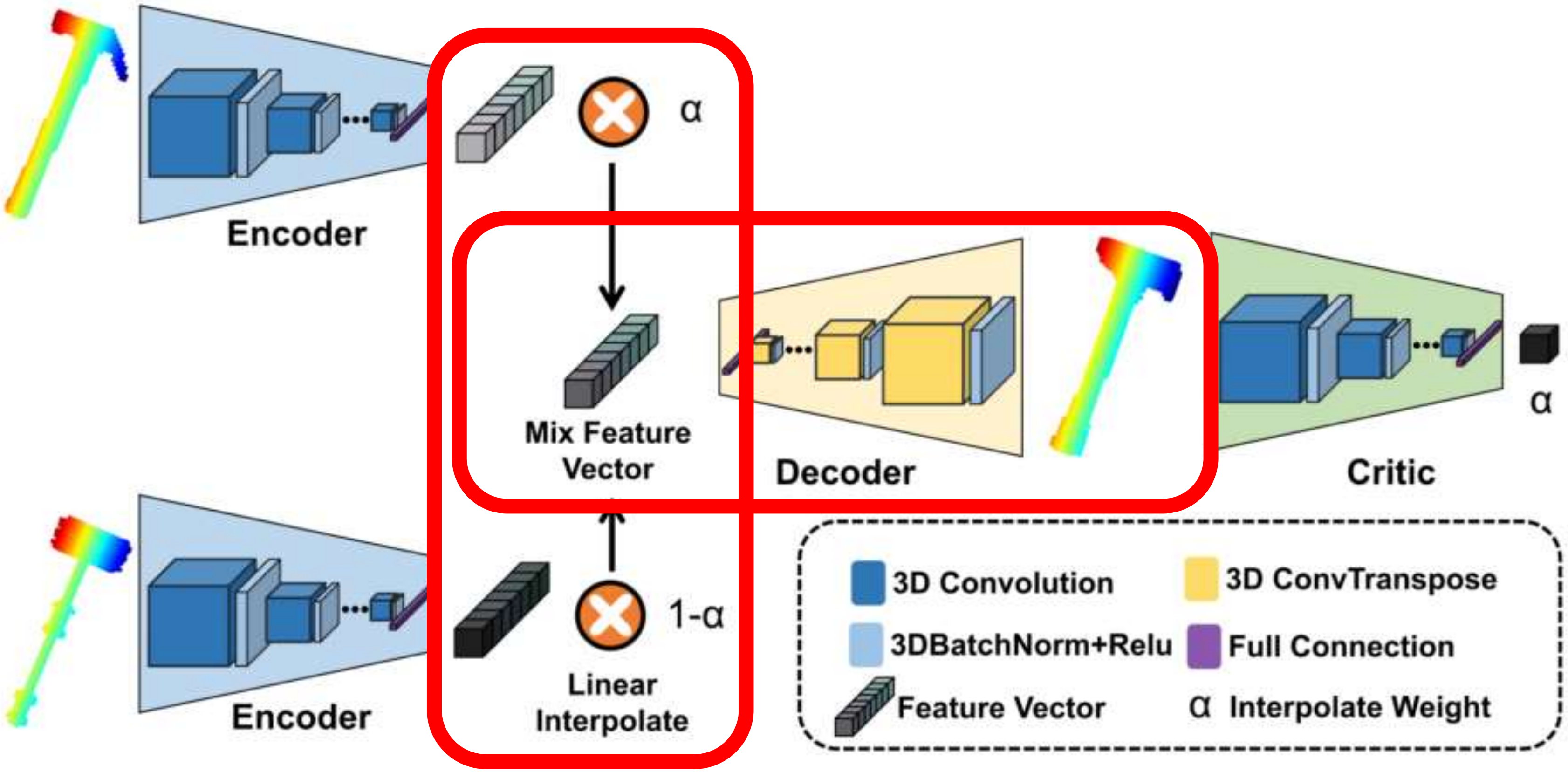
Object shape encoding

- **Encode shapes into feature vectors**
- **Interpolate feature vectors**

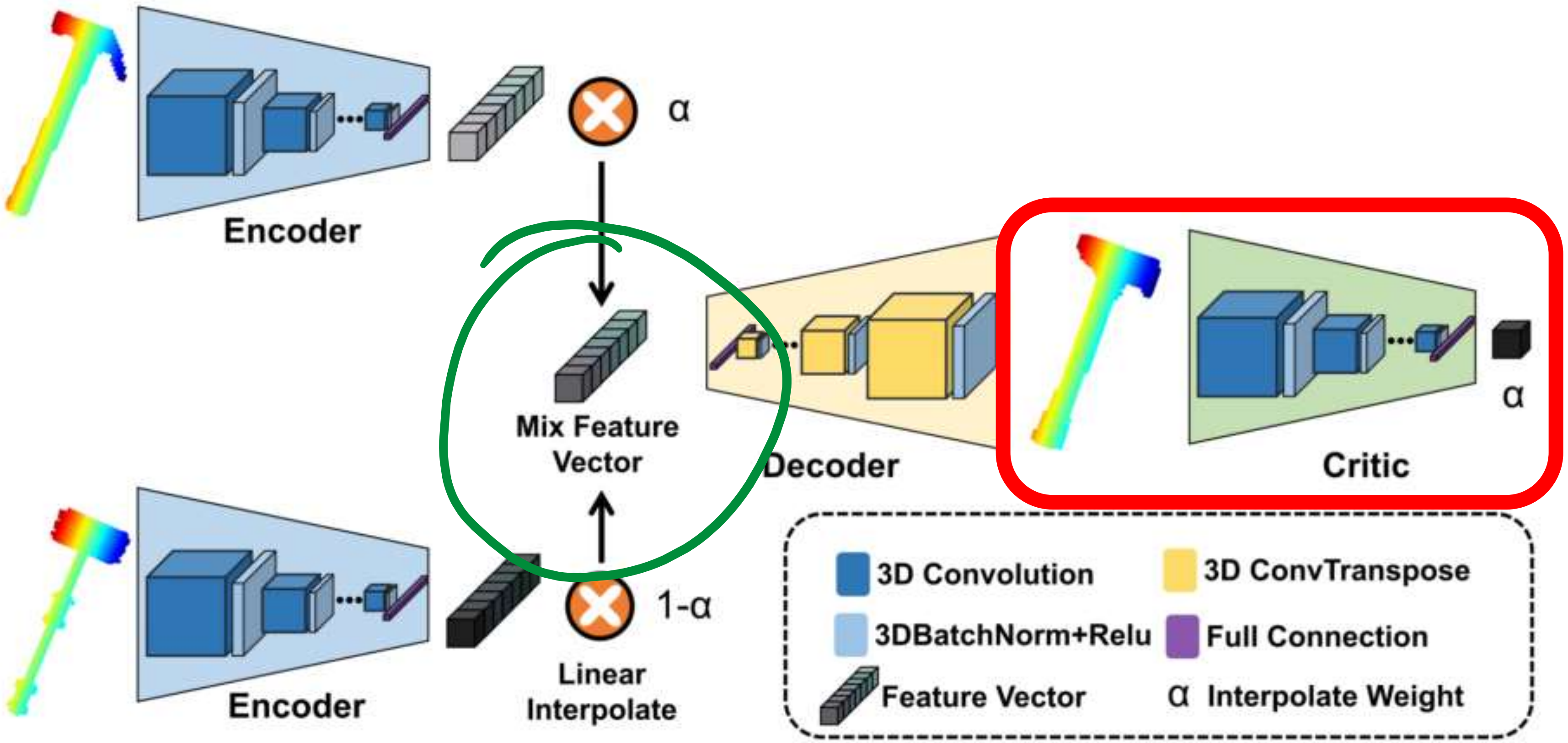


Object shape encoding

- **Encode shapes into feature vectors**
- **Interpolate feature vectors**
- **Decode mix feature vector and generate new shapes**



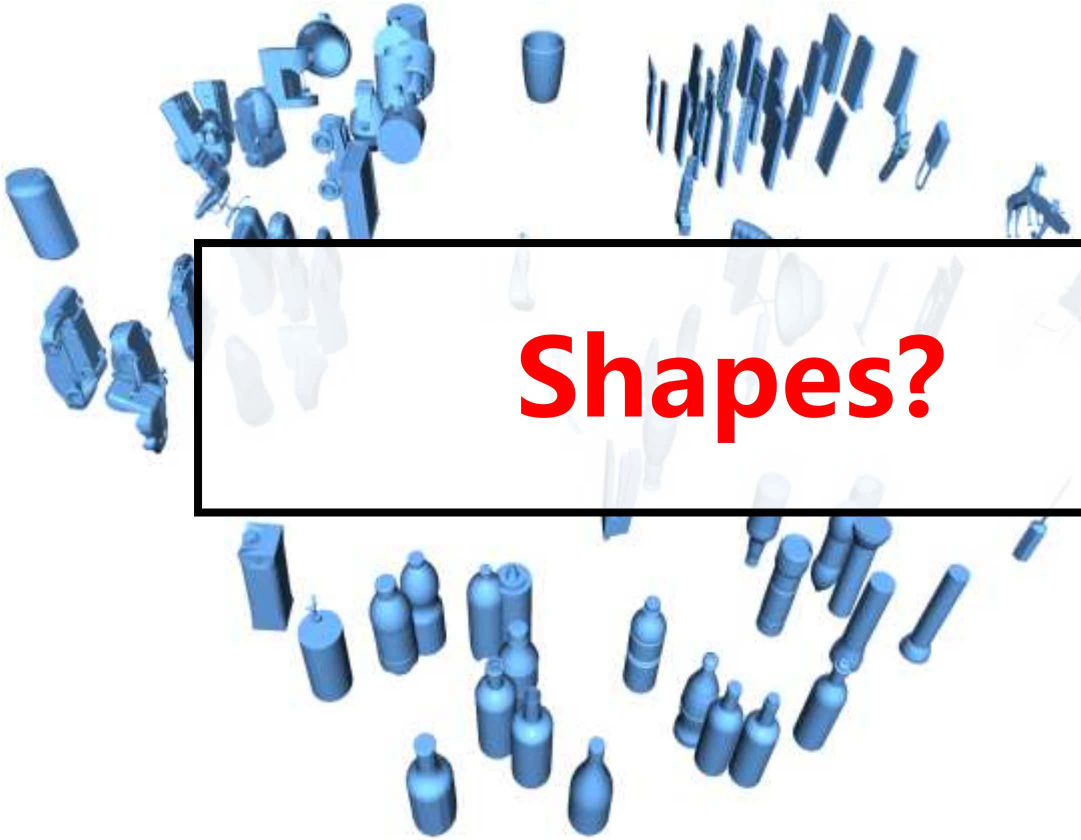
Object shape encoding



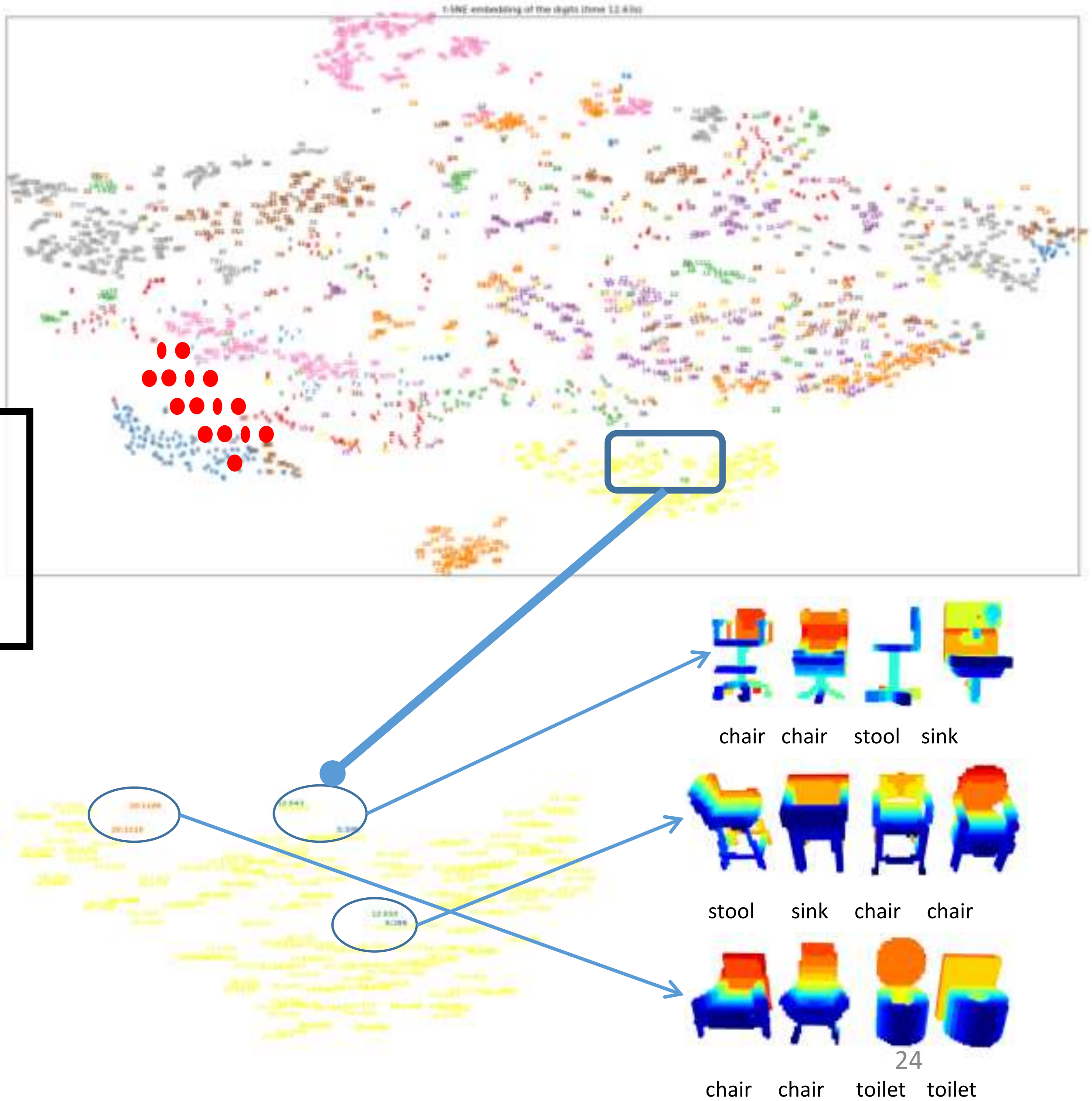
Critic is used to predict the interpolate weights and regularize the generated shapes to be more realistic

Object shape distribution

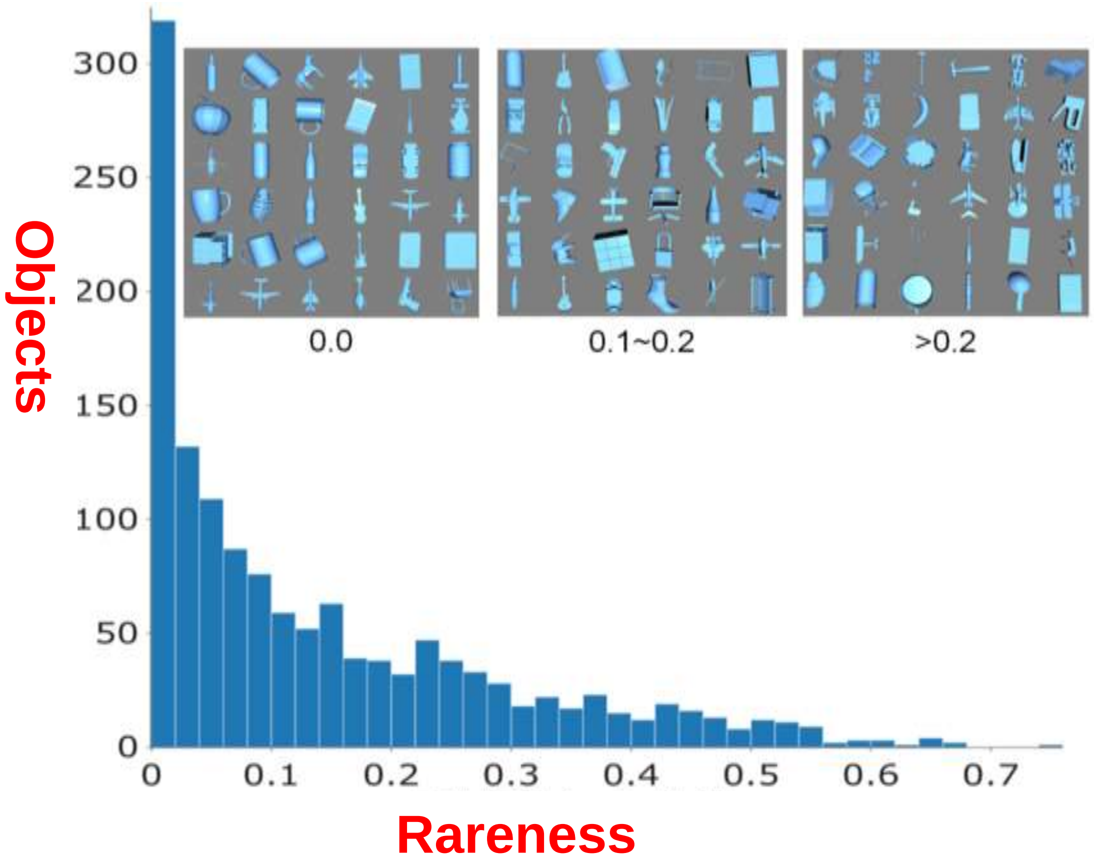
Shape Distribution

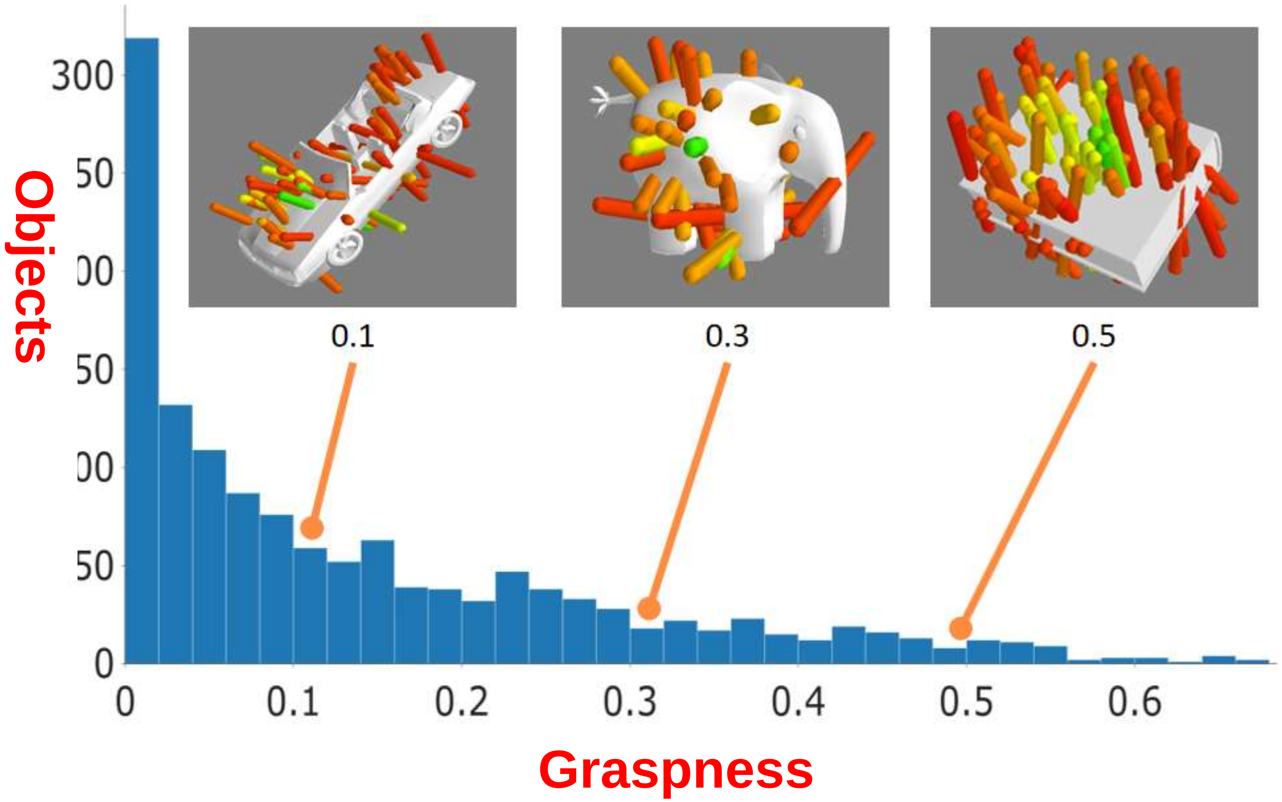


Shapes?



Shape generation

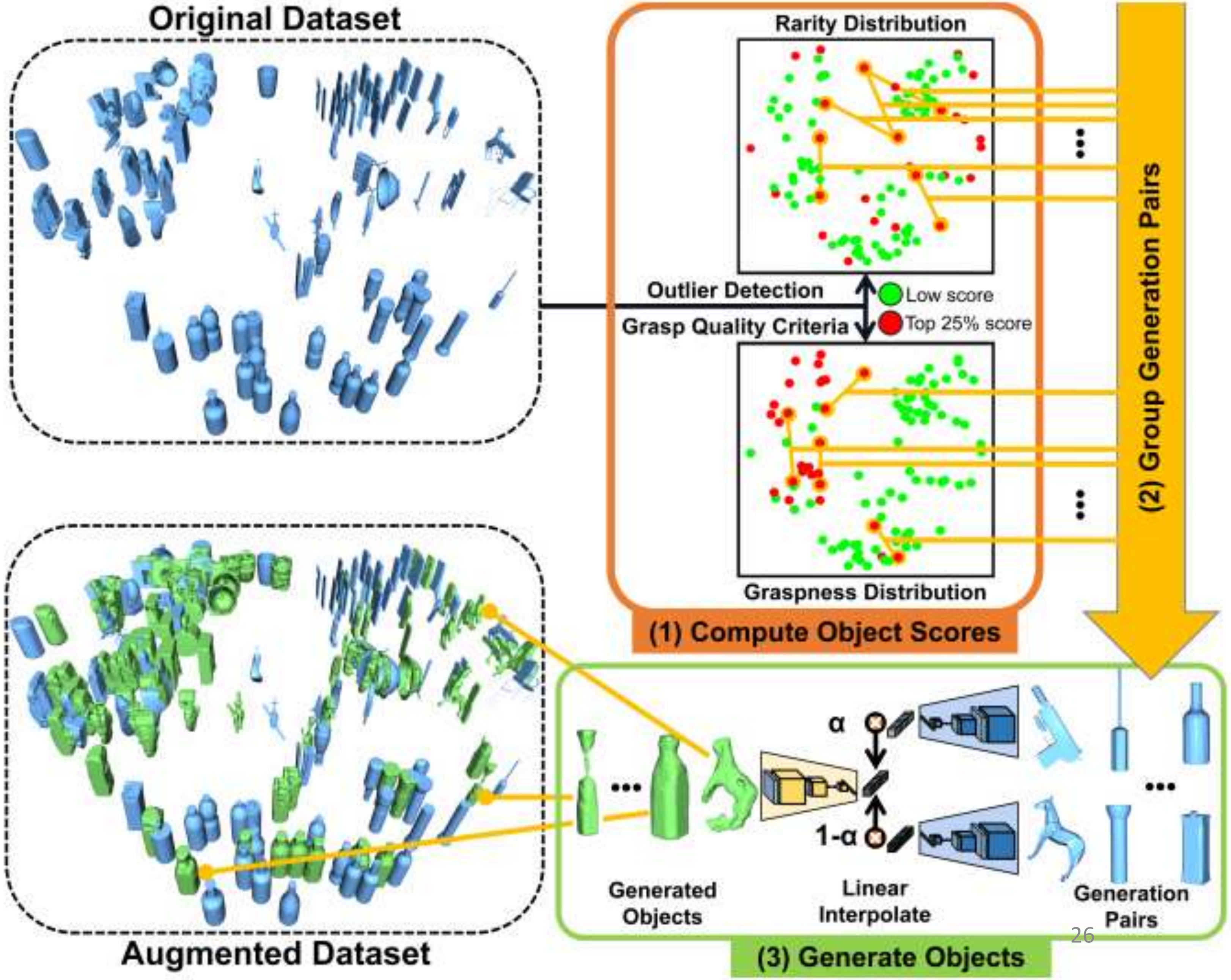




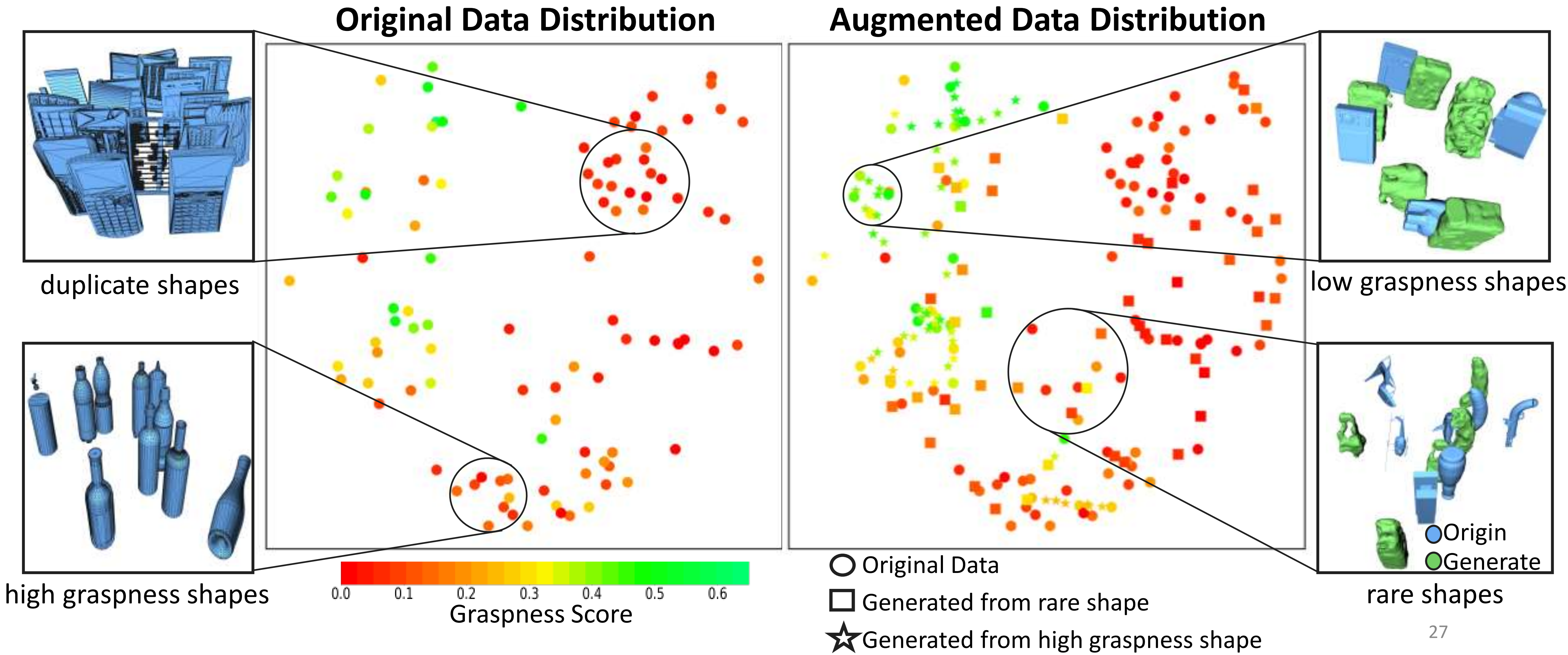
Rare + Difficult to grasp

Shape generation

- **Compute object scores through outlier detection and grasp-quality criteria**
- **Group high-scoring data as generation pair**
- **Generate new objects by AE-Critic network**

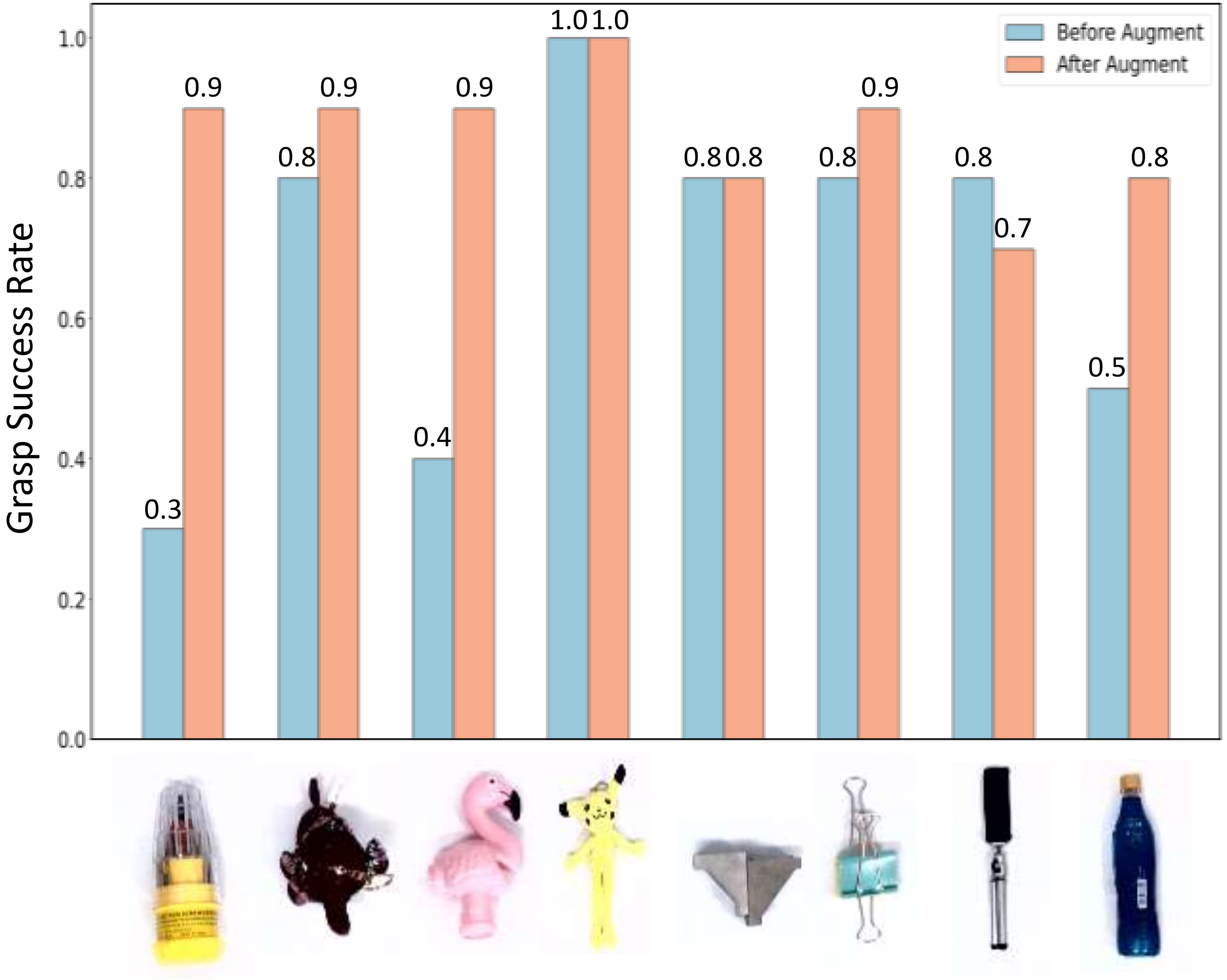


Shape generation

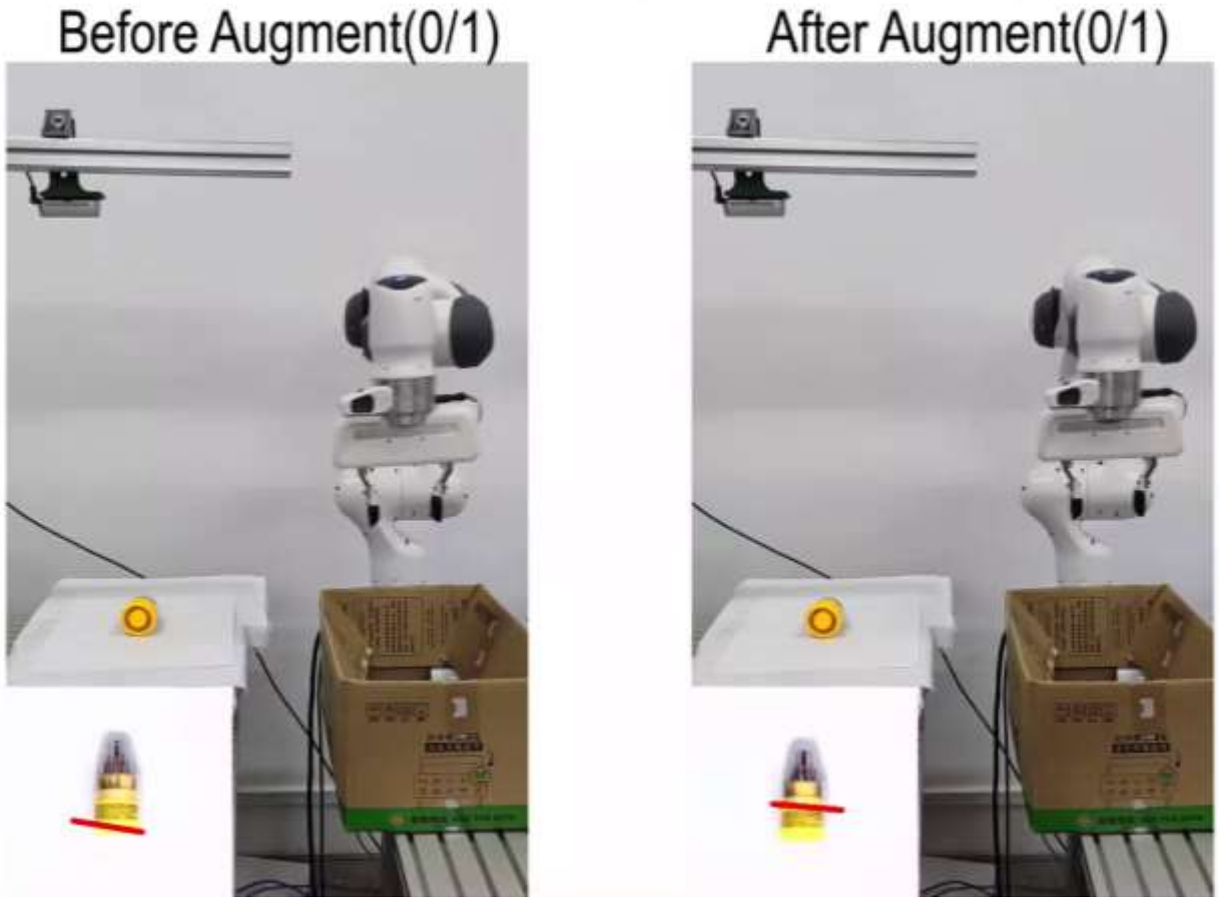


Real-Robot Experiment

Grasp Success Rate Comparison between
Before and After Augmentation (Average 68% : 86%)

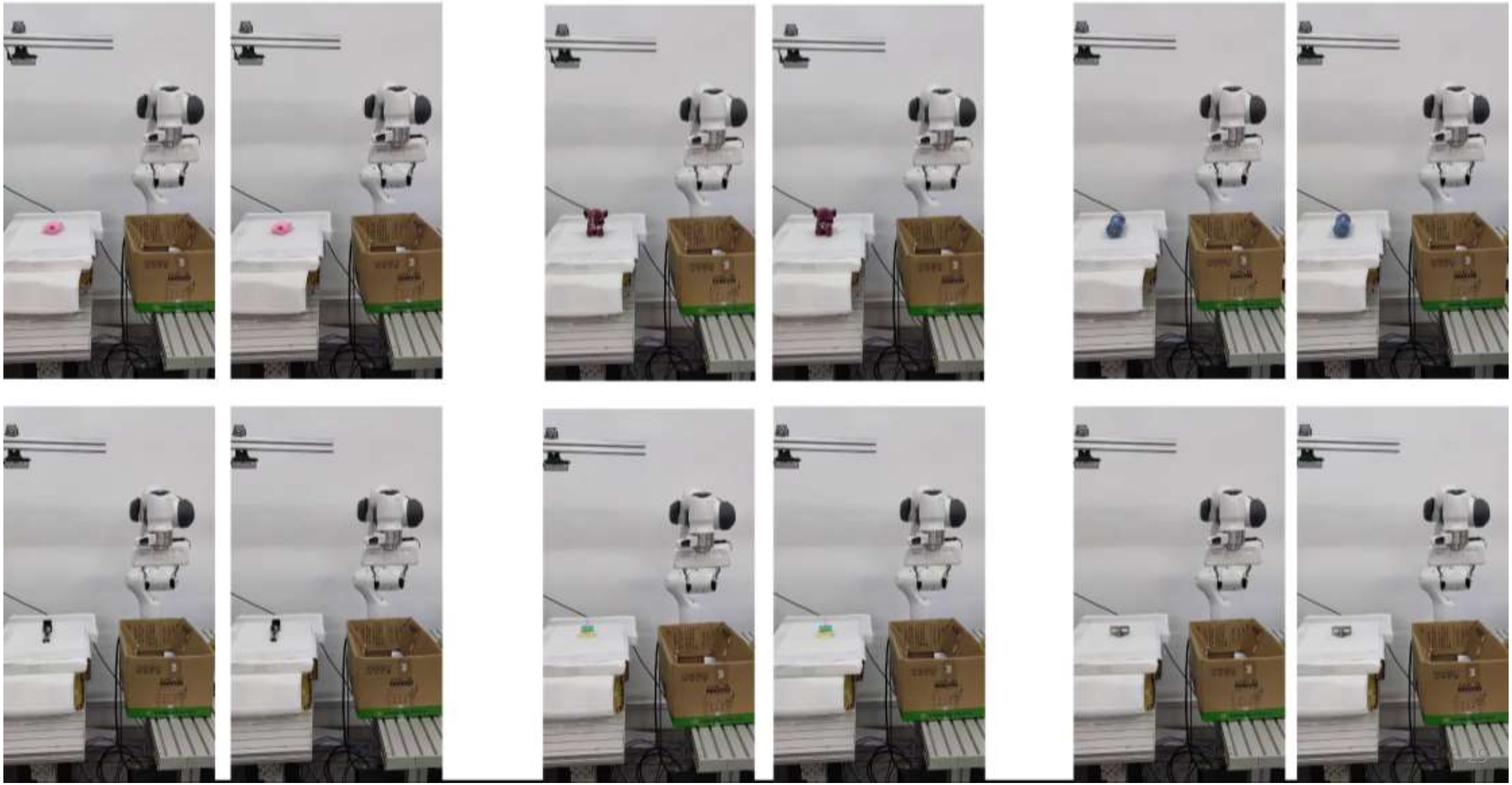


Grasp Trials(Success/All)

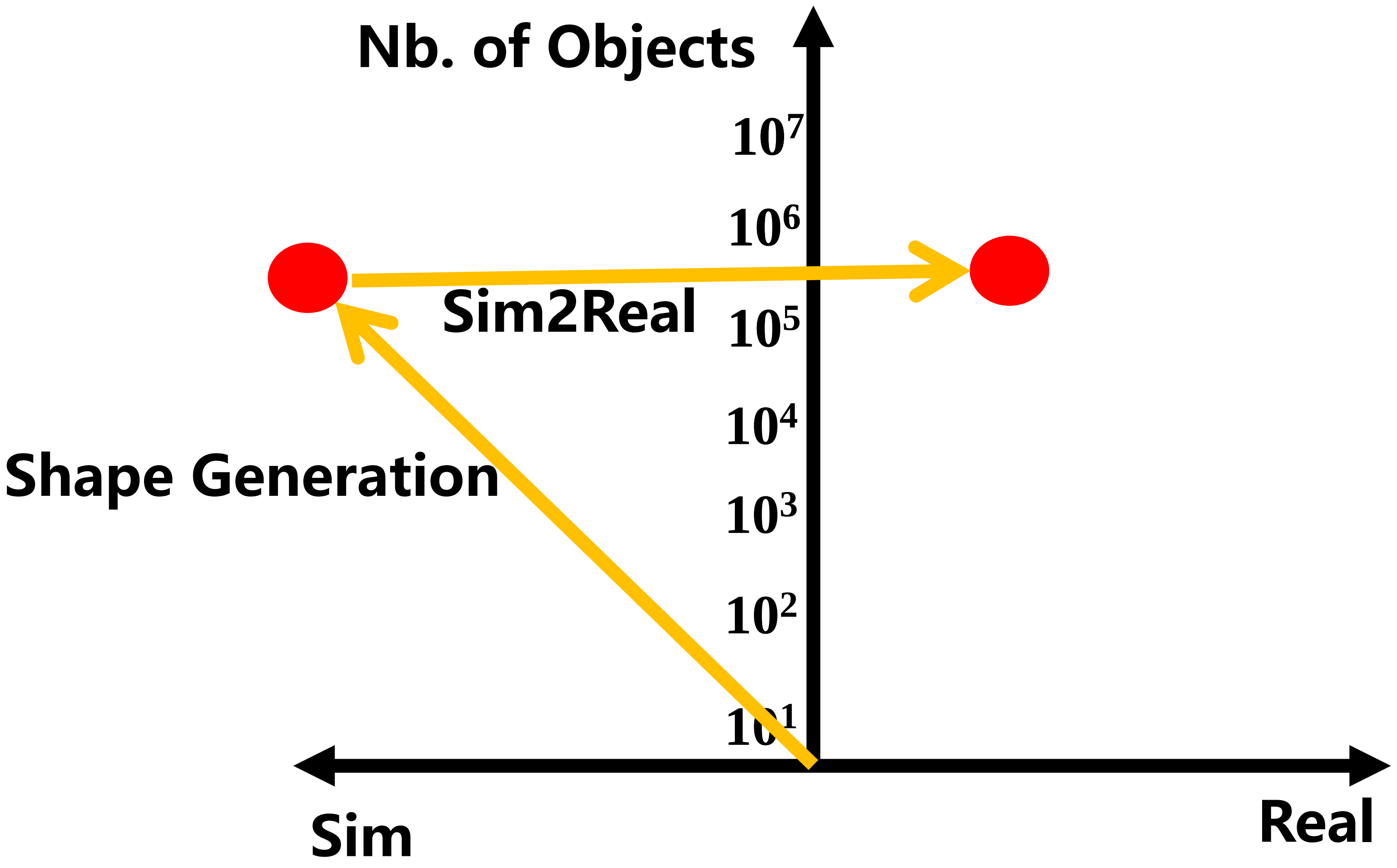


Experiments Results

Before After Before After Before After



Recap of the grasp planning (3)



Big data in simulation + Sim2Real

Grasp transfer

Household **Factory** **Simulation**

Real-world Dataset:
Labeled training data from real-world scenarios

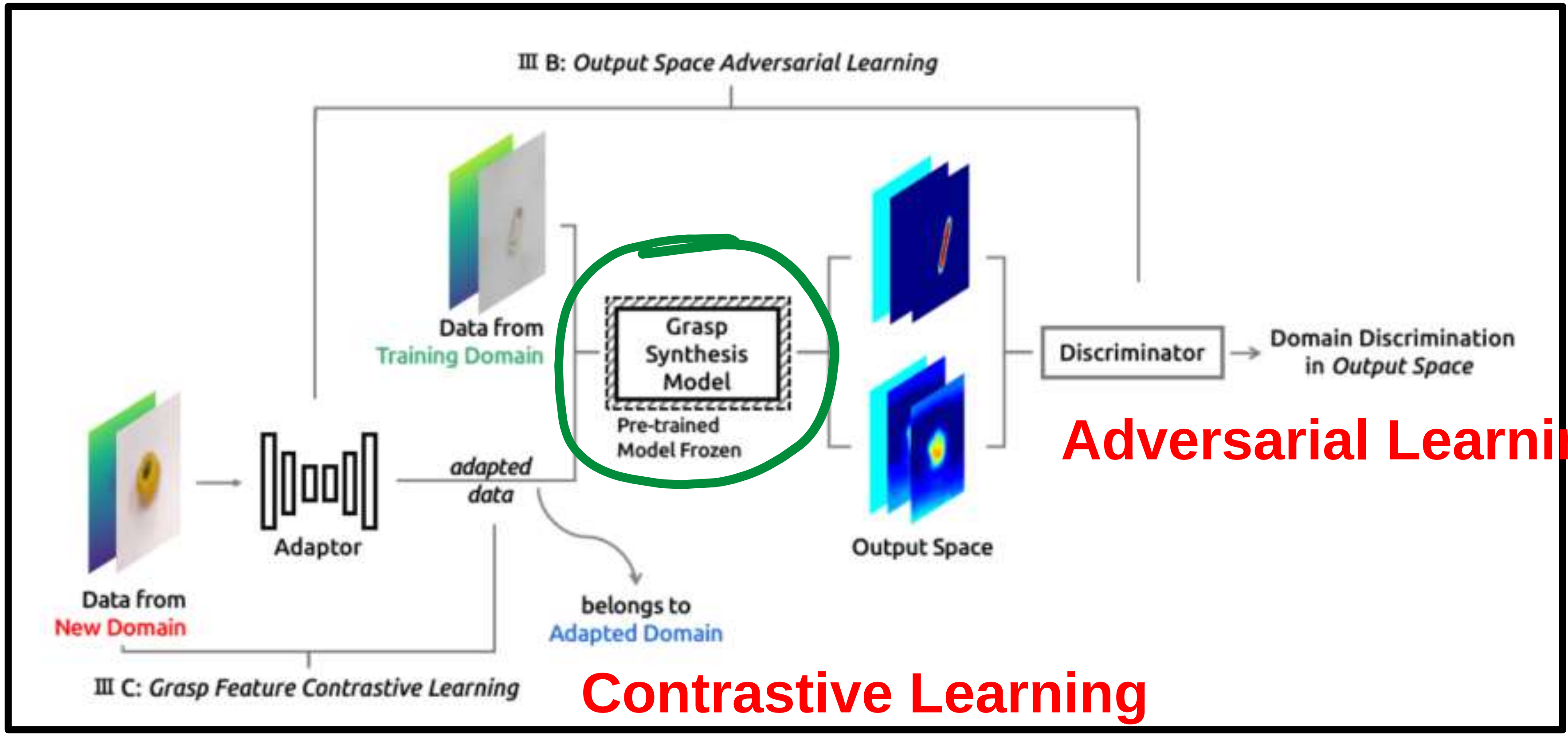
Drawbacks: *expensive, gap between different scenarios*

Synthetic Dataset:
Generating Labeled Data from Simulation

Drawbacks: *simulation-to-reality gap*

How to transfer the learned grasping ability to new domains?

Sim2Real (Joint Work with Fei Chen, Yasemin Bekiroglu)

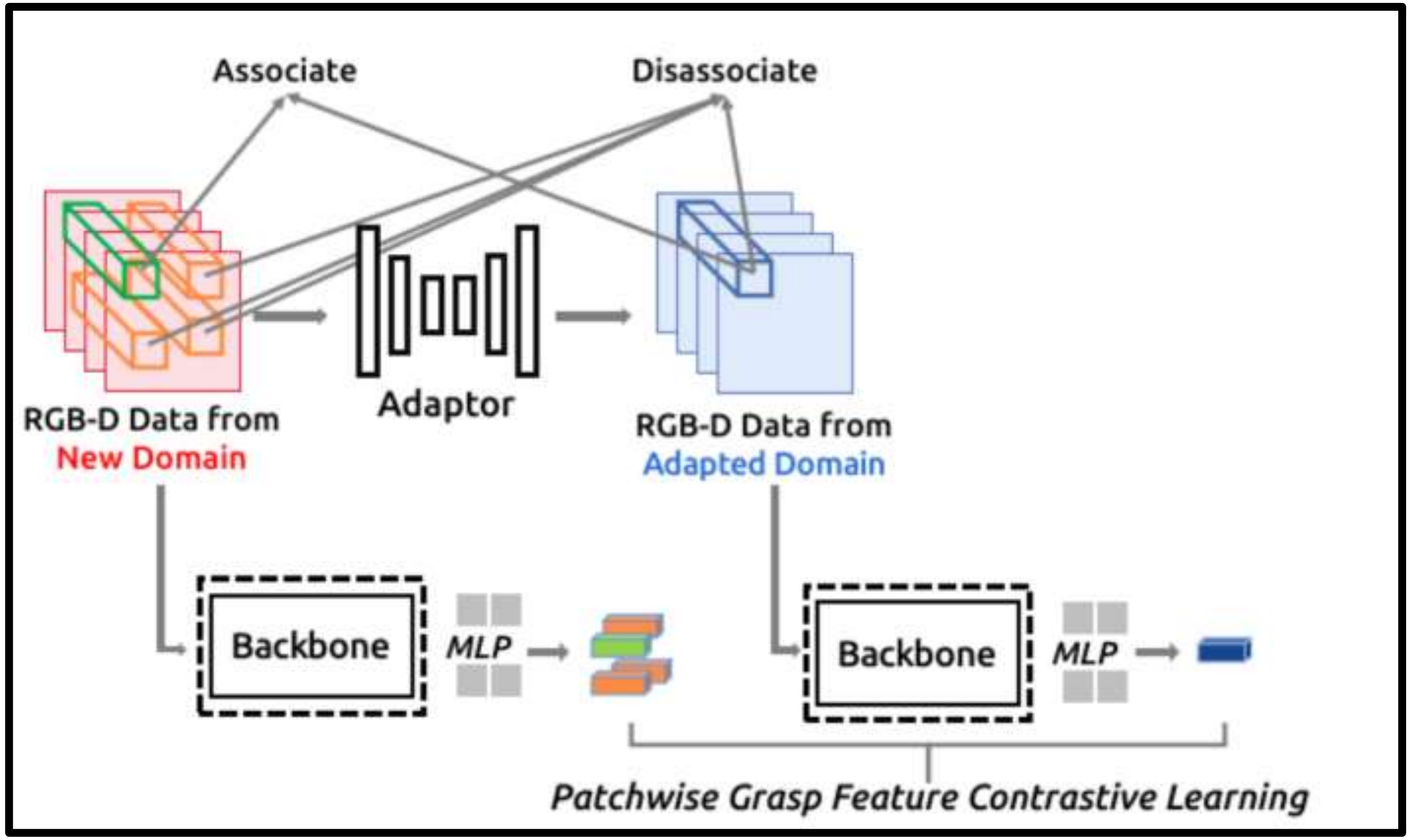


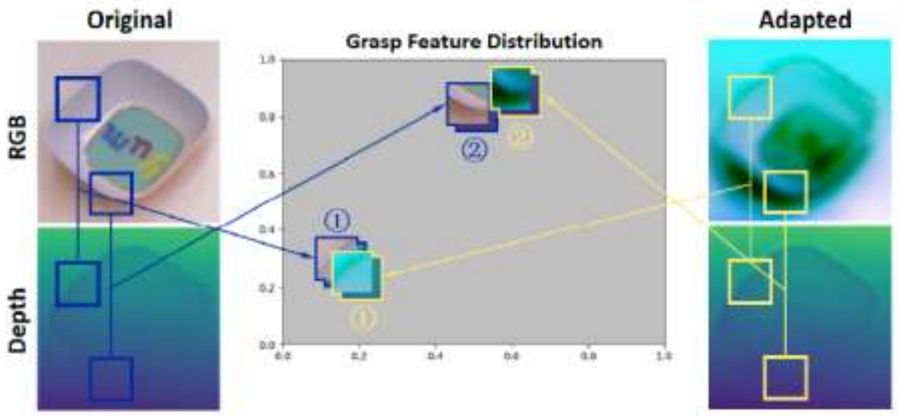
Grasp Adaptation

- **Output Space Adversarial Learning** — **What to transfer?**
- and
- **Grasp Feature Contrastive Learning** — **How to transfer?**

GraspAda: Deep Grasp Adaptation through Domain Transfer, ICRA 2023

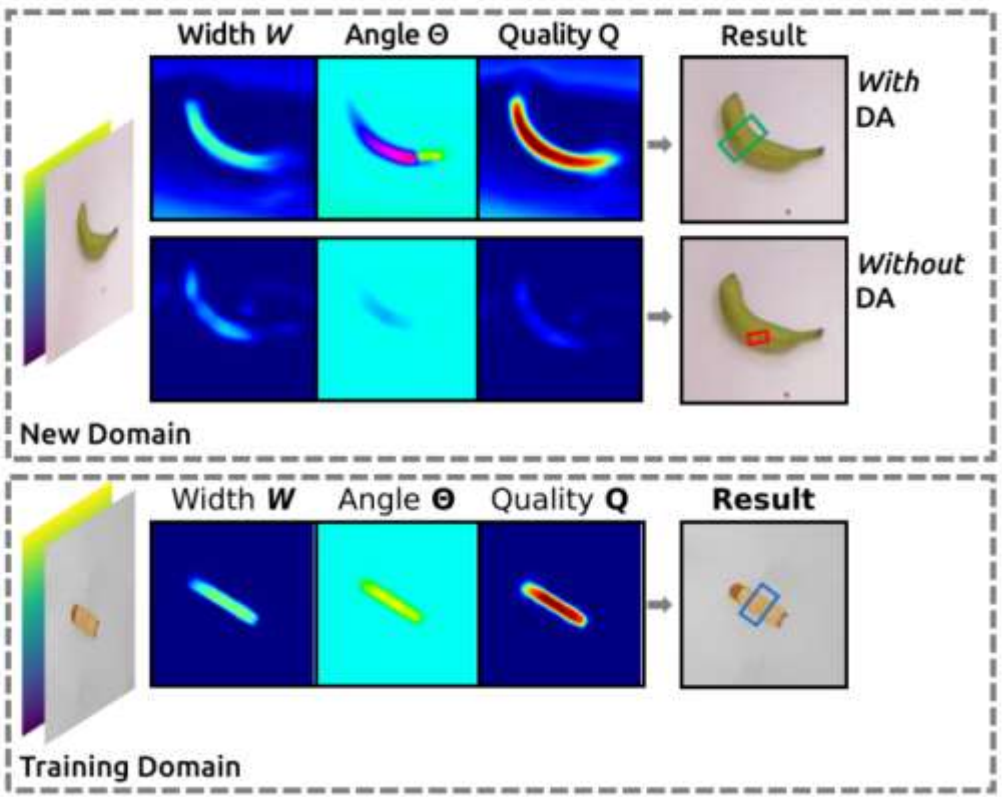
GraspAda





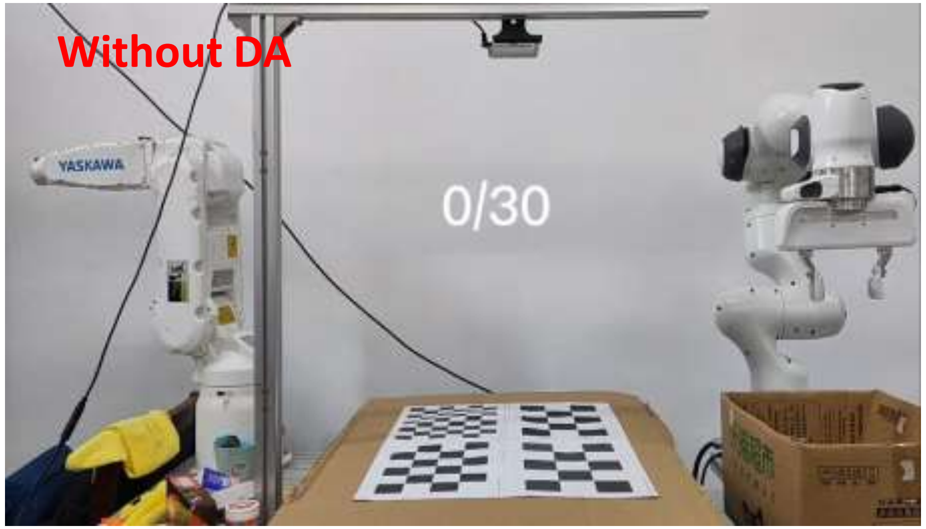
A **feature-level contrastive learning scheme** is developed to enforce the grasp relative feature **consistency** during adaptation.

GraspAda

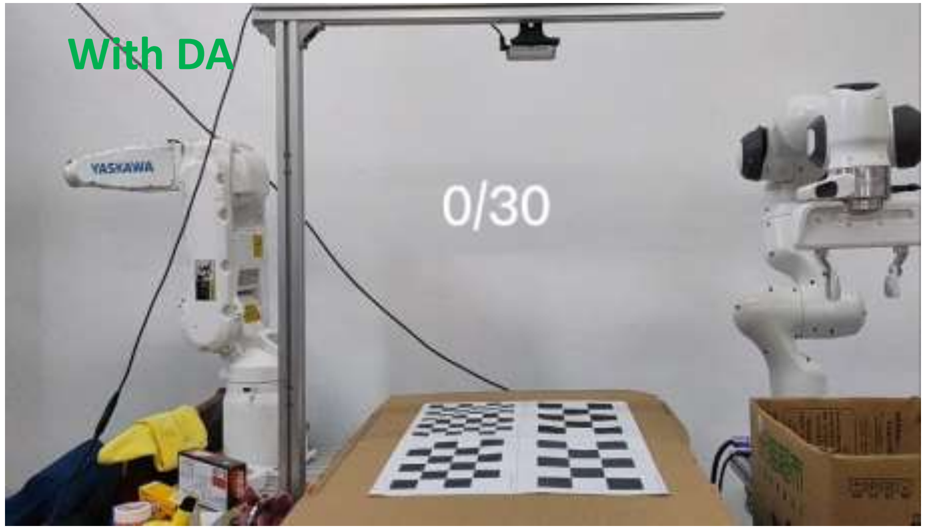

Width W
Angle Θ
Quality Q
Result
With DA
Without DA
New Domain
Width W
Angle Θ
Quality Q
Result
Training Domain

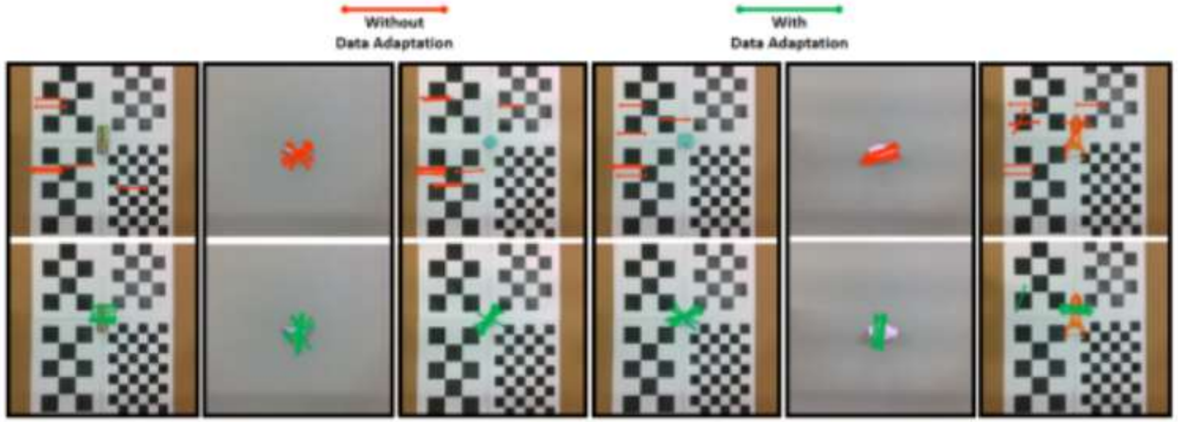
GraspAda

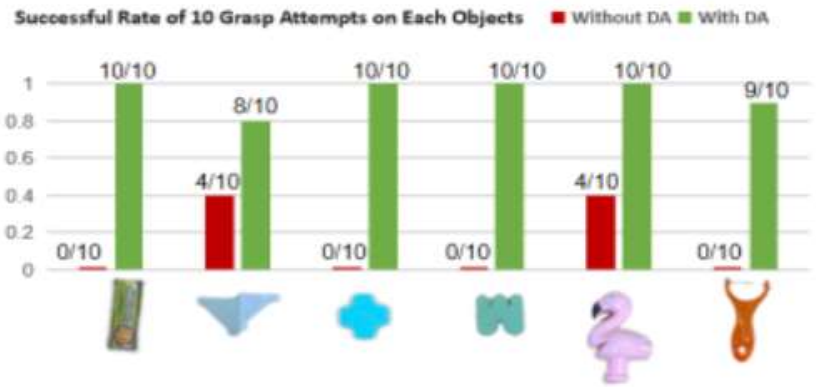
Overall grasping success rate: 40%



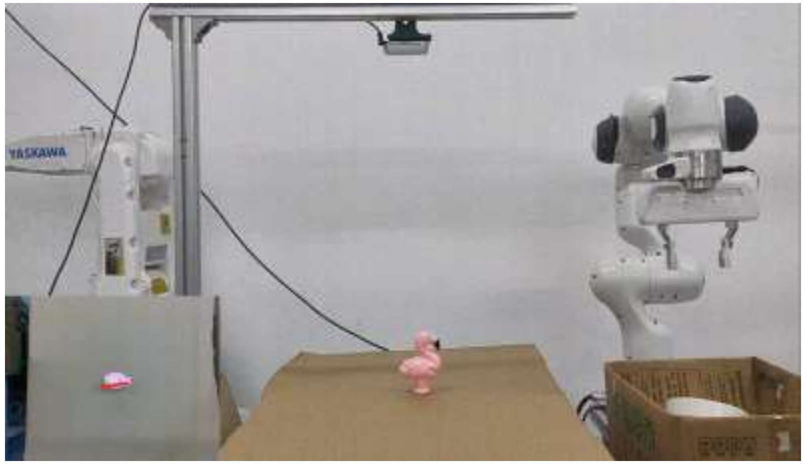
Overall grasping success rate: 80%



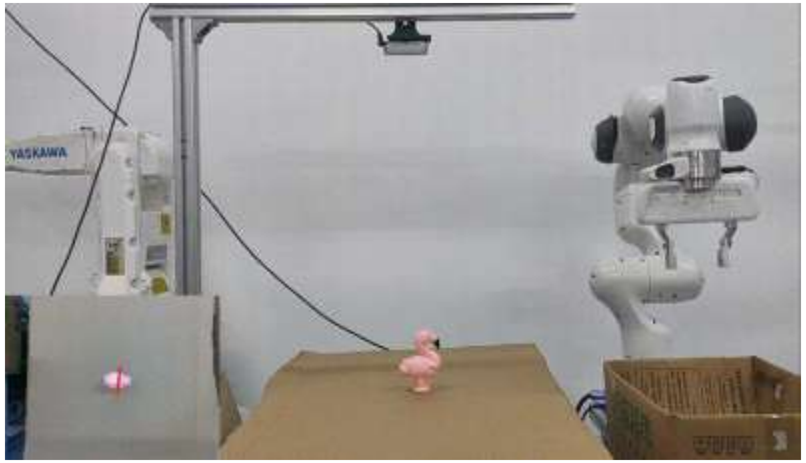




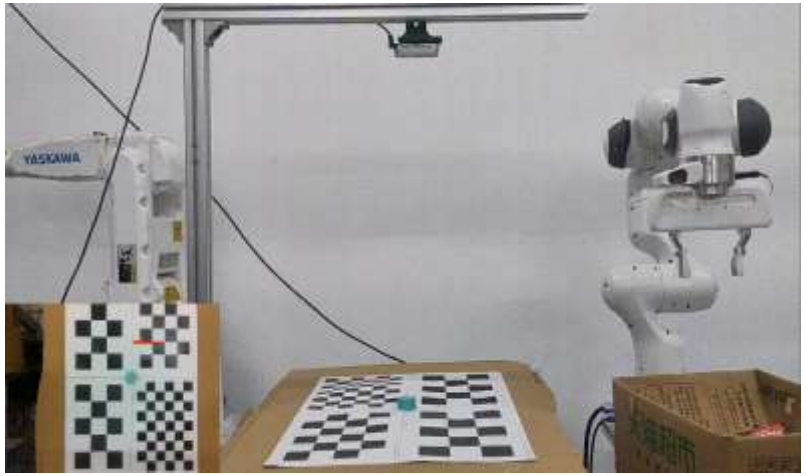
GraspAda



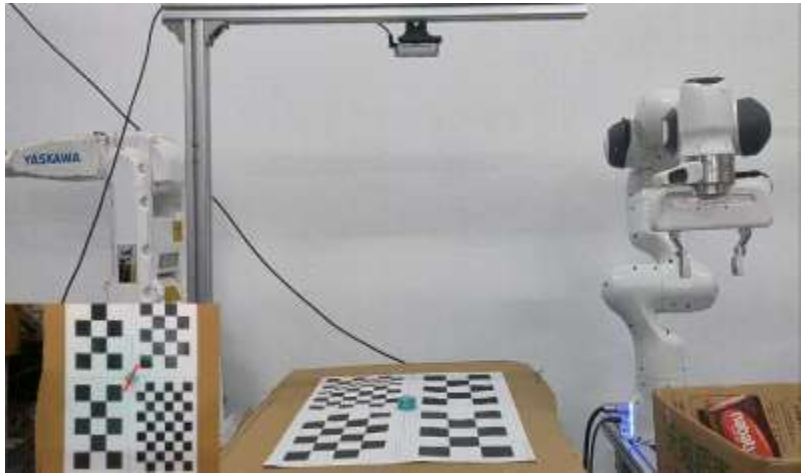
Without Data Adaptation
Successful Rate: 4/10



With Data Adaptation
Successful Rate: 10/10

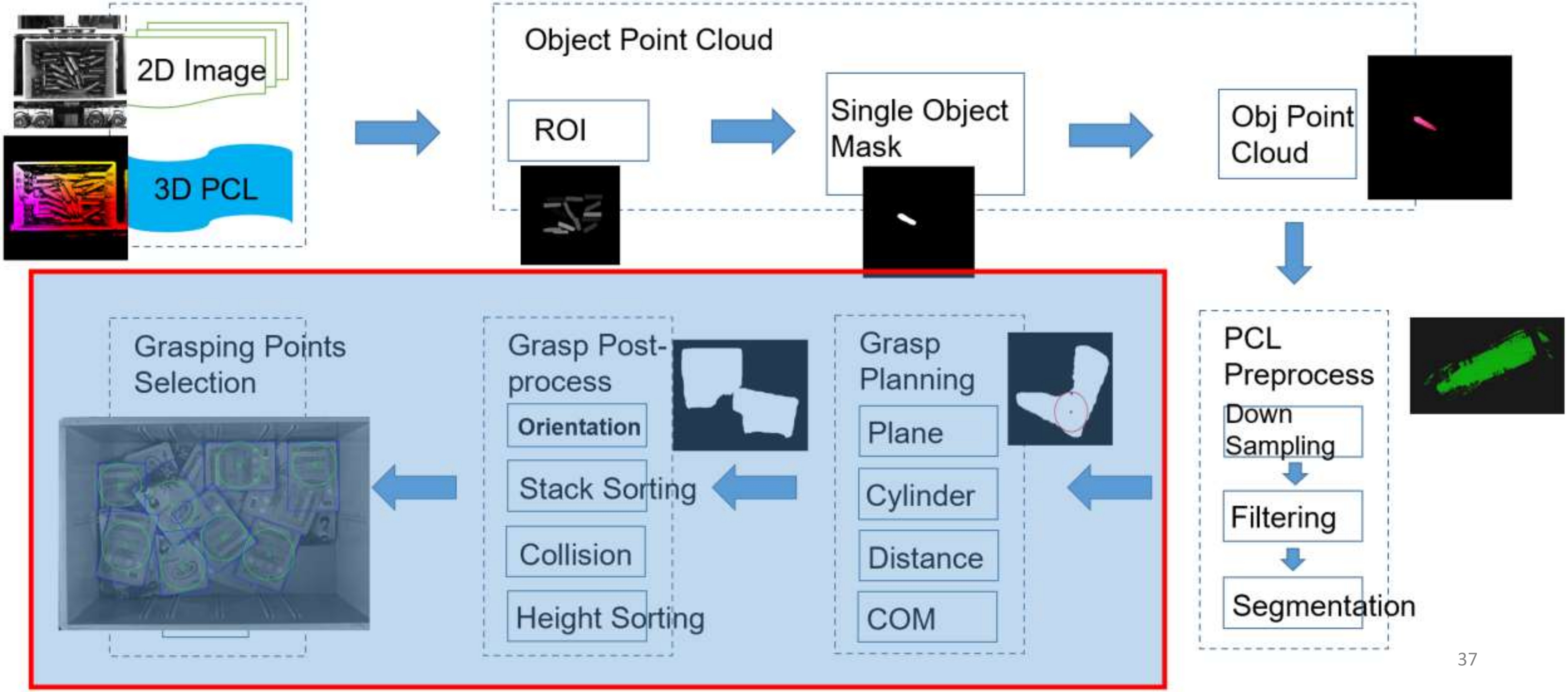


Without Data Adaptation
(strong background noise)
Successful Rate: 0/10



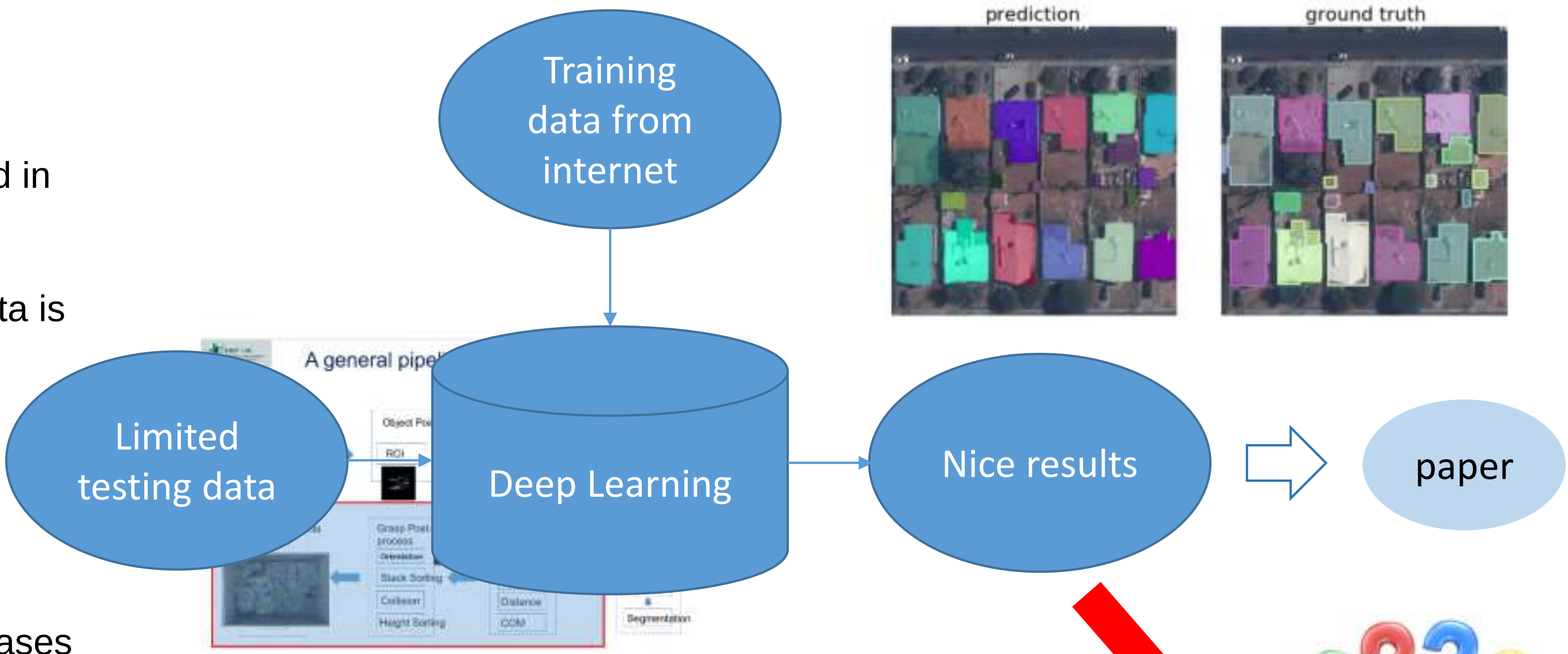
With Data Adaptation
(strong background noise)
Successful Rate: 10/10

A general pipeline of robotic grasping

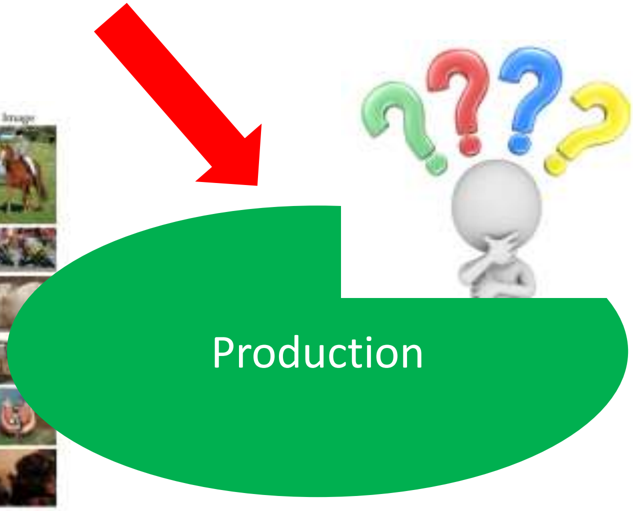


Transfer grasping to real world

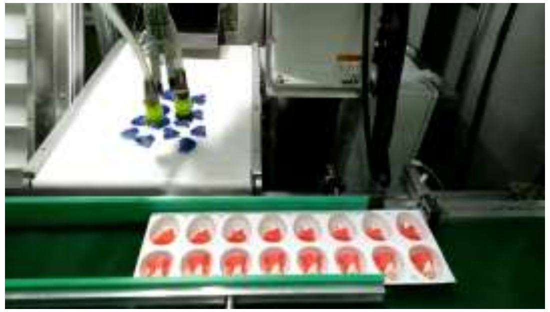
- All the algorithms from DL is probabilistic (deterministic required in production)
- A large number of well-labelled data is required (data is expensive in production)
- 90% success rate could be a nice paper (99% is not enough)
- Ignore the corner cases (Corner cases must be taken into account, e.g self-driving)
- 1ms in prediction time is not perceived (1 ms could make a huge difference)
- Low (zero) stake vs high stake
- ……..

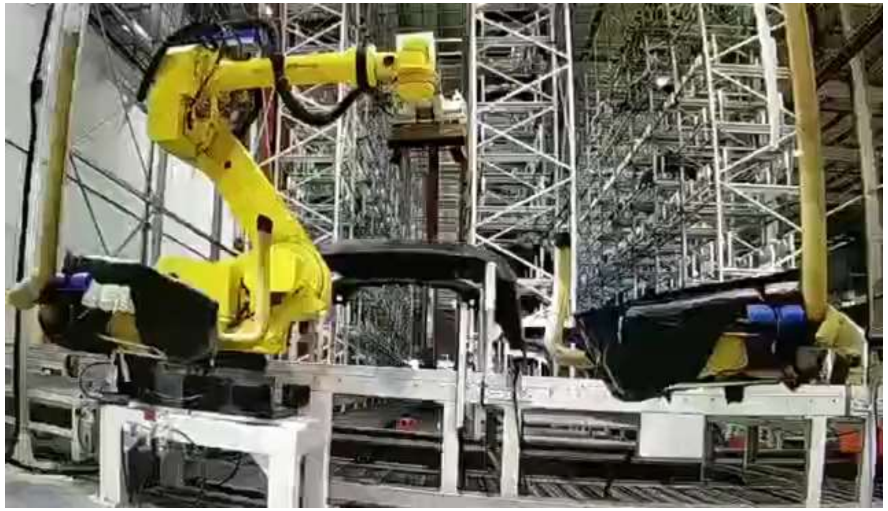


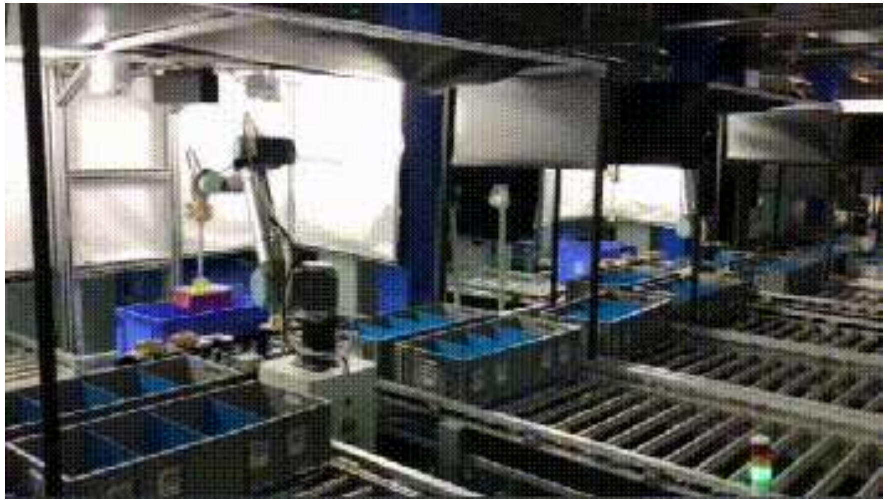
Example of scene segmentation:
Is there any guarantee that we can segment the image correctly all the time?



Grasping in the real world

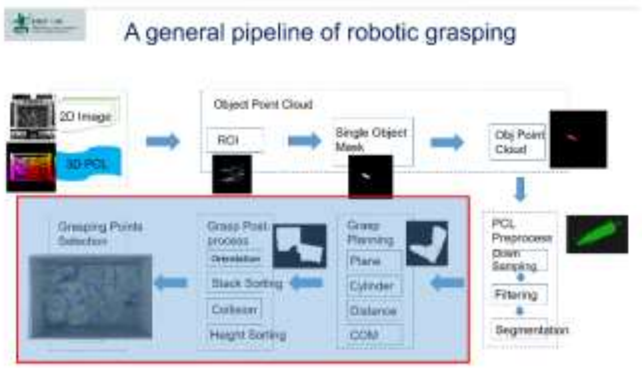






Transfer grasping to real world

- ✓ Data is expensive
- ✓ 99.99% is still far away! (4s per grasp)
- ✓ 1ms is important
- ✓ System is important



1. Robustness – Grasping is solved without this constraint
2. Speed – What makes grasping useful in real life
3. Adaptability – What makes grasping intelligent

Thanks for your attention!

Email: miao.li@whu.edu.cn