# Form2Fit: Learning Shape Priors for Generalizable Assembly from Disassembly

Kevin Zakka, Andy Zeng, Johnny Lee, Shuran Song

Google Research

Stanford University

Columbia University

# Form2Fit: Learning Shape Priors for Generalizable Assembly from Disassembly

Kevin Zakka, Andy Zeng, Johnny Lee, Shuran Song

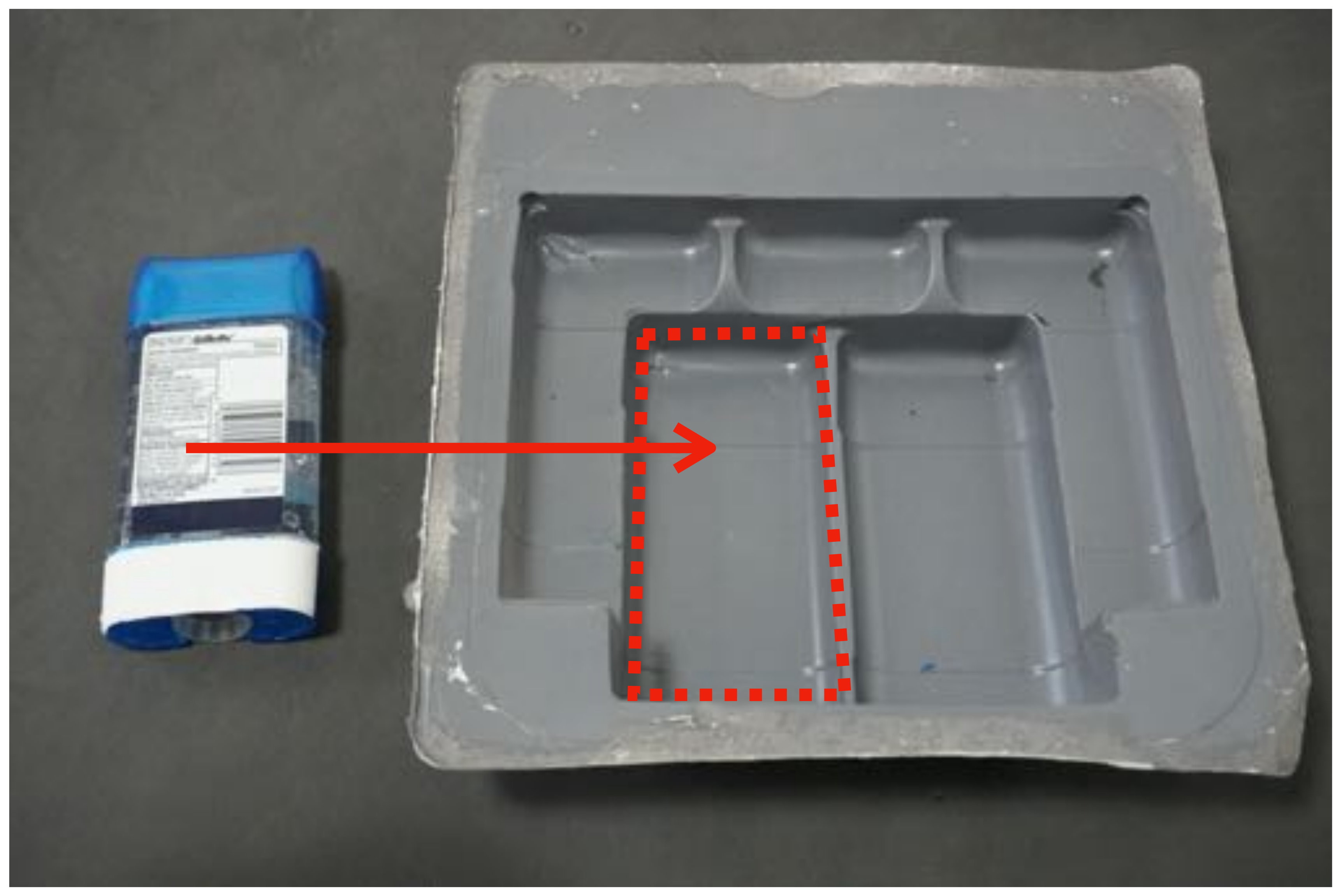
*includes video narration*

Google Research

Stanford University

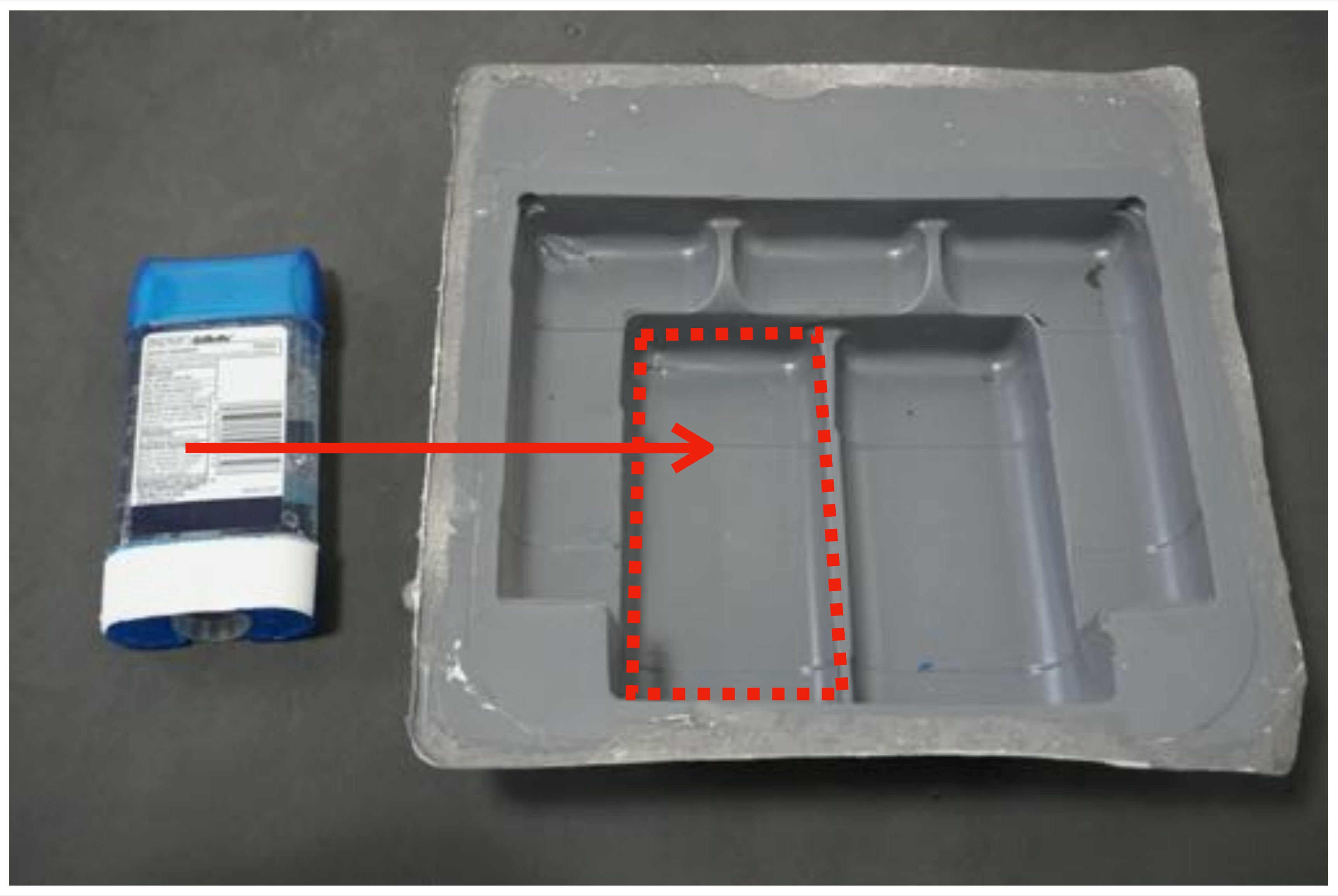
Columbia University

# Shape Matching



kit assembly

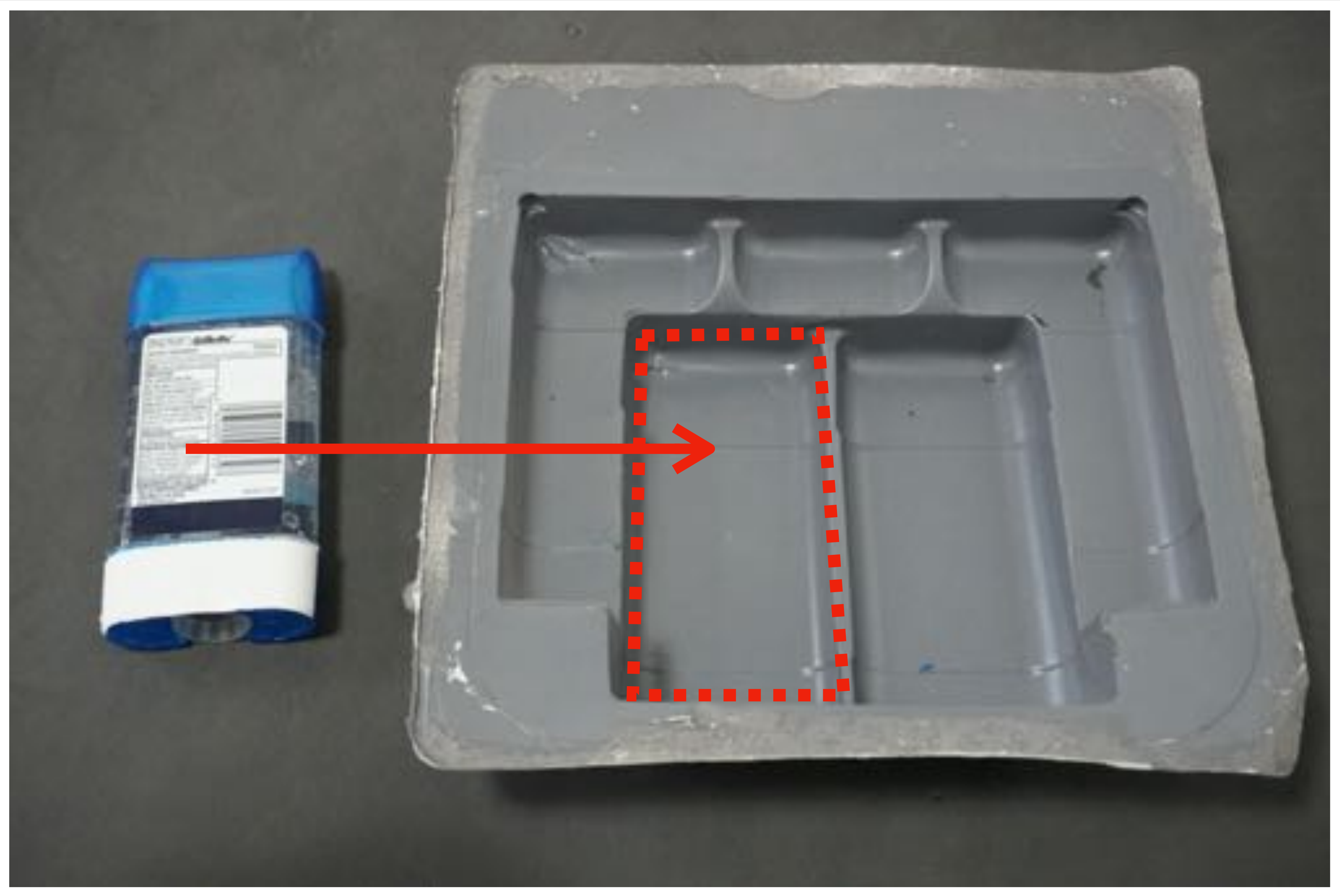
# Shape Matching



kit assembly

everyday interactions

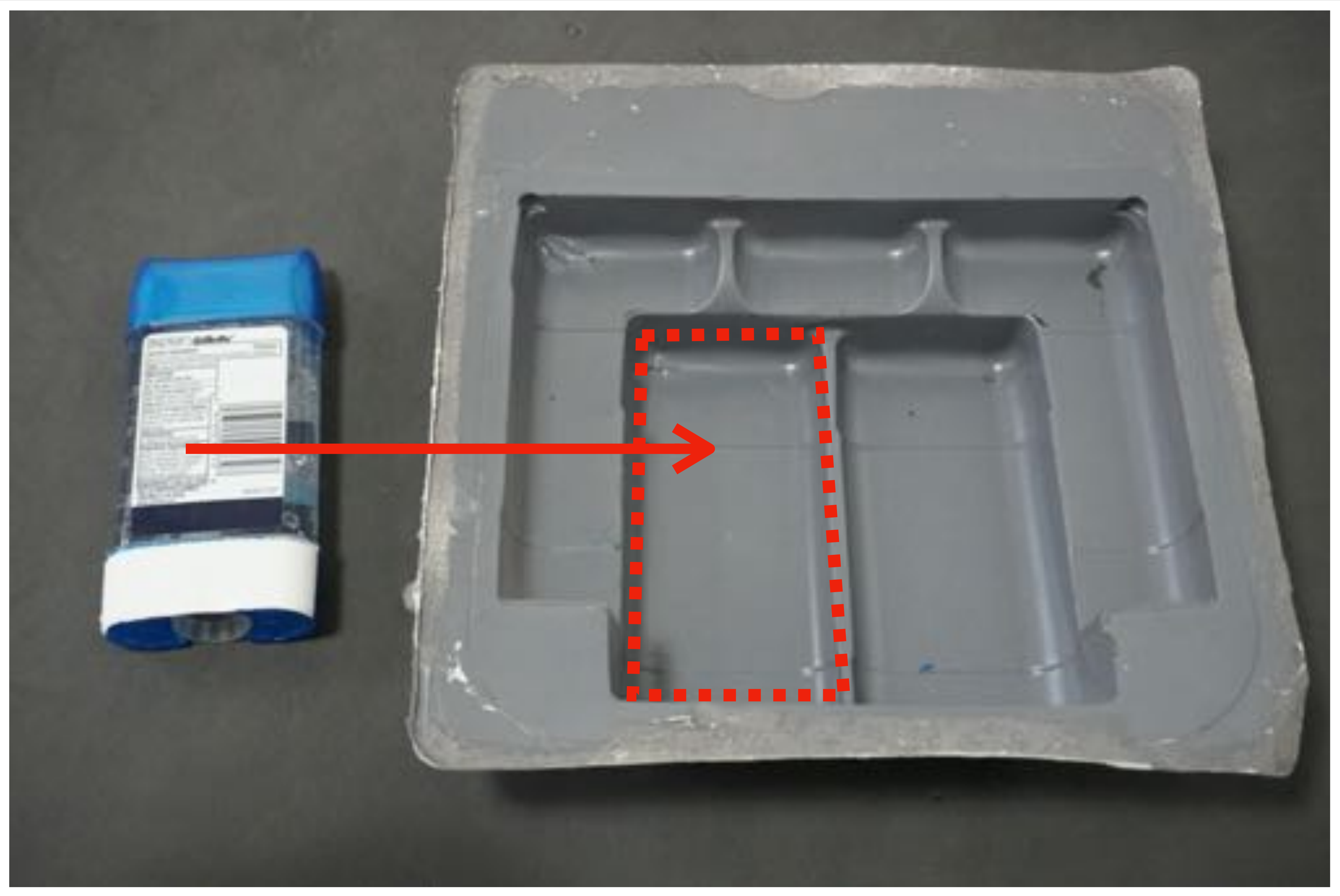
# Shape Matching



kit assembly

everyday interactions

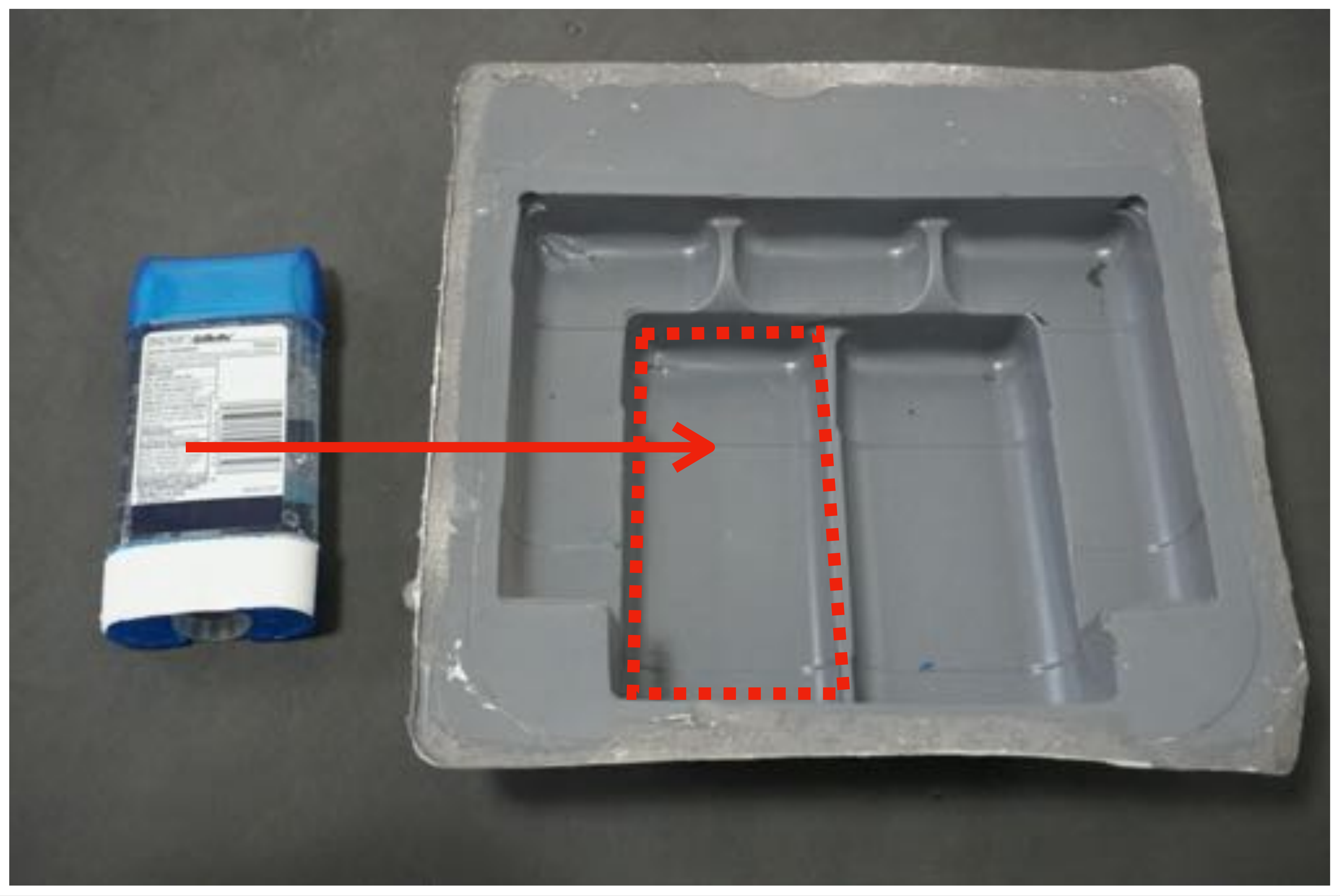
# Shape Matching



kit assembly

everyday interactions

# Shape Matching

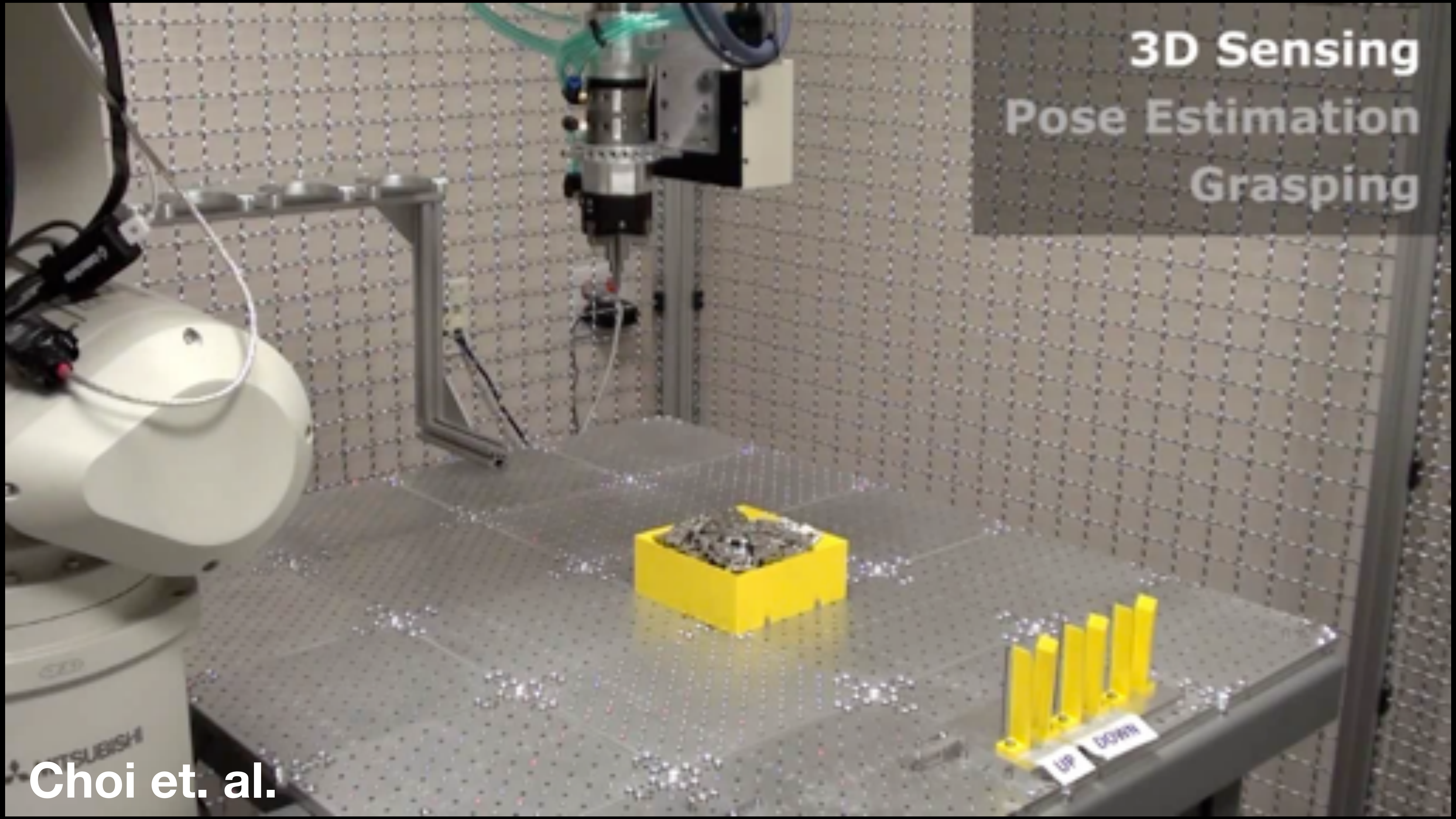


kit assembly

everyday interactions

# Towards Flexible Assembly

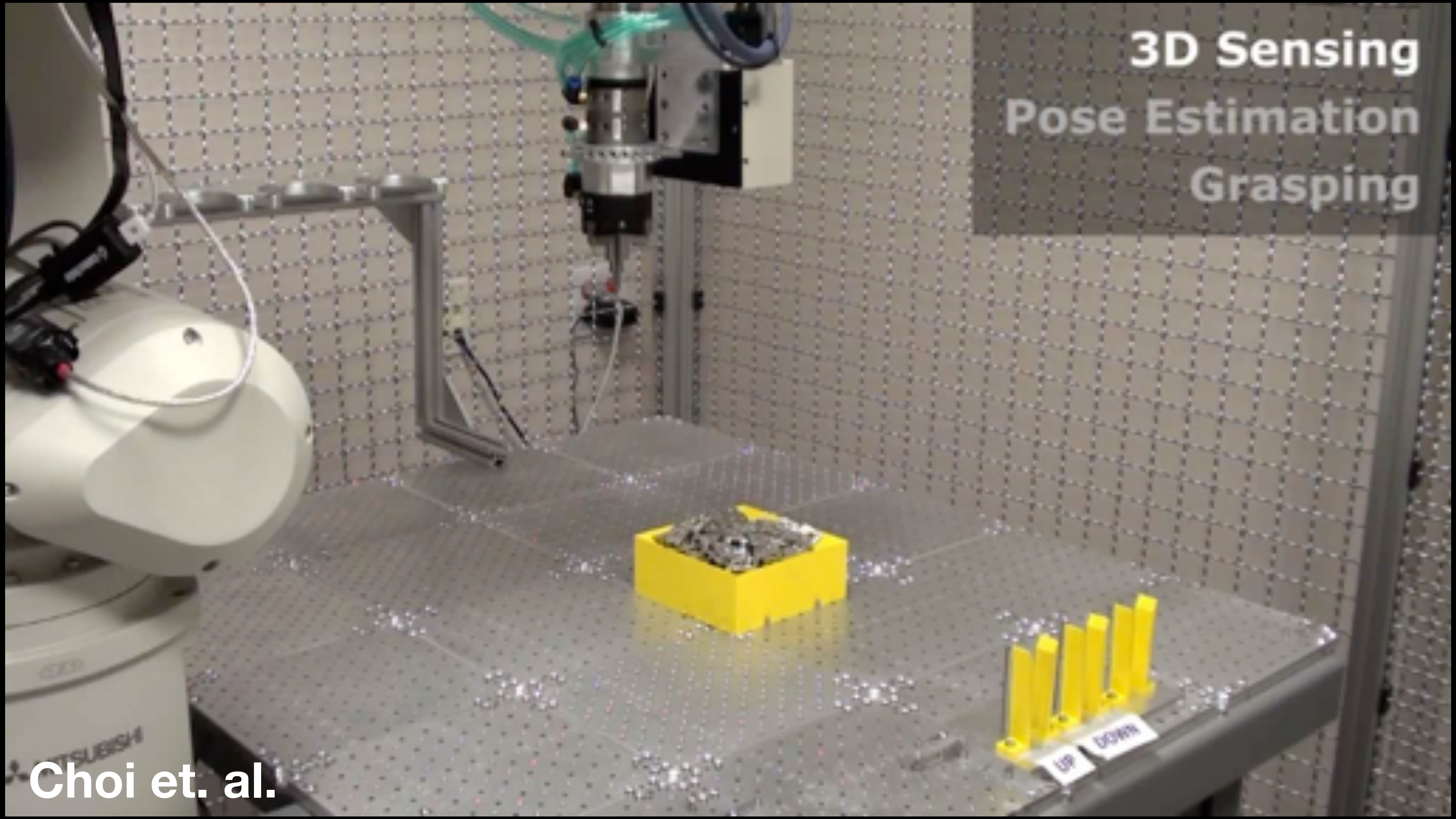
state-of-the-art
robo-kitting
solution




CAD Model

# Towards Flexible Assembly

state-of-the-art
robo-kitting
solution



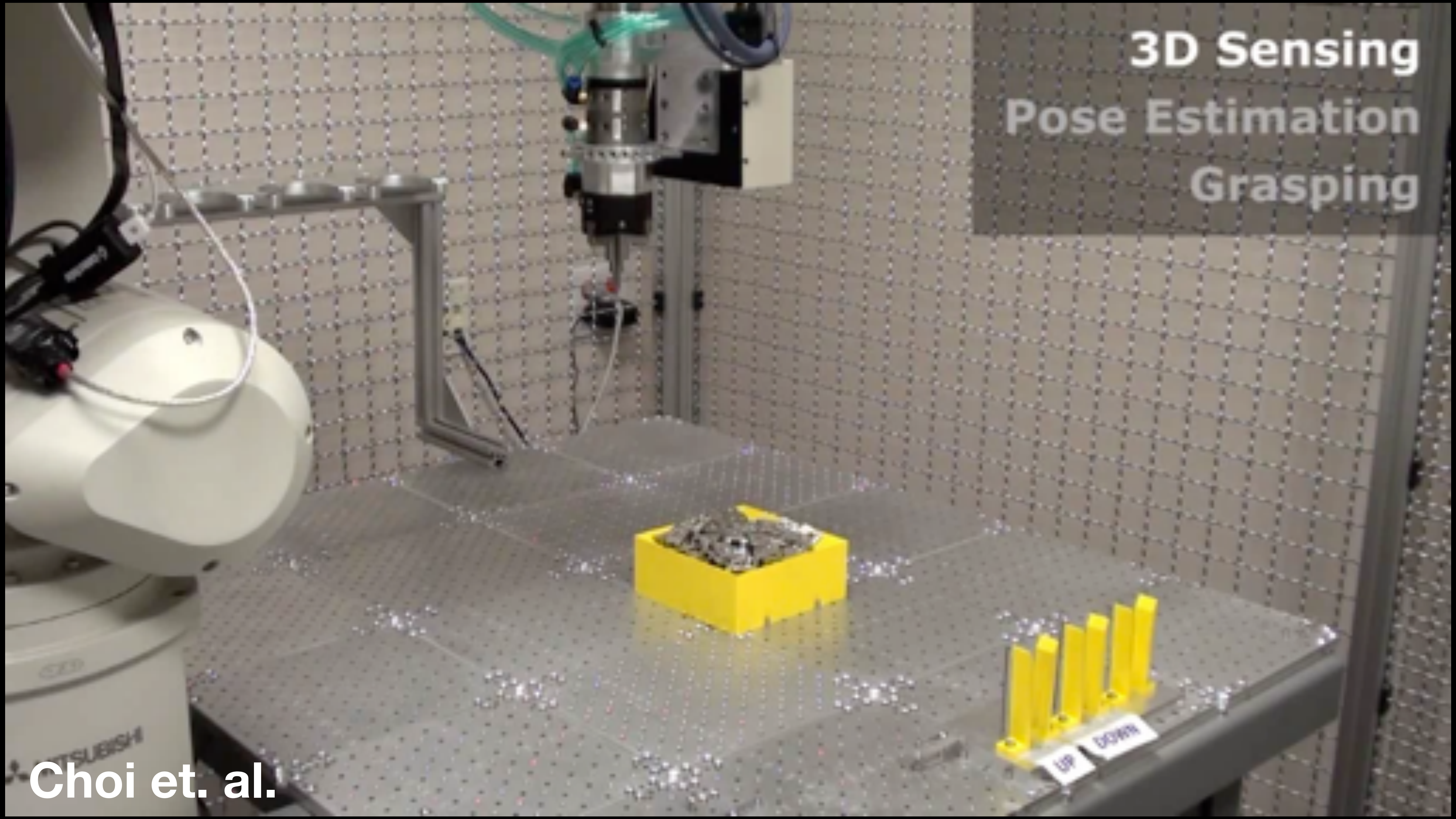

CAD Model

# Towards Flexible Assembly

state-of-the-art

robo-kitting

solution




CAD Model

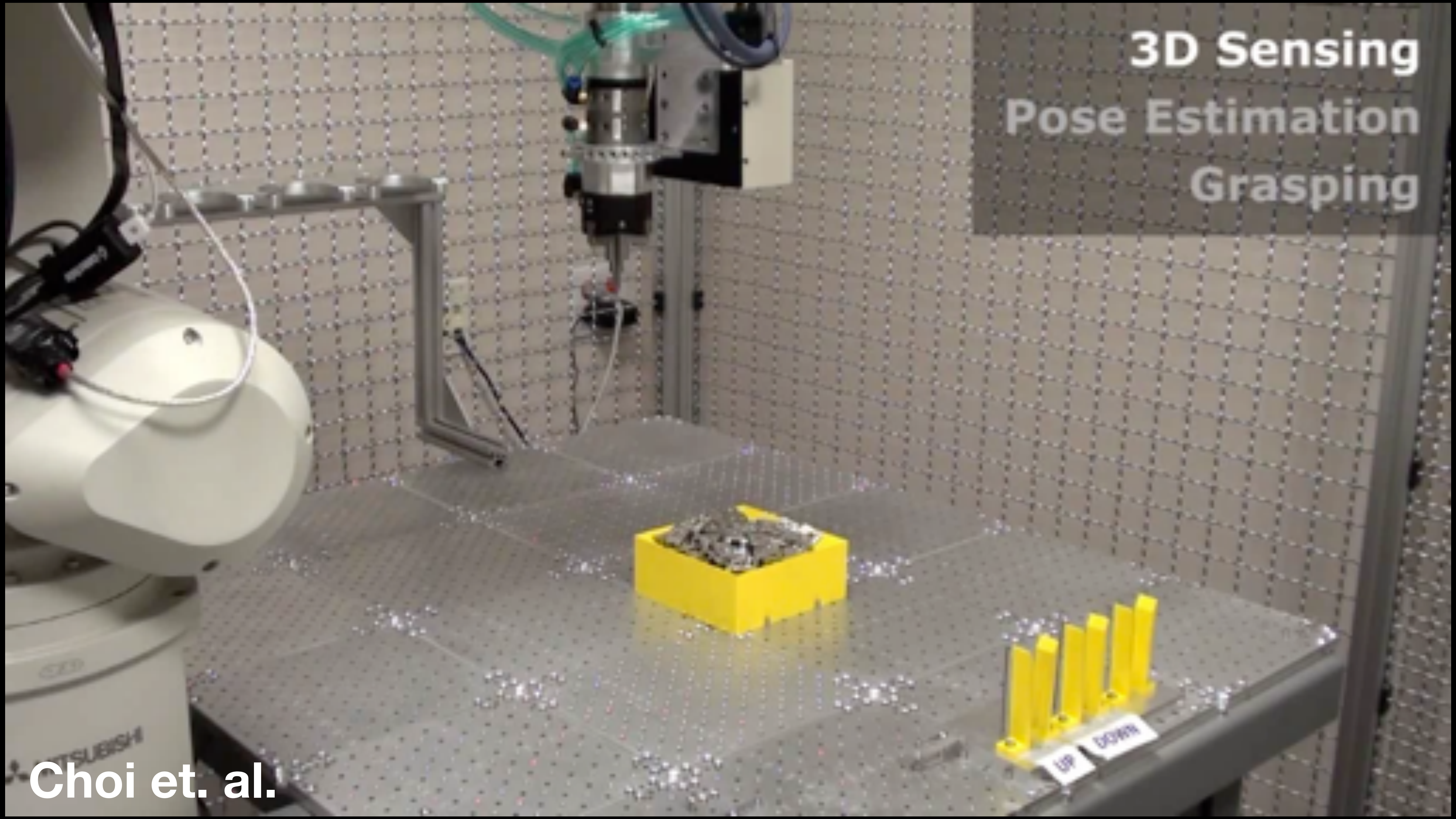
- require prior knowledge and manual engineering

# Towards Flexible Assembly

state-of-the-art

robo-kitting

solution

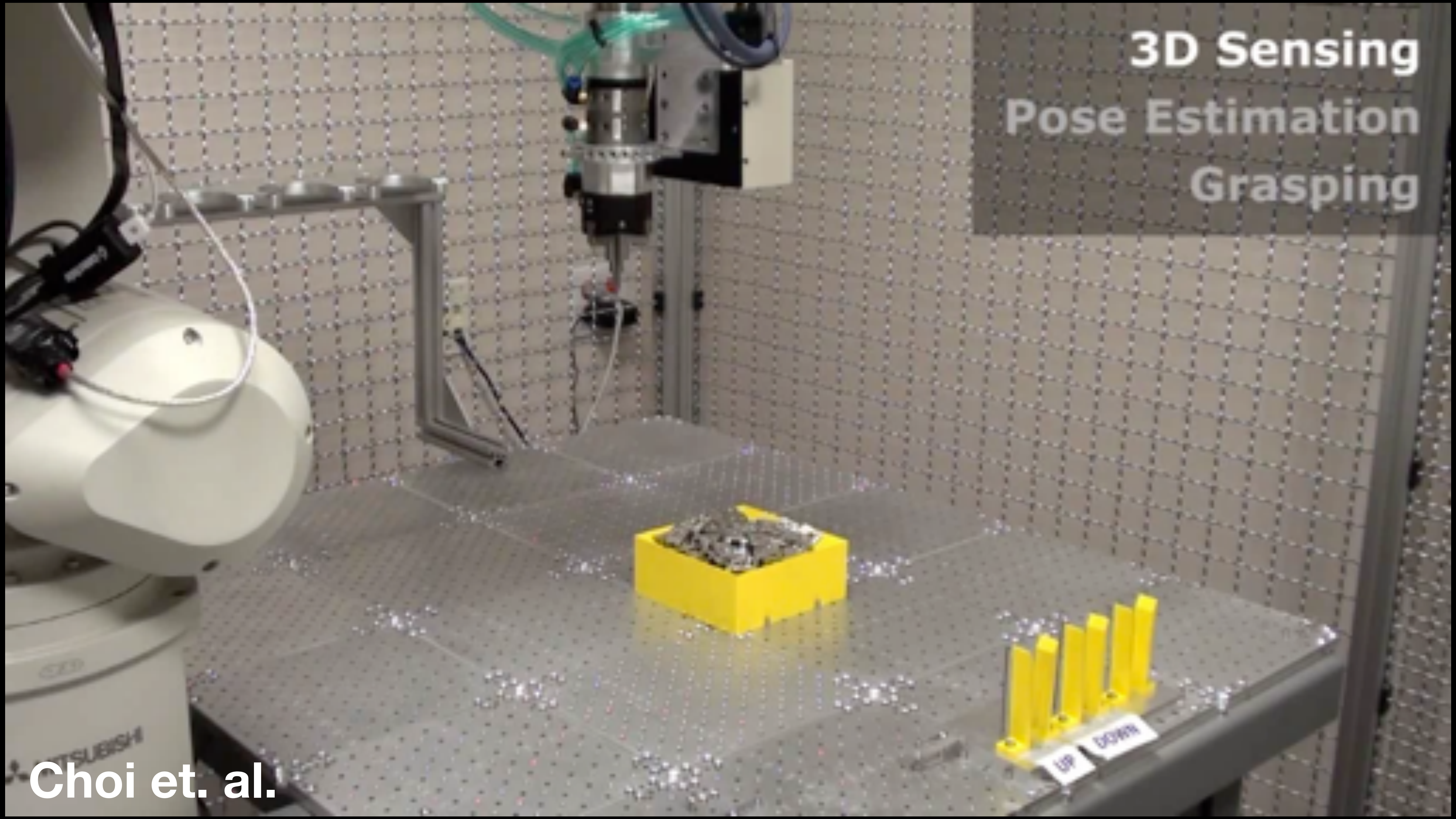



CAD Model

- require prior knowledge and manual engineering
- cannot quickly adapt to new objects and settings

# Towards Flexible Assembly

state-of-the-art
robo-kitting
solution

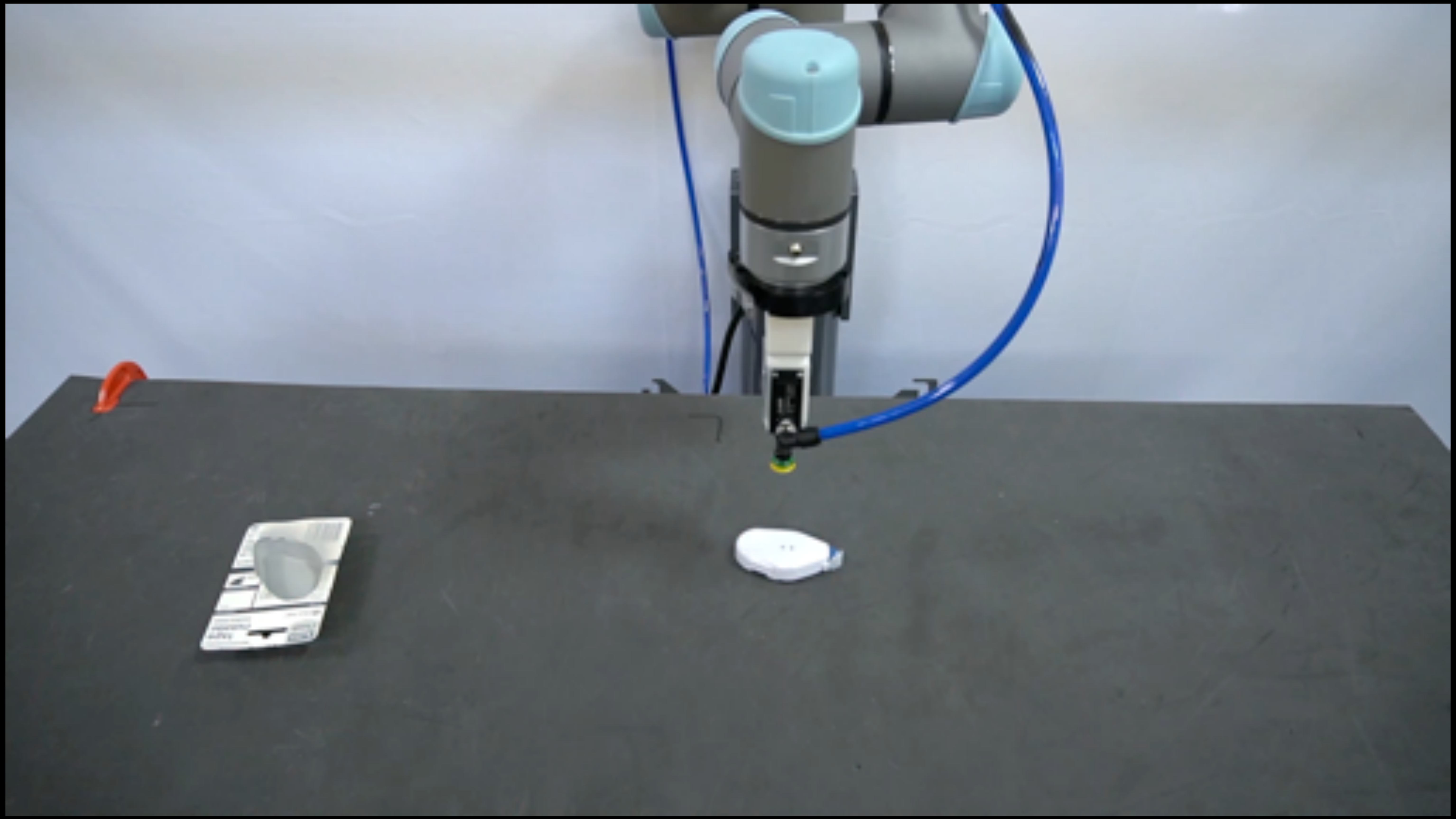



CAD Model

- require prior knowledge and manual engineering
- cannot quickly adapt to new objects and settings

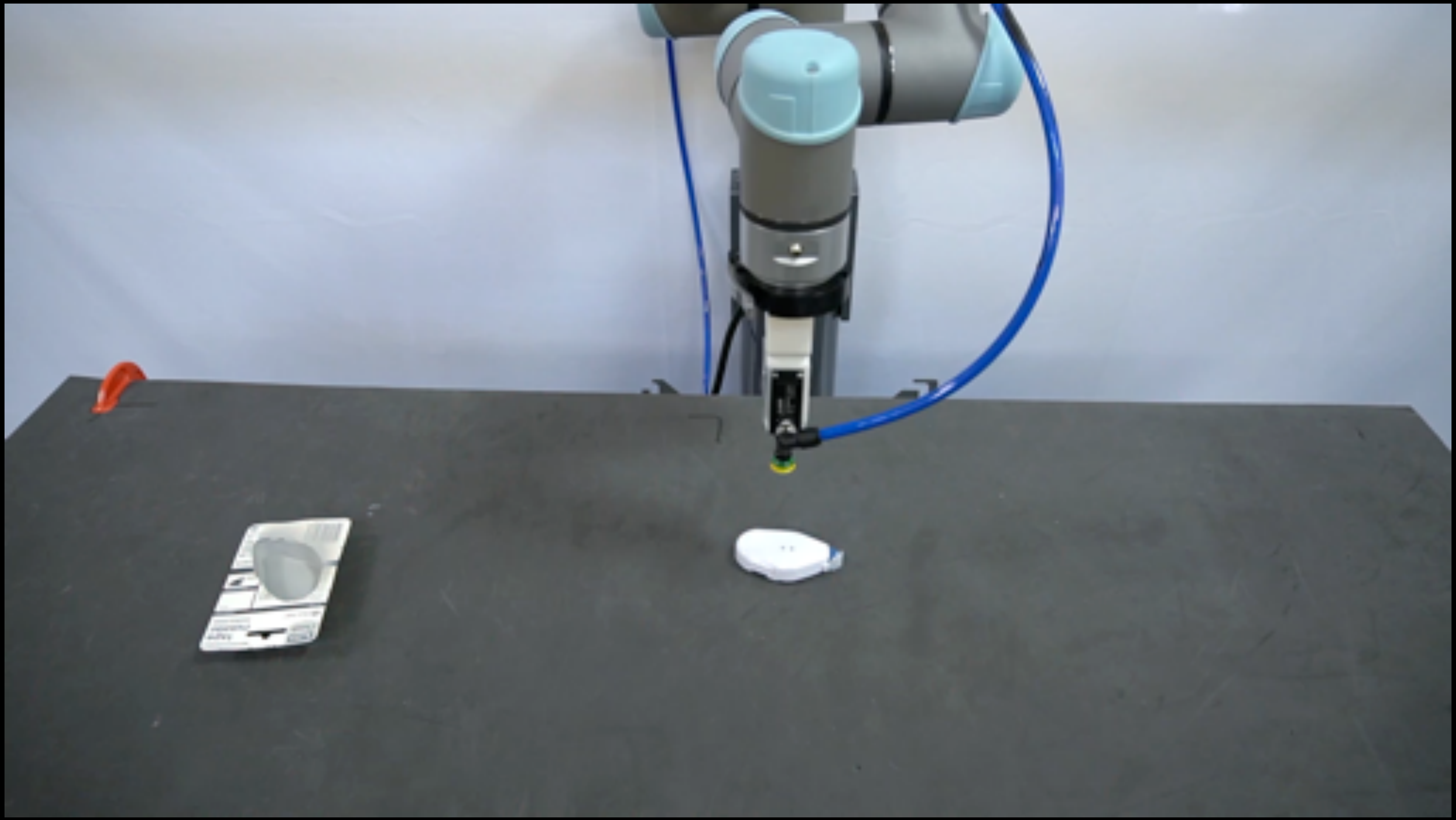
can we endow them with generalization abilities?

# Generalizable Assembly



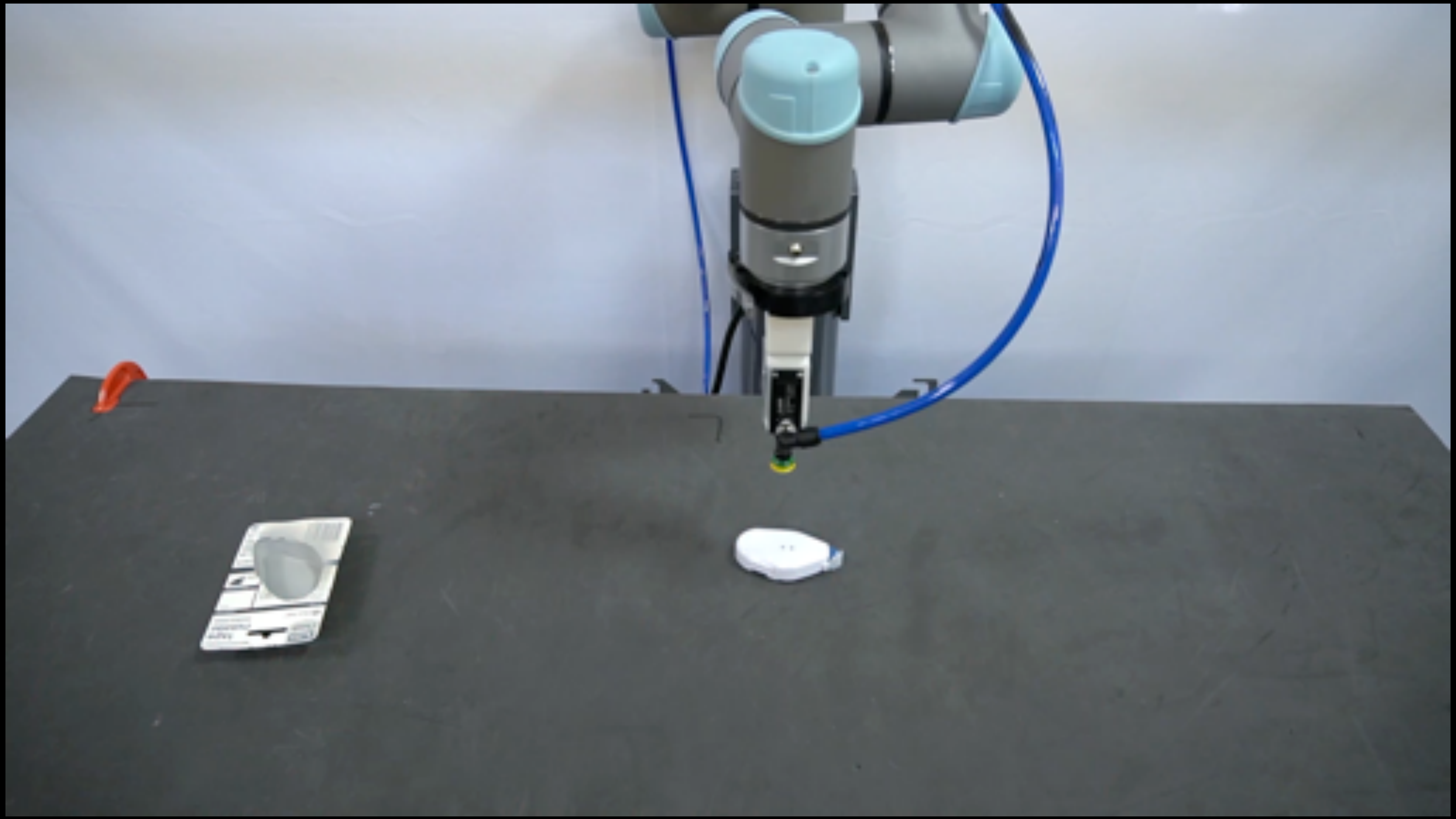
# Generalizable Assembly

**through Shape Matching & Self-Supervision**



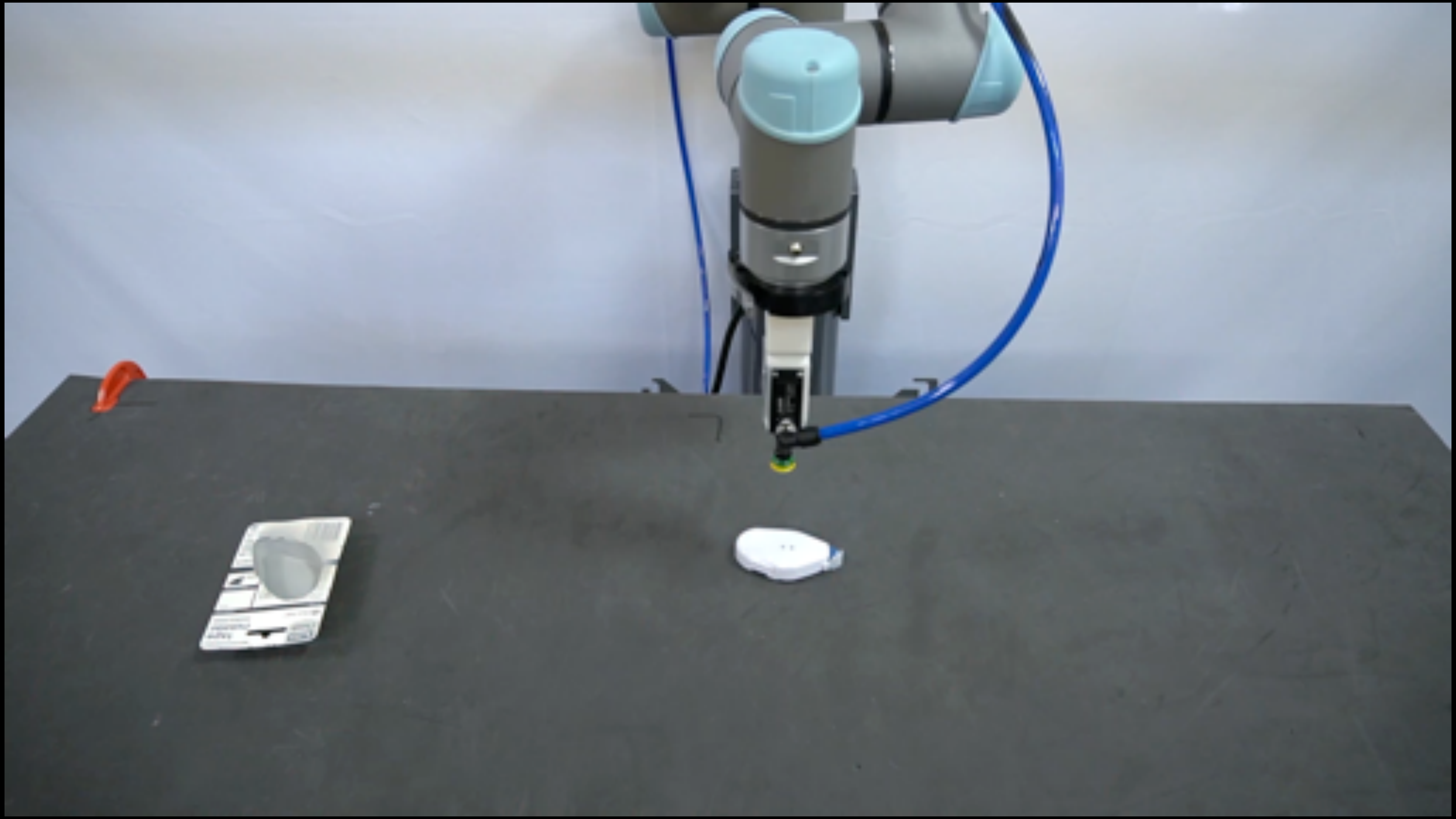
# Generalizable Assembly

**through Shape Matching & Self-Supervision**



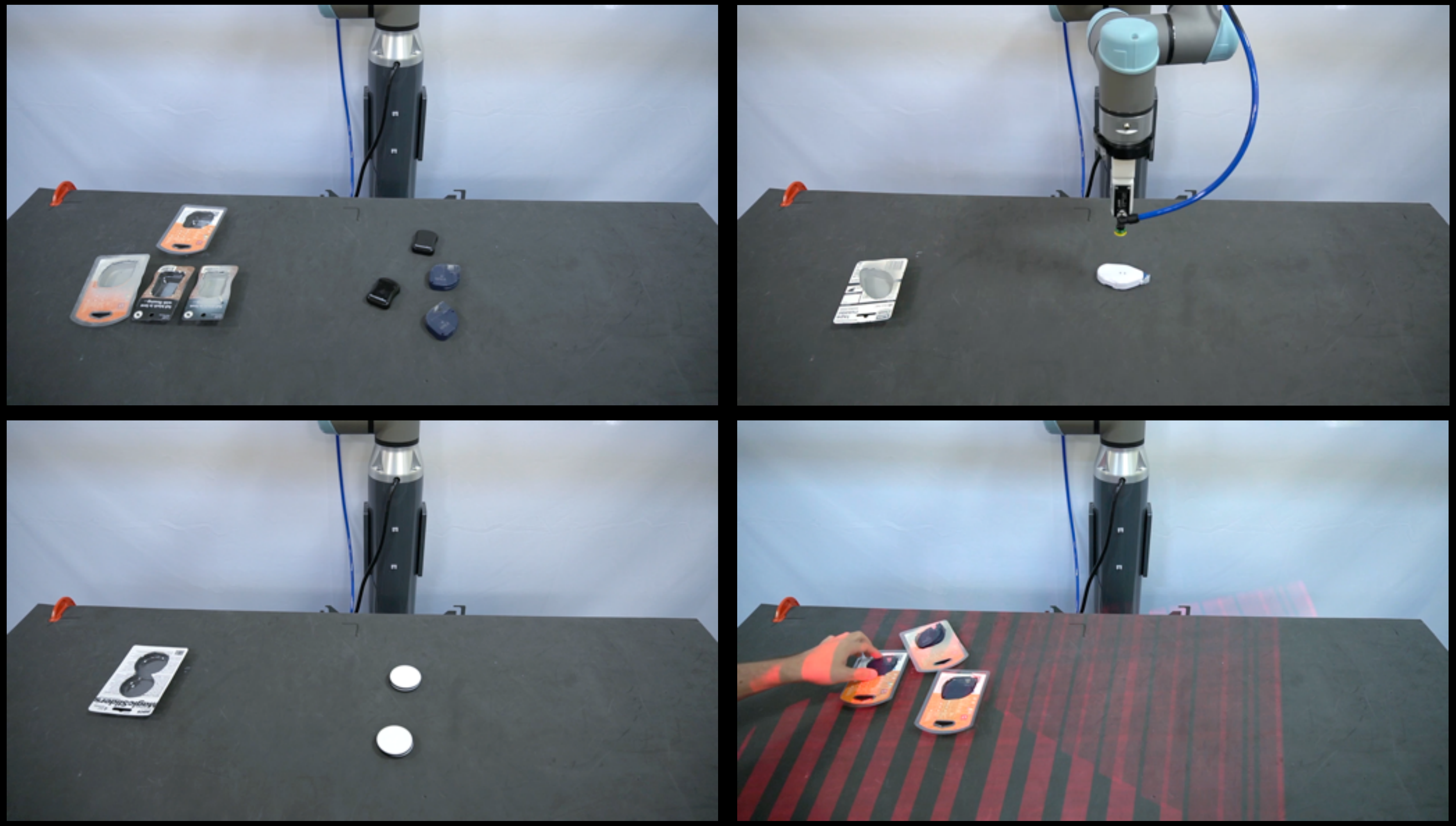
# Generalizable Assembly

**through Shape Matching & Self-Supervision**



# Generalizable Assembly

**through Shape Matching & Self-Supervision**

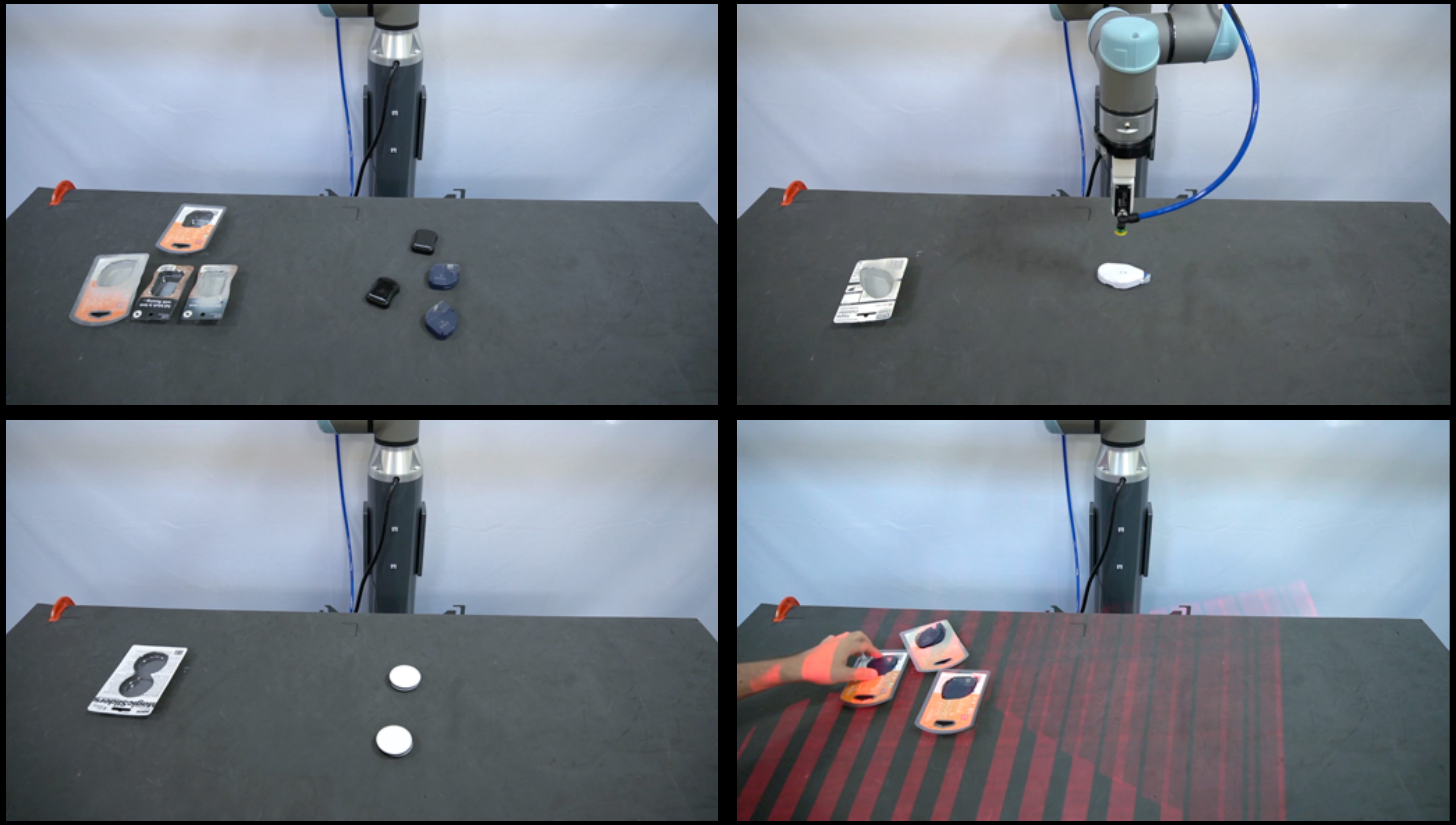


never before seen

# Generalizable Assembly

through **Shape Matching** & **Self-Supervision**

**Form2Fit**



never before seen

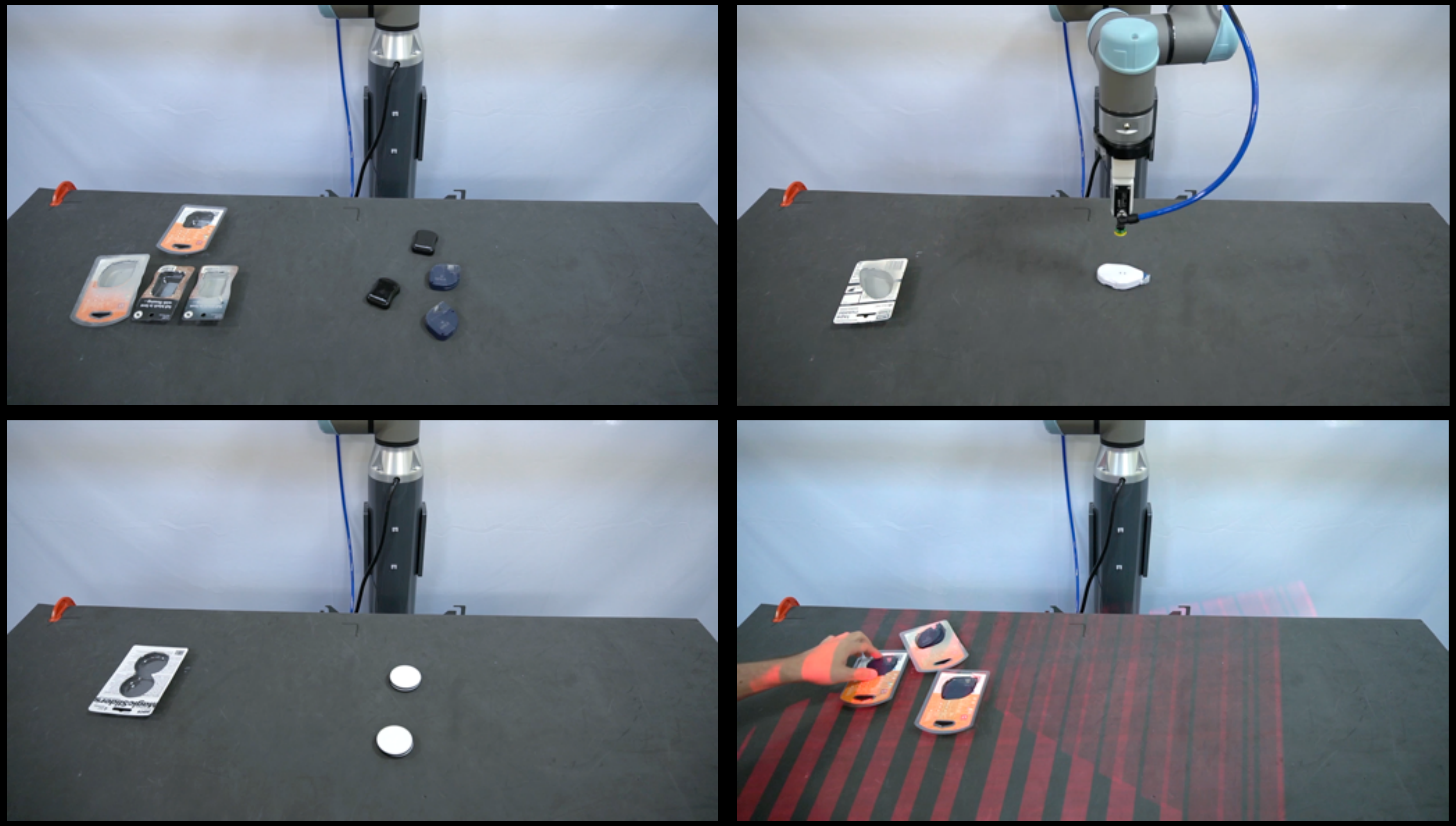
# Generalizable Assembly

through Shape Matching & Self-Supervision

**Form2Fit**

94% novel configurations

86% novel objects & kits



never before seen

# Generalizable Assembly

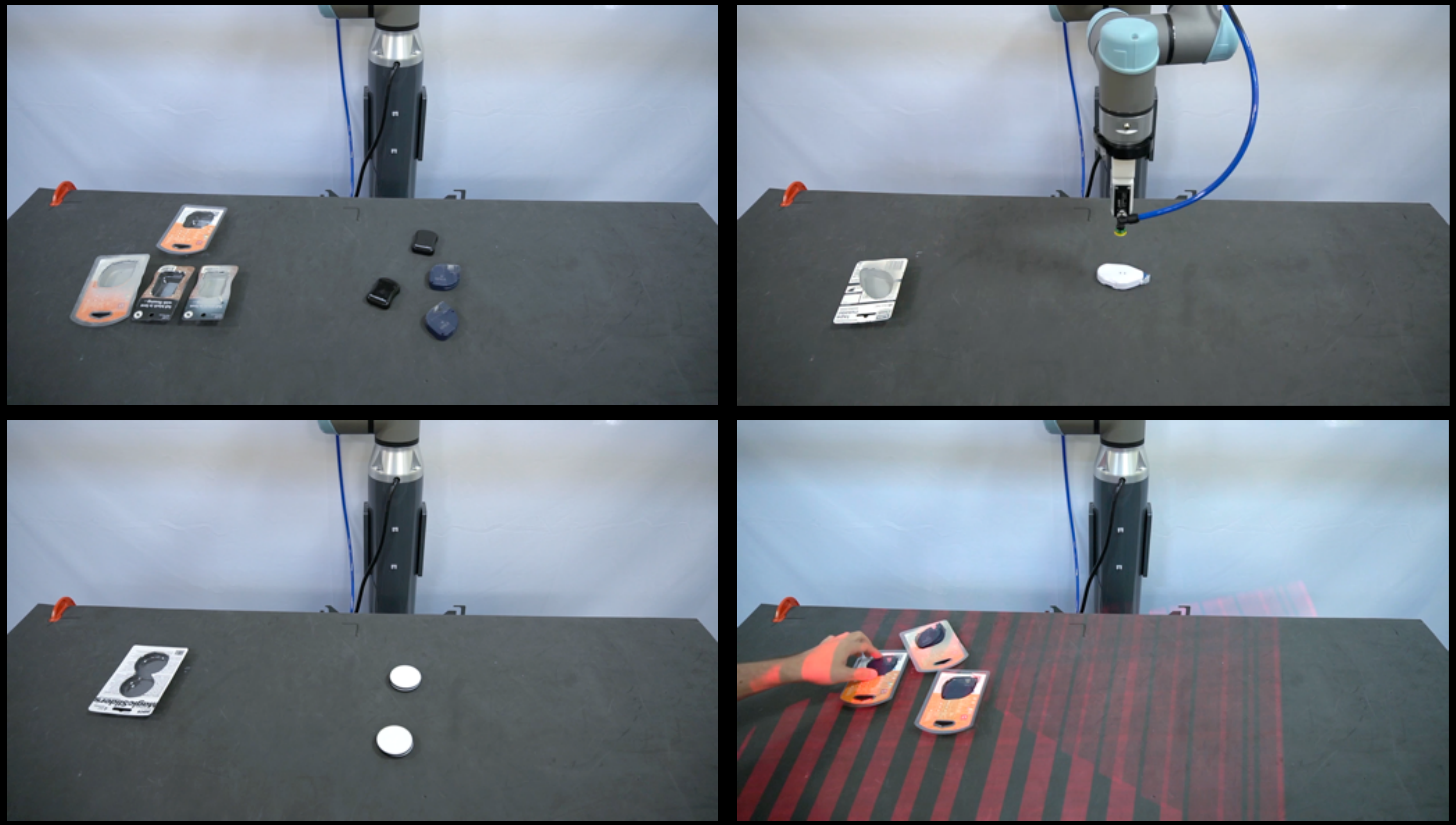
## through Shape Matching & Self-Supervision

**Form2Fit**

94% novel configurations

86% novel objects & kits

~12 hours training



never before seen

# Key Ideas

Kit Assembly → Shape Matching

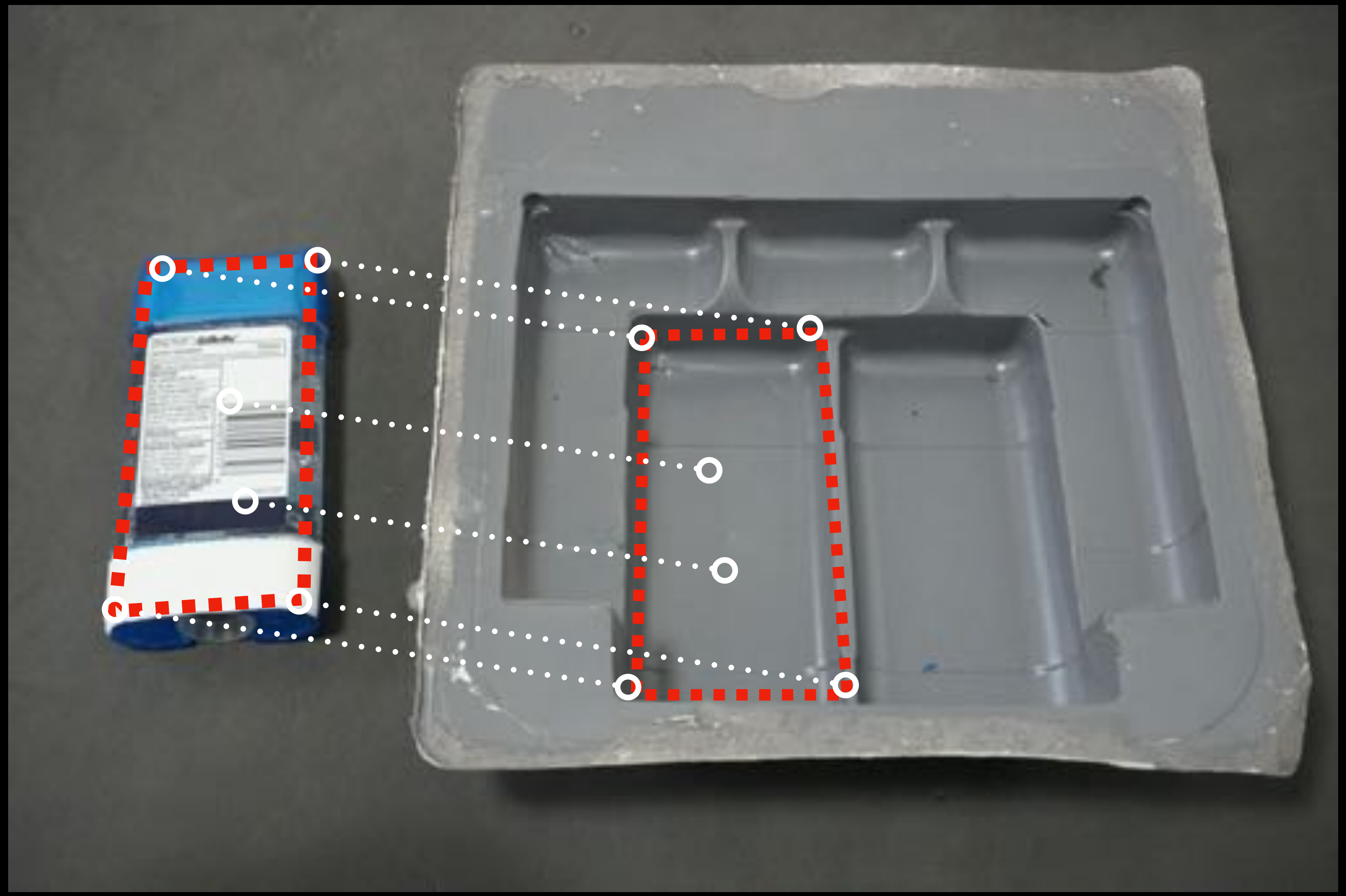
# Key Ideas

Kit Assembly → Shape Matching

- learns geometric shape descriptors

# Key Ideas

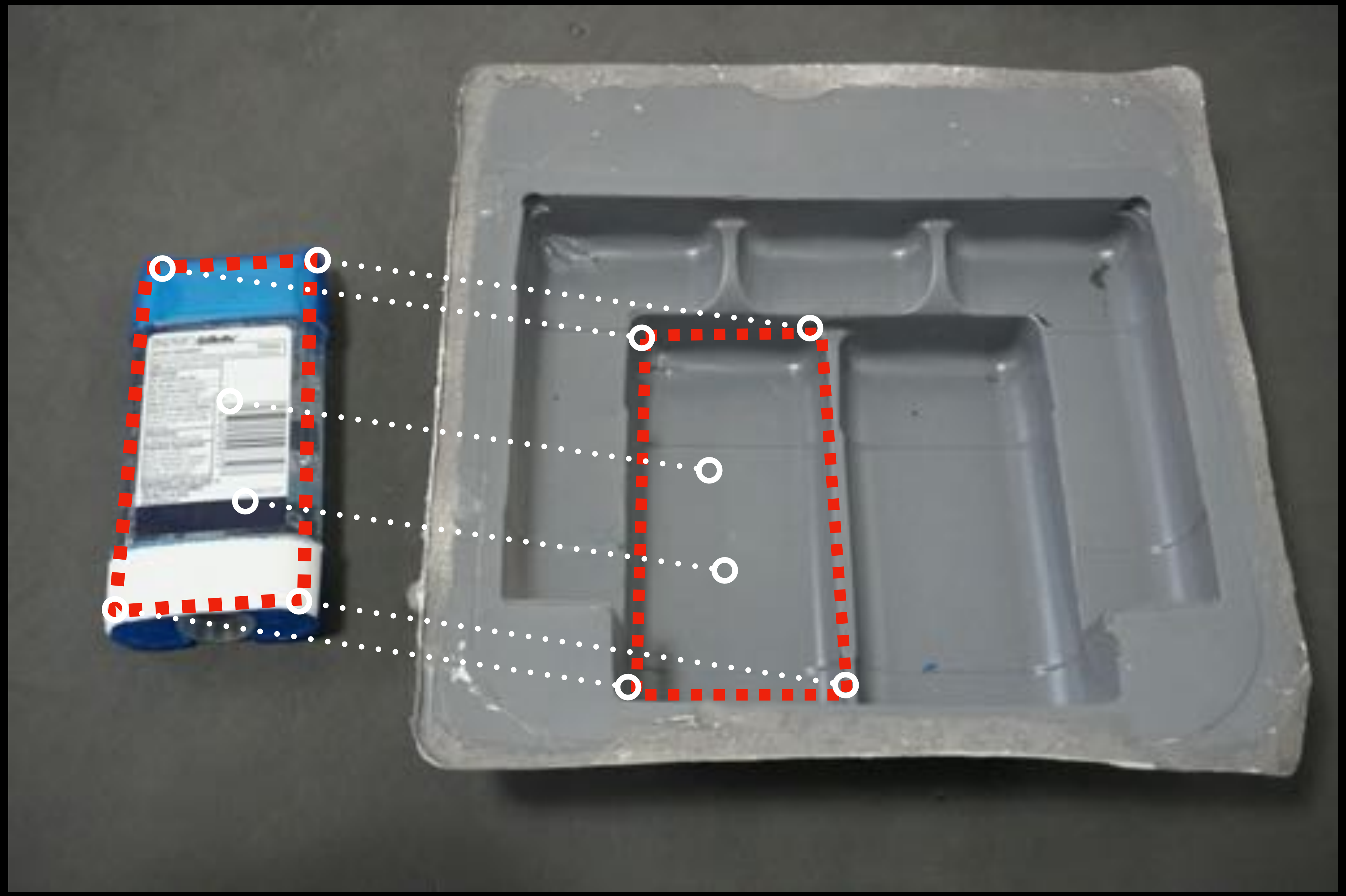
Kit Assembly → Shape Matching



- learns geometric shape descriptors

# Key Ideas

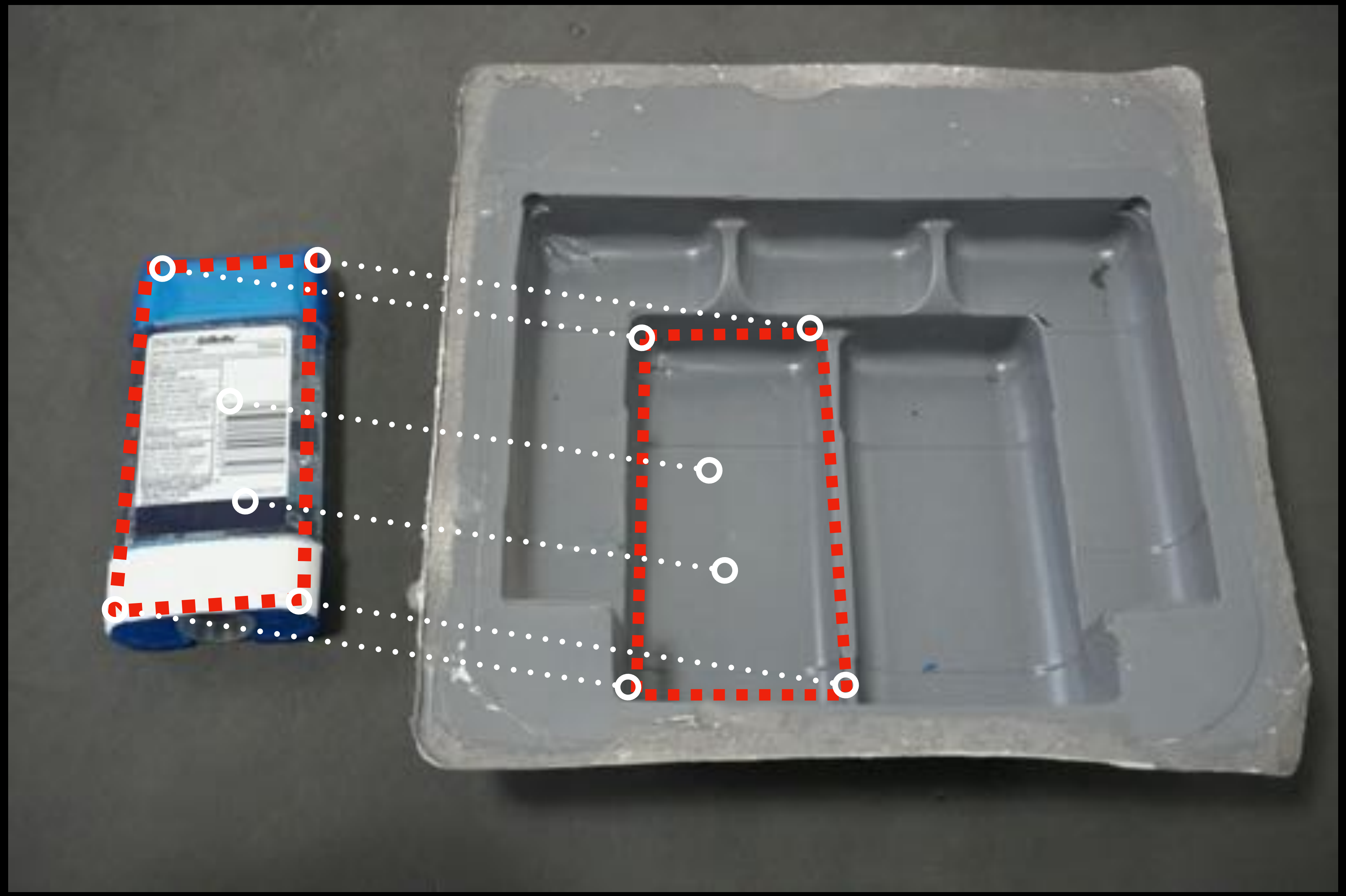
Kit Assembly → Shape Matching



- learns geometric shape descriptors
- generalizes to new shapes

# Key Ideas

Kit Assembly → Shape Matching

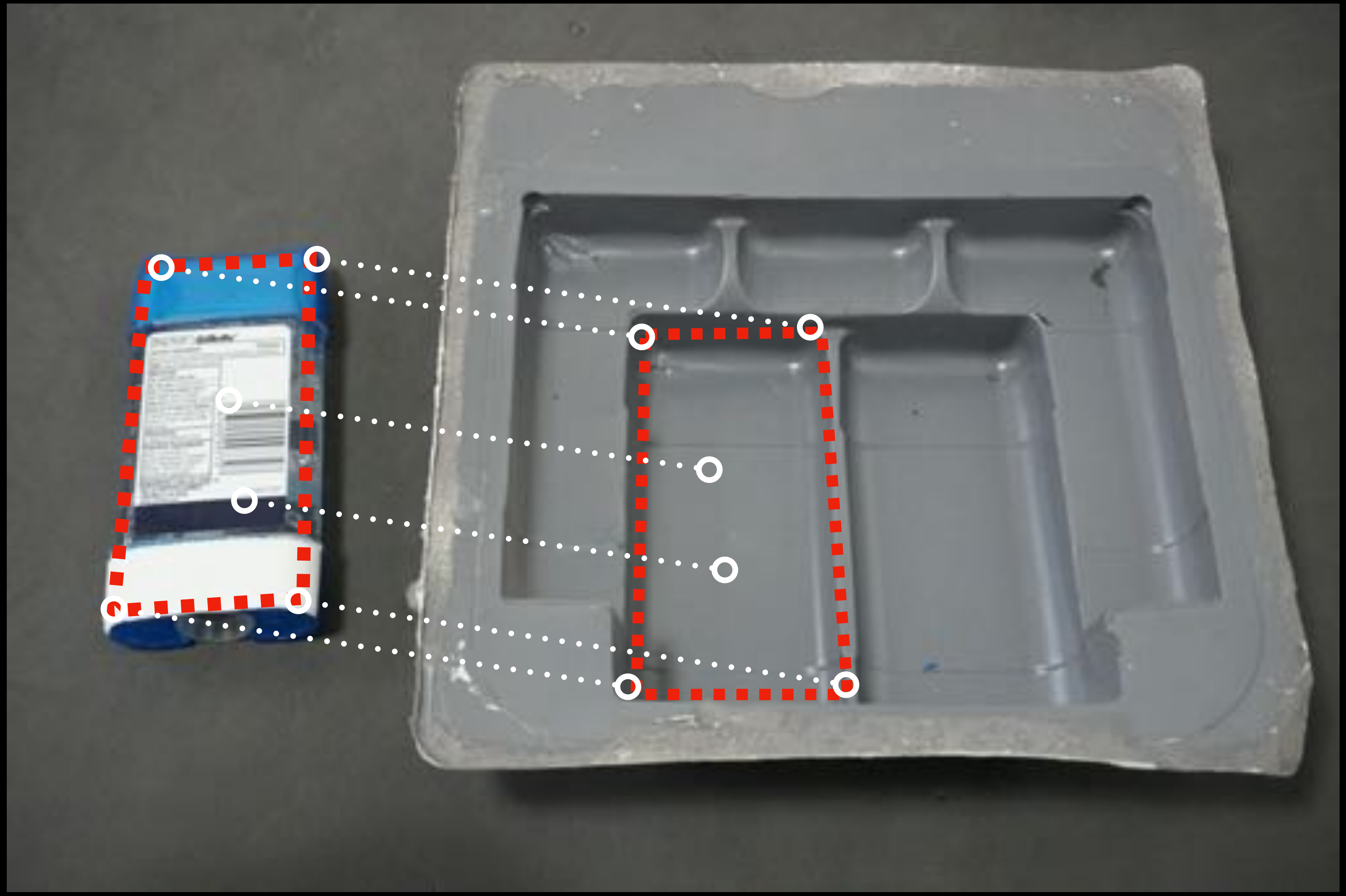


- learns geometric shape descriptors
- generalizes to new shapes

Assembly from Disassembly

# Key Ideas

### Kit Assembly → Shape Matching



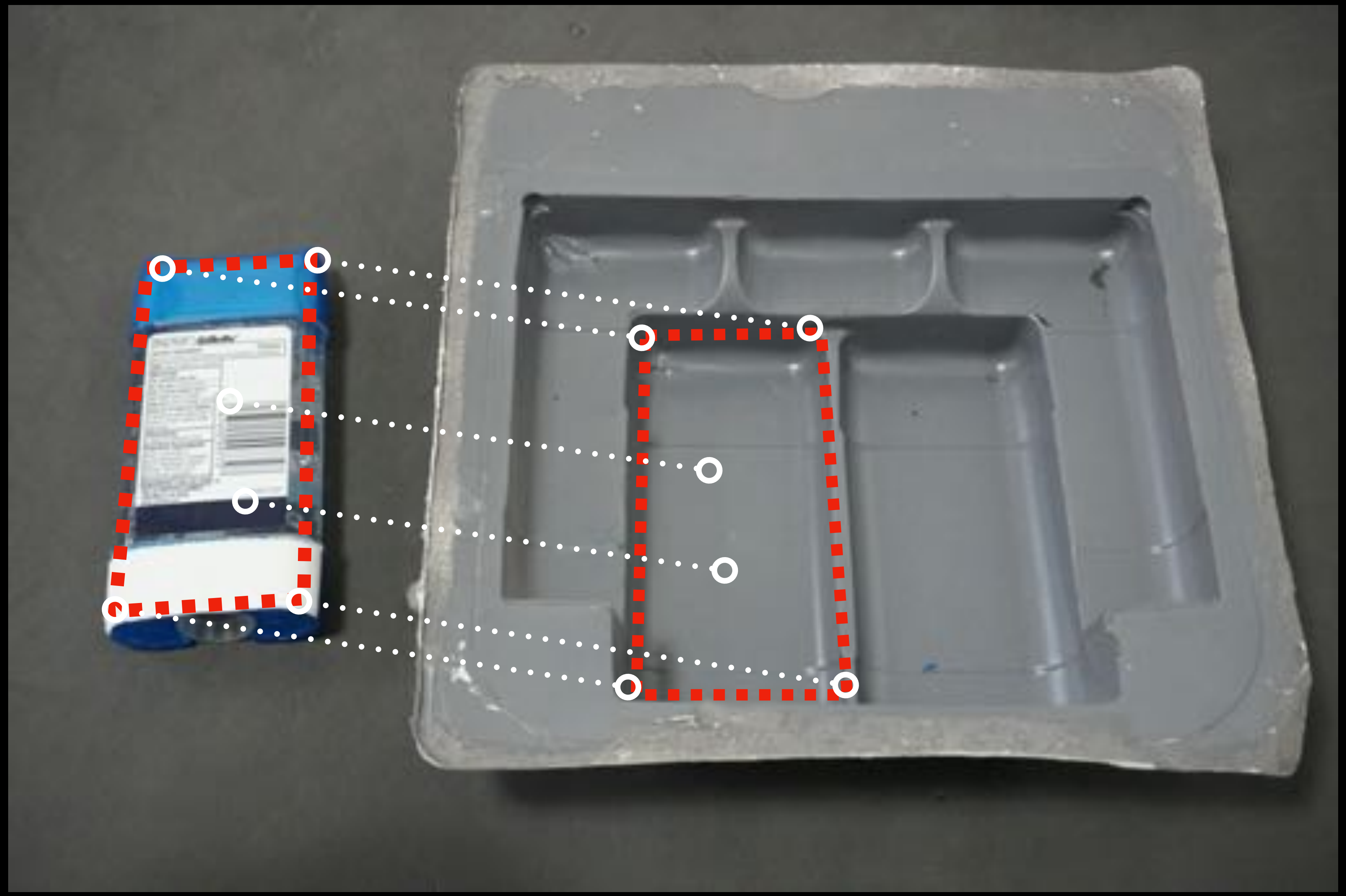
- learns geometric shape descriptors
- generalizes to new shapes

### Assembly from Disassembly


- fully self-supervised

# Key Ideas

Kit Assembly → Shape Matching



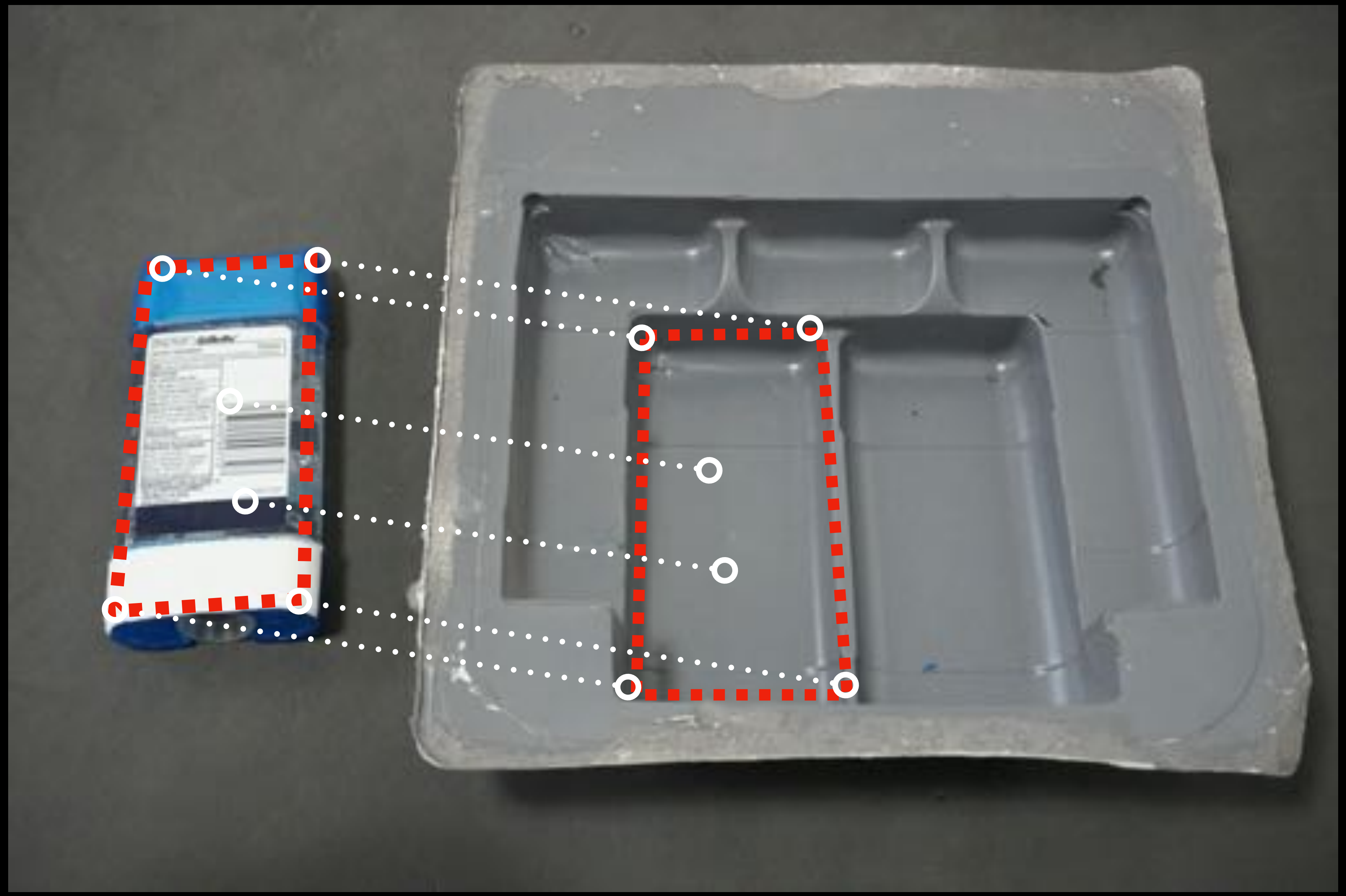
- learns geometric shape descriptors
- generalizes to new shapes

Assembly from Disassembly


- fully self-supervised
- trial and error

# Key Ideas

Kit Assembly → Shape Matching



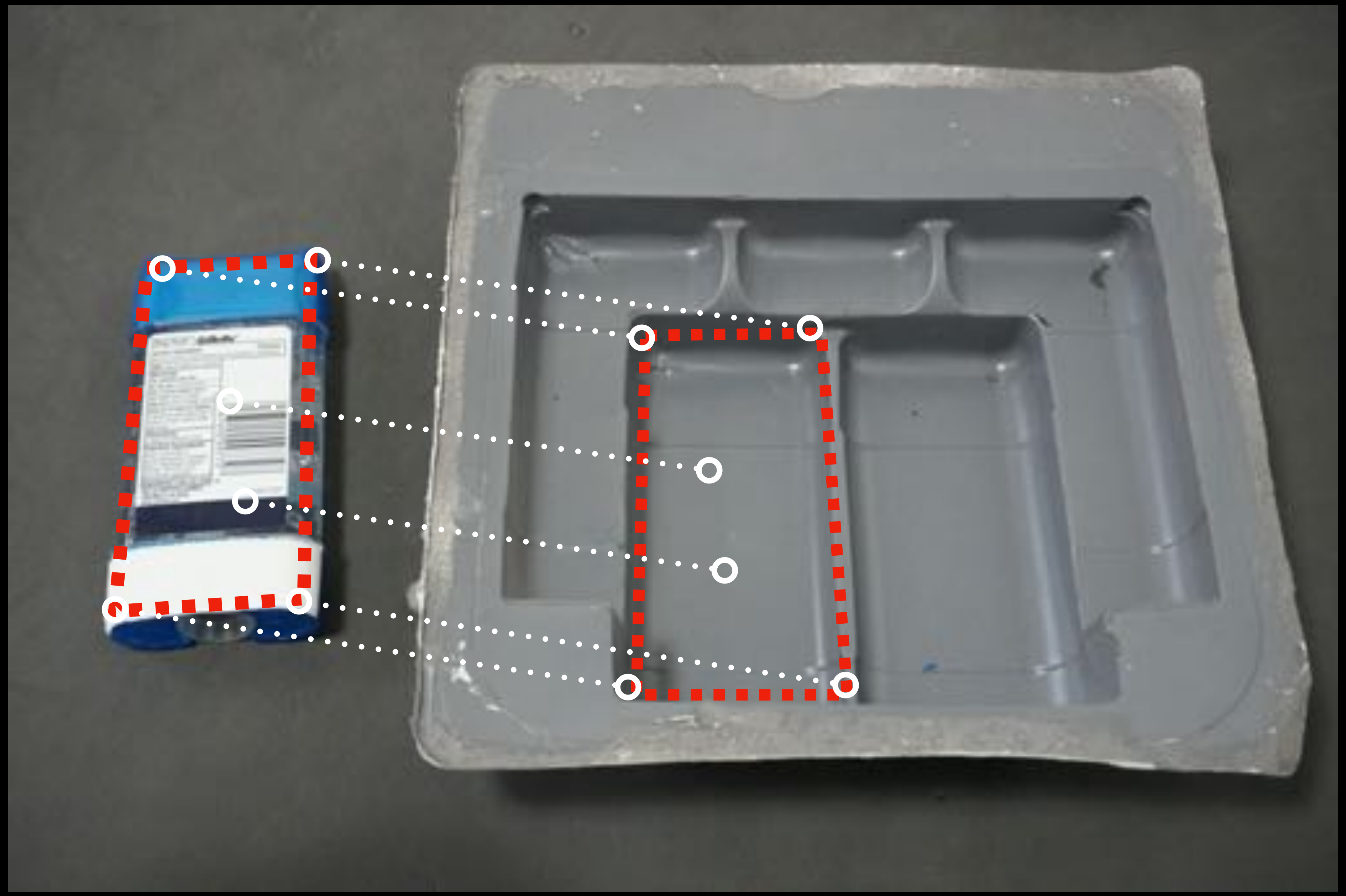
- learns geometric shape descriptors
- generalizes to new shapes

Assembly from Disassembly


- fully self-supervised
- trial and error

# Key Ideas

Kit Assembly → Shape Matching



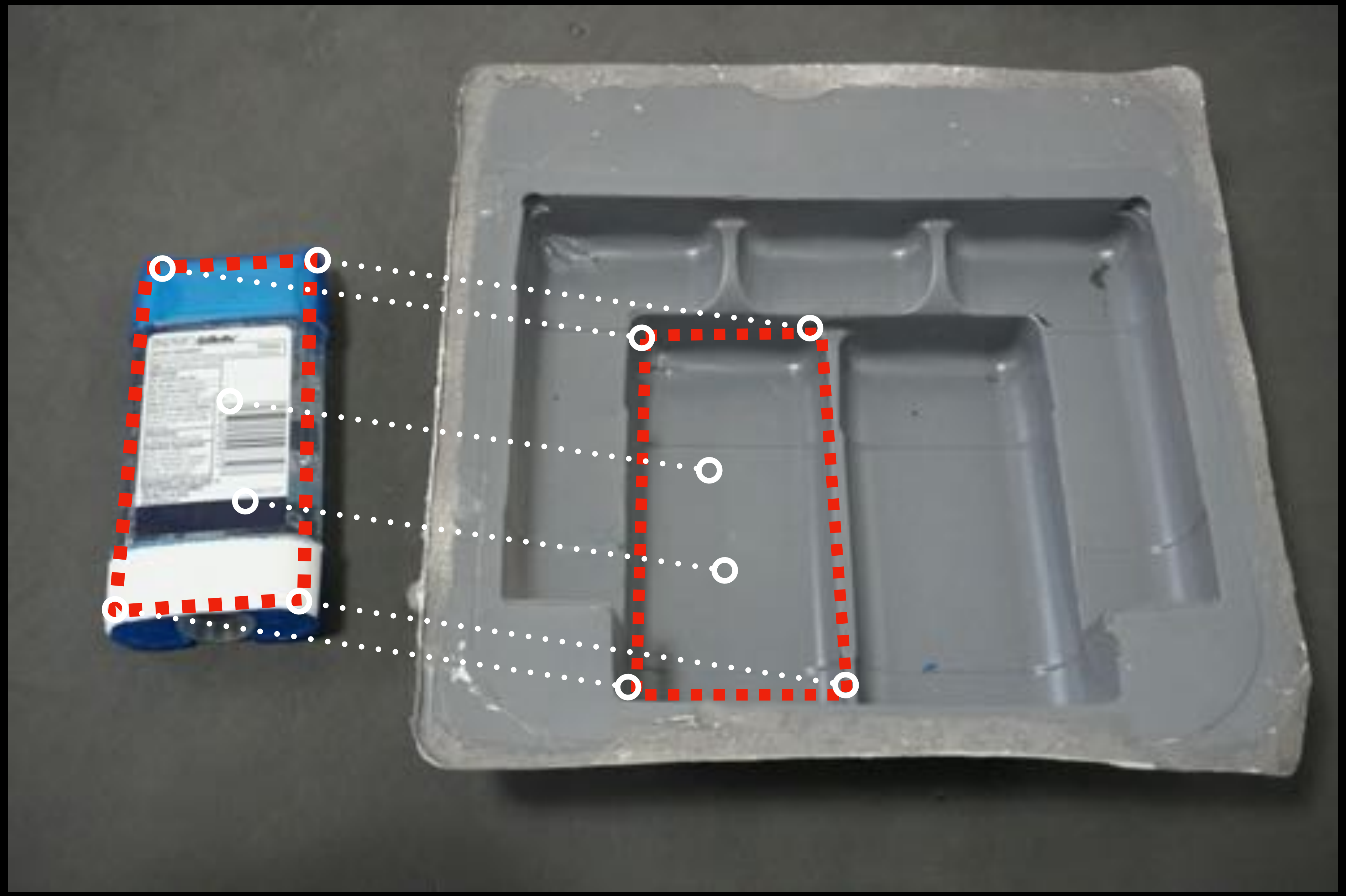
- learns geometric shape descriptors
- generalizes to new shapes

Assembly from Disassembly


- fully self-supervised
- trial and error

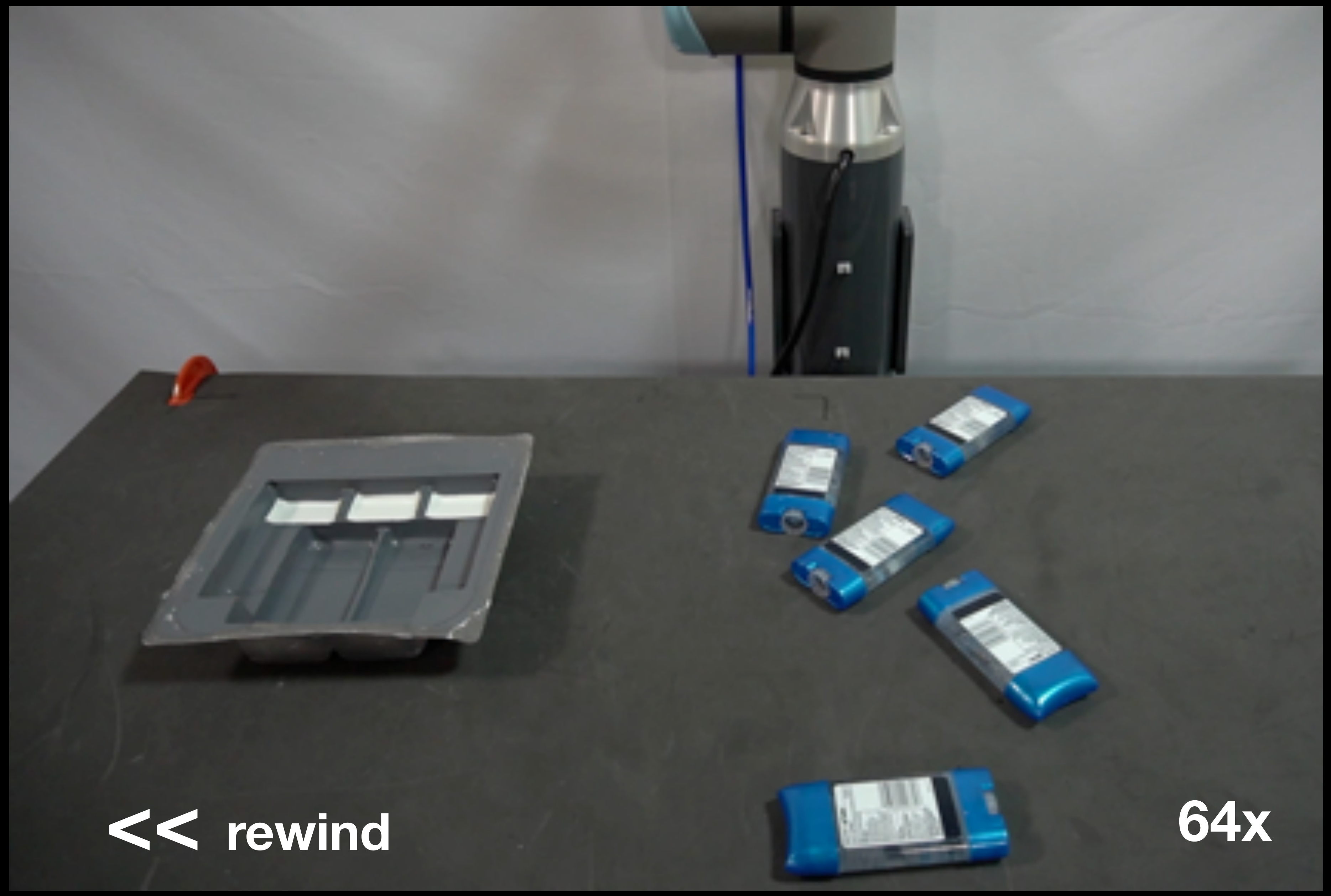
# Key Ideas

Kit Assembly → Shape Matching



- learns geometric shape descriptors
- generalizes to new shapes

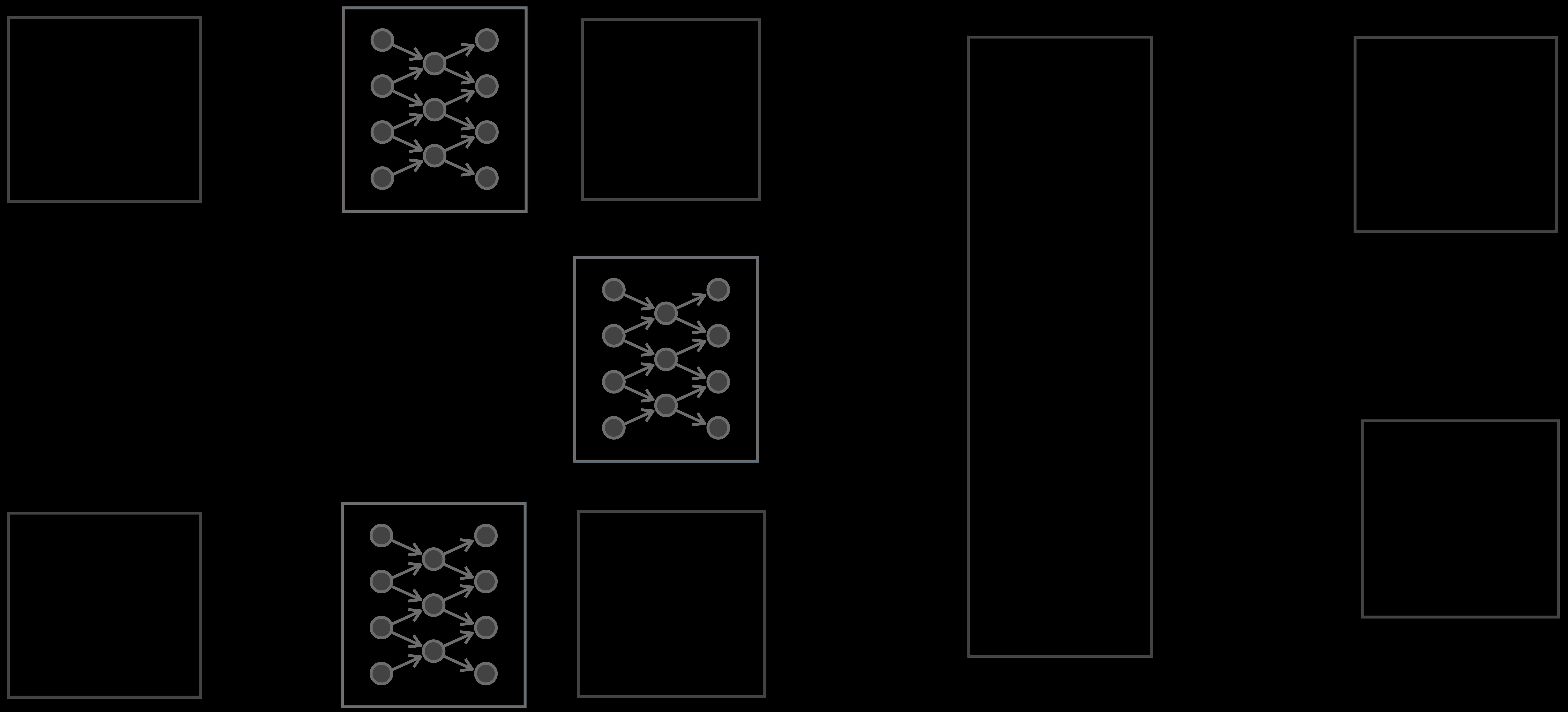
Assembly from Disassembly




- fully self-supervised
- trial and error

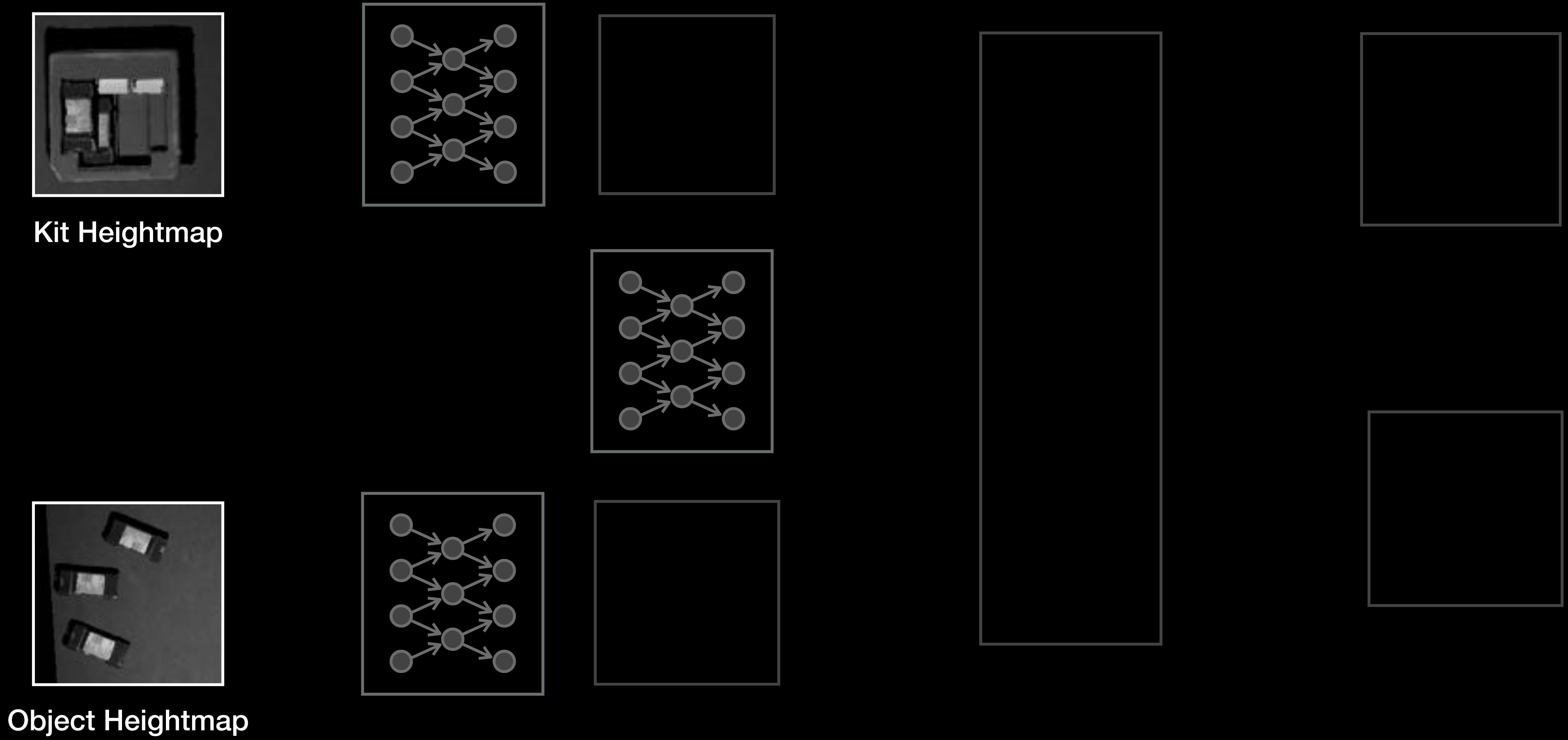
# Method

# Overview of Form2Fit



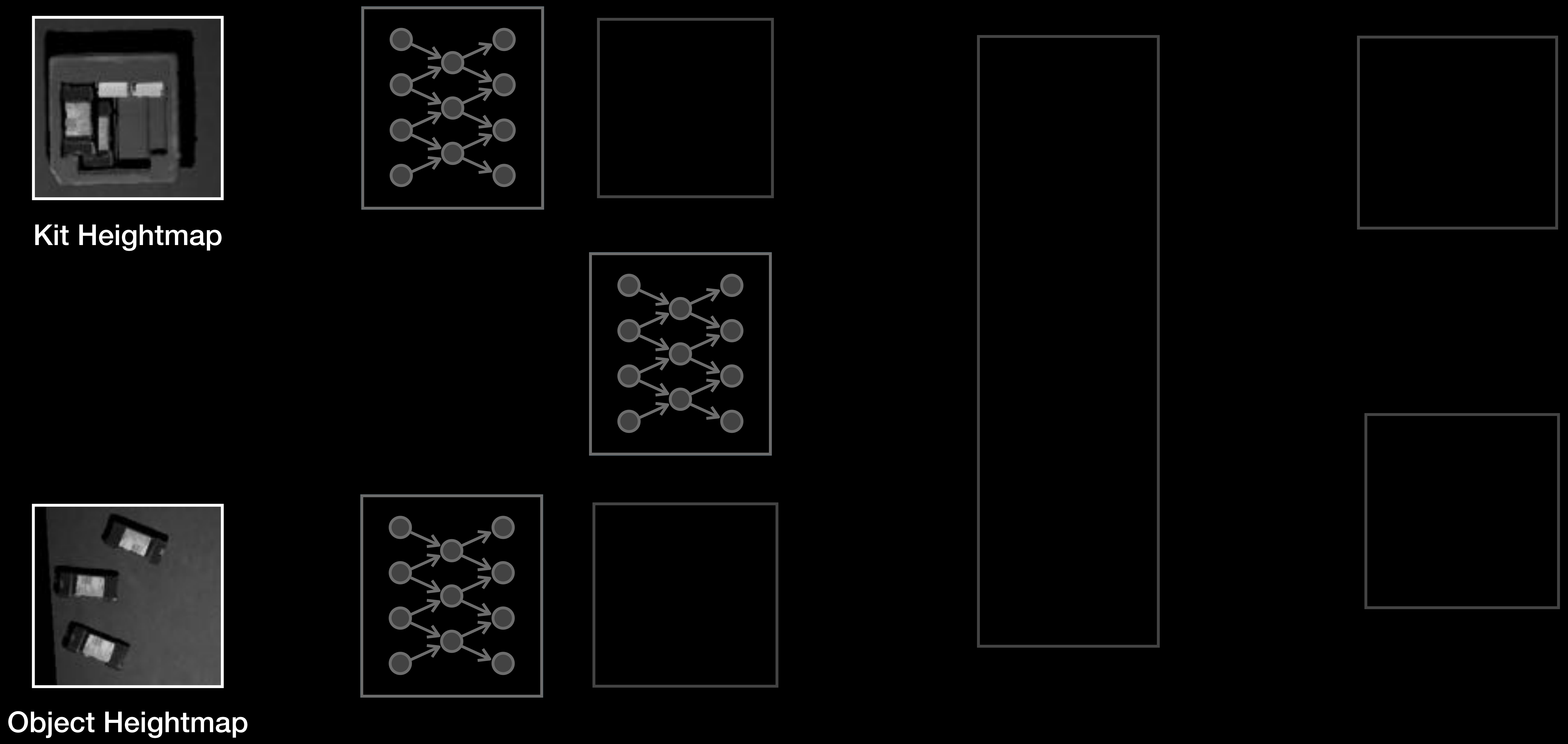
grayscale-depth heightmaps are generated from 3D pointcloud data

# Overview of Form2Fit



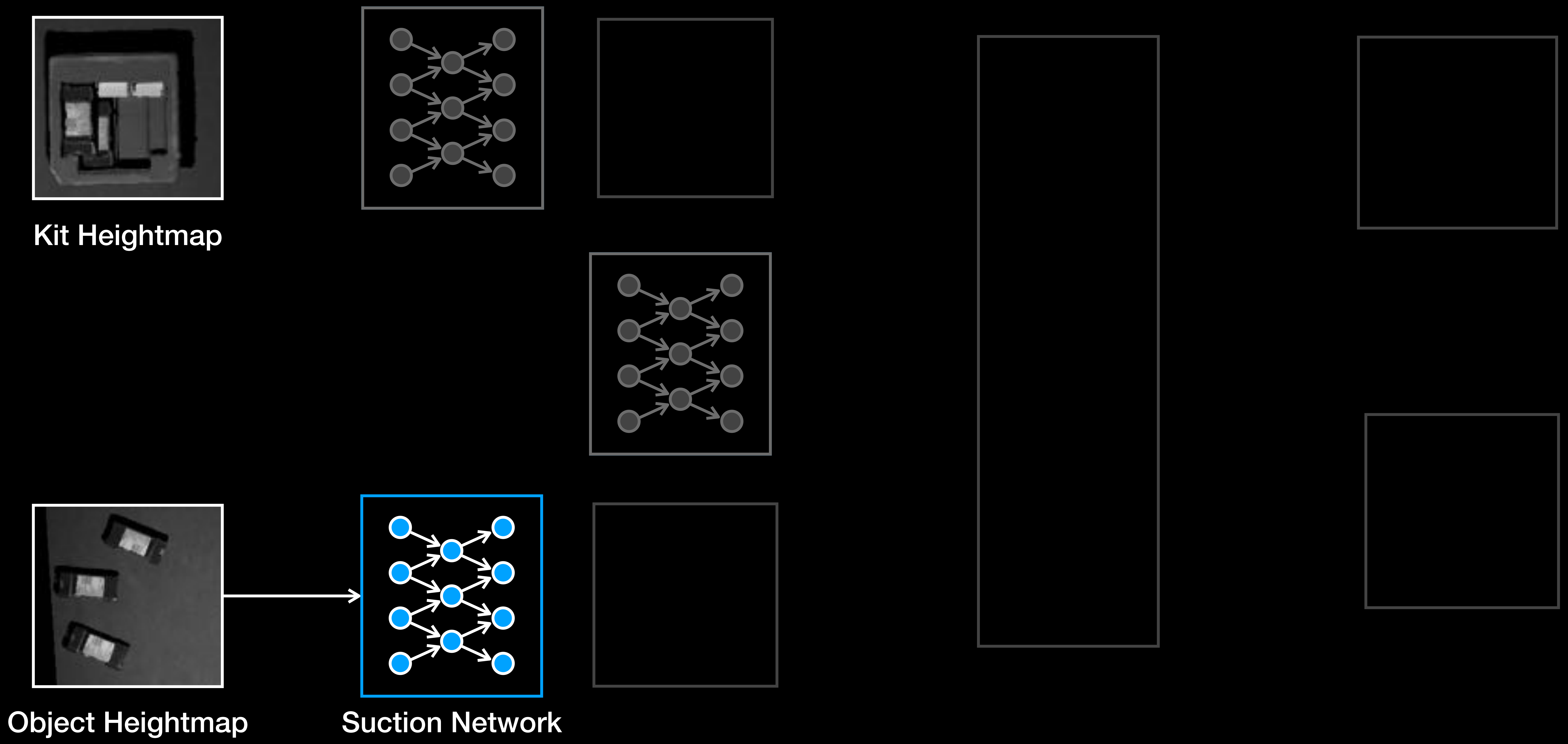

grayscale-depth heightmaps are generated from 3D pointcloud data

# Overview of Form2Fit



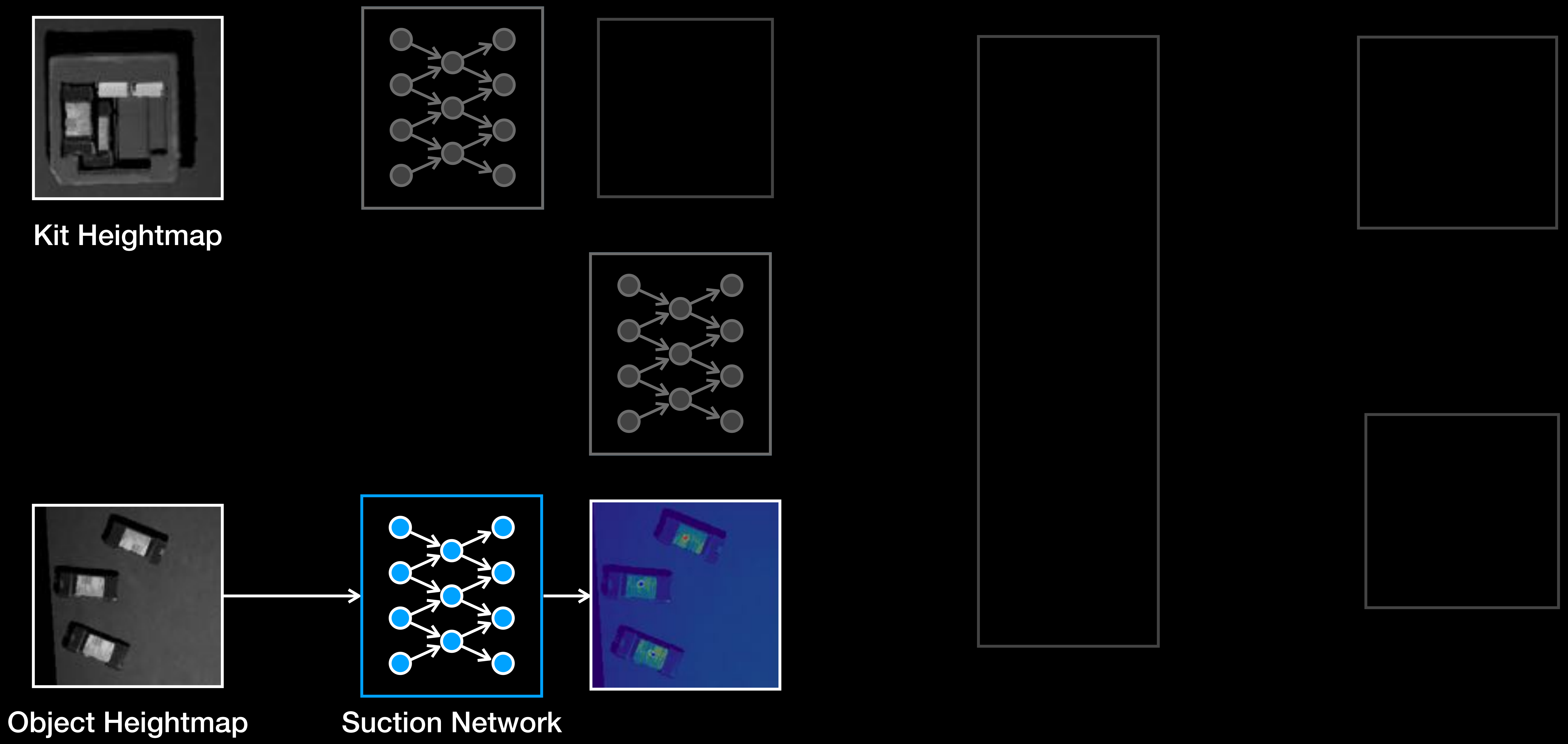

suction network ingests object heightmap and outputs suction heatmap

# Overview of Form2Fit



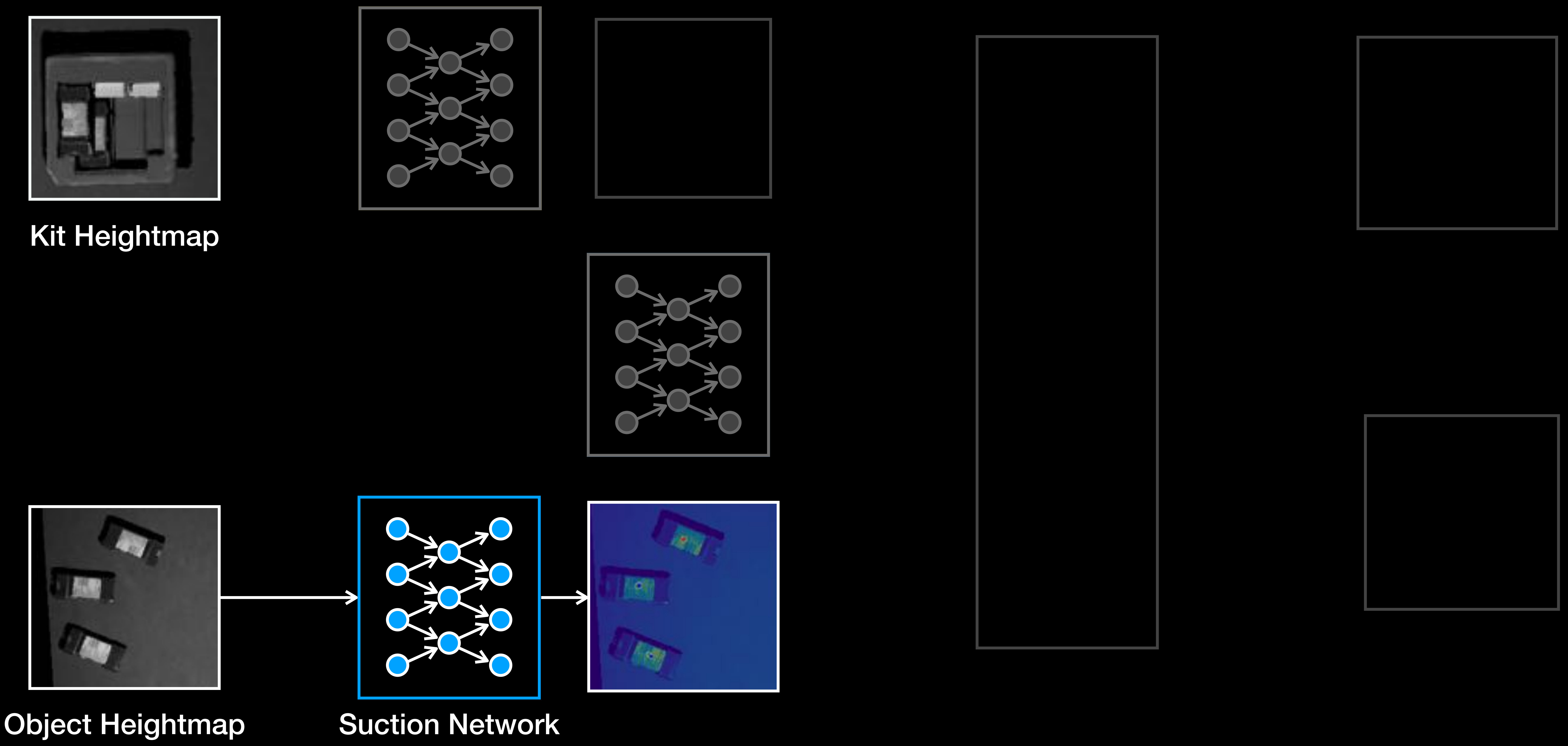

suction network ingests object heightmap and outputs suction heatmap

# Overview of Form2Fit



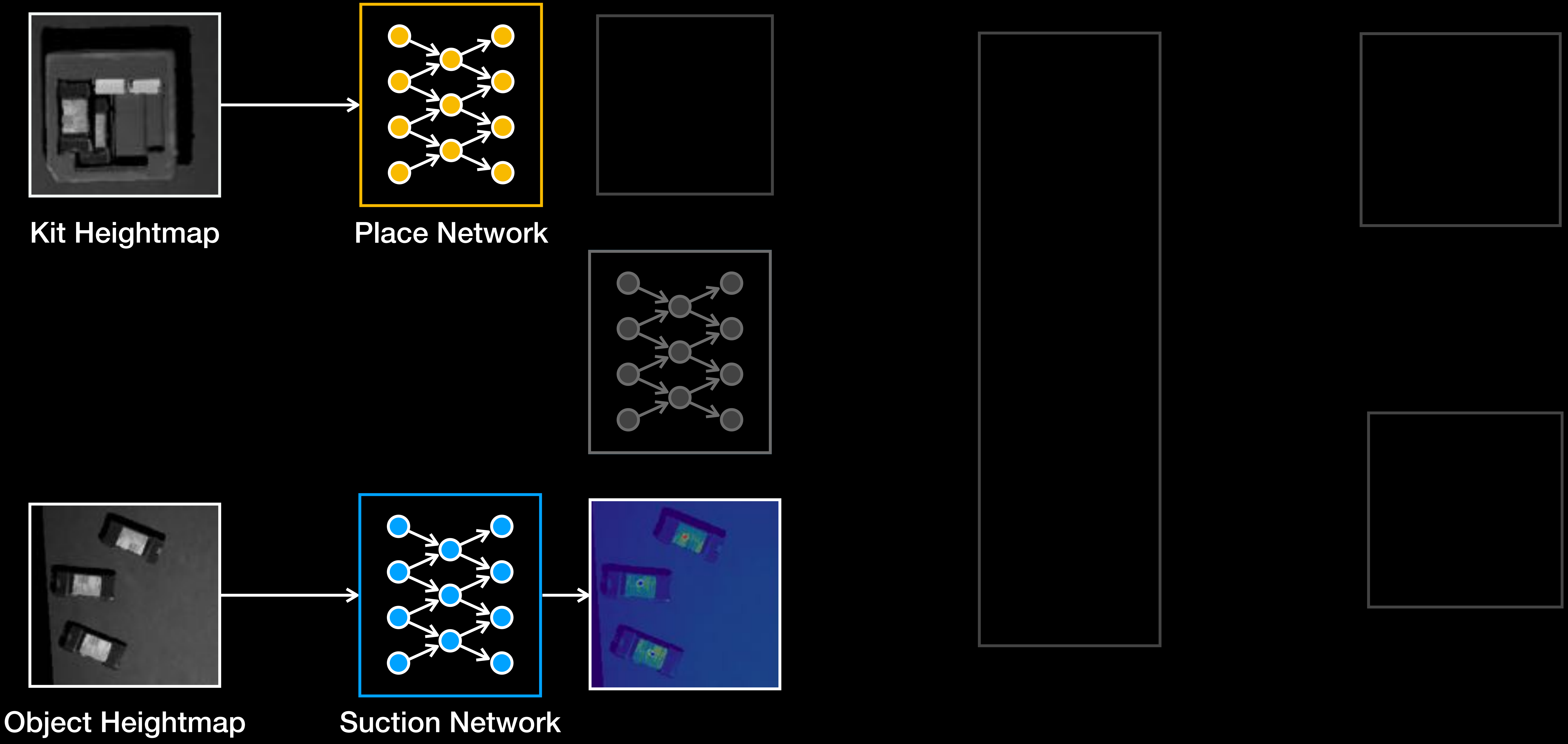

suction network ingests object heightmap and outputs suction heatmap

# Overview of Form2Fit



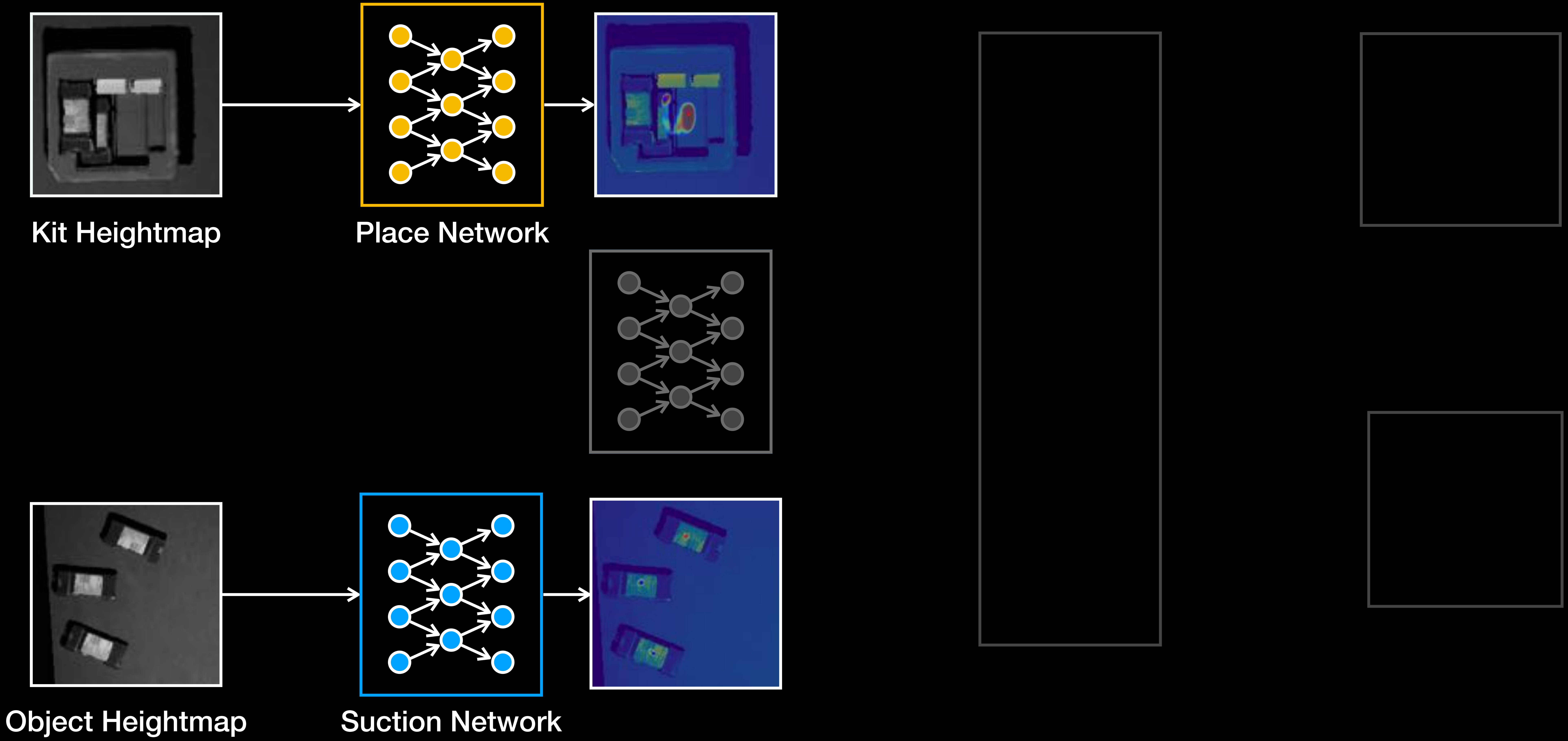

place network ingests kit heightmap and outputs place heatmap

# Overview of Form2Fit



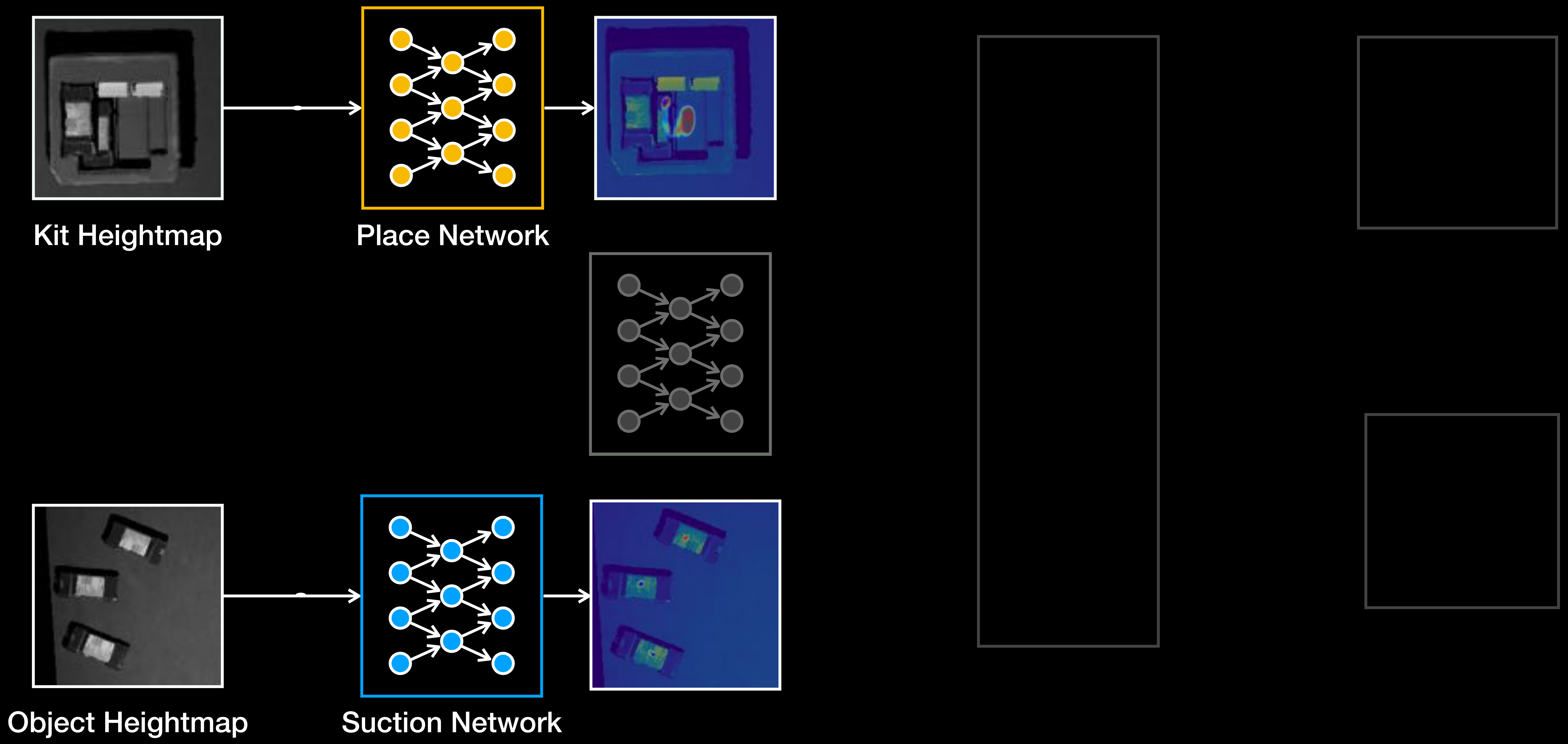

place network ingests kit heightmap and outputs place heatmap

# Overview of Form2Fit



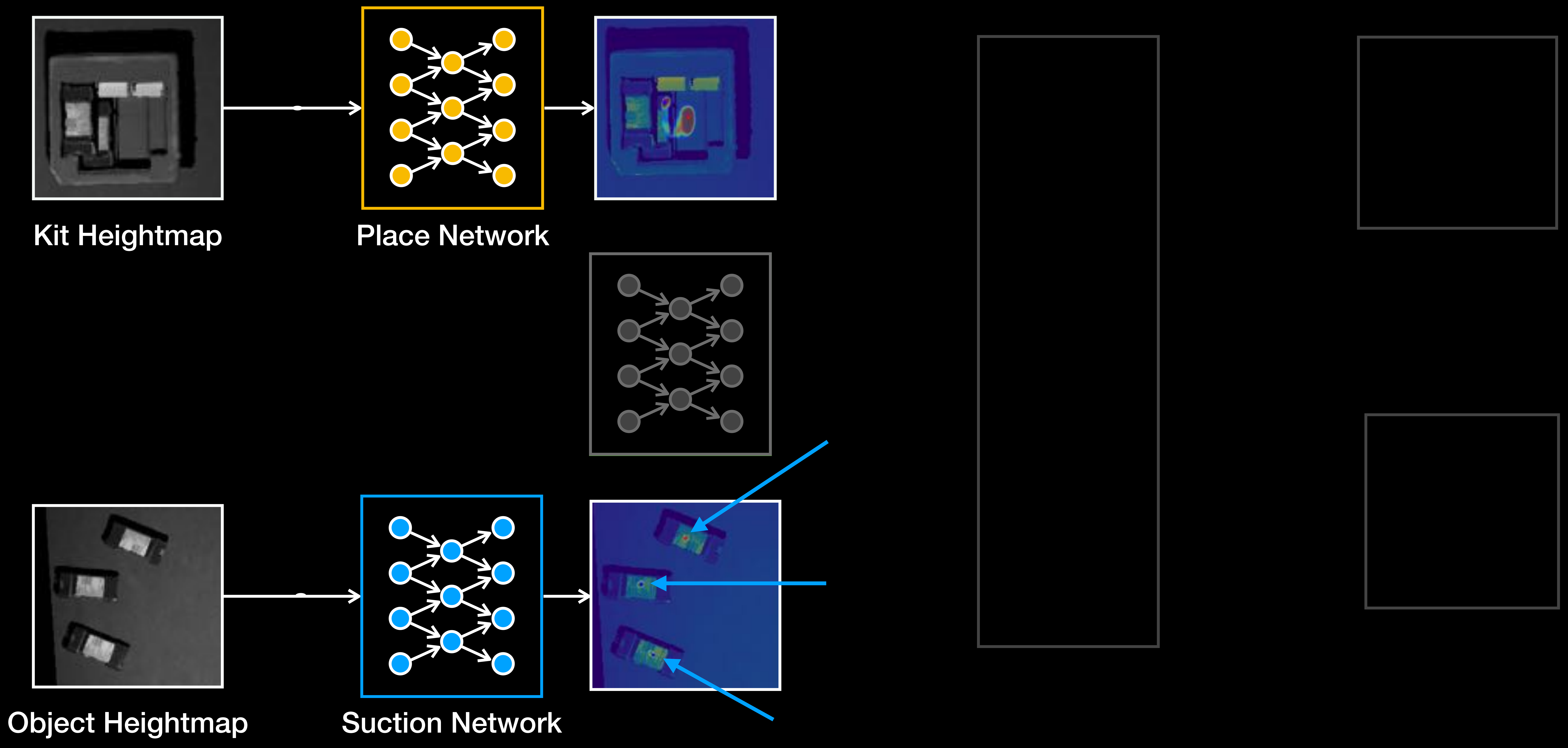

place network ingests kit heightmap and outputs place heatmap

# Overview of Form2Fit



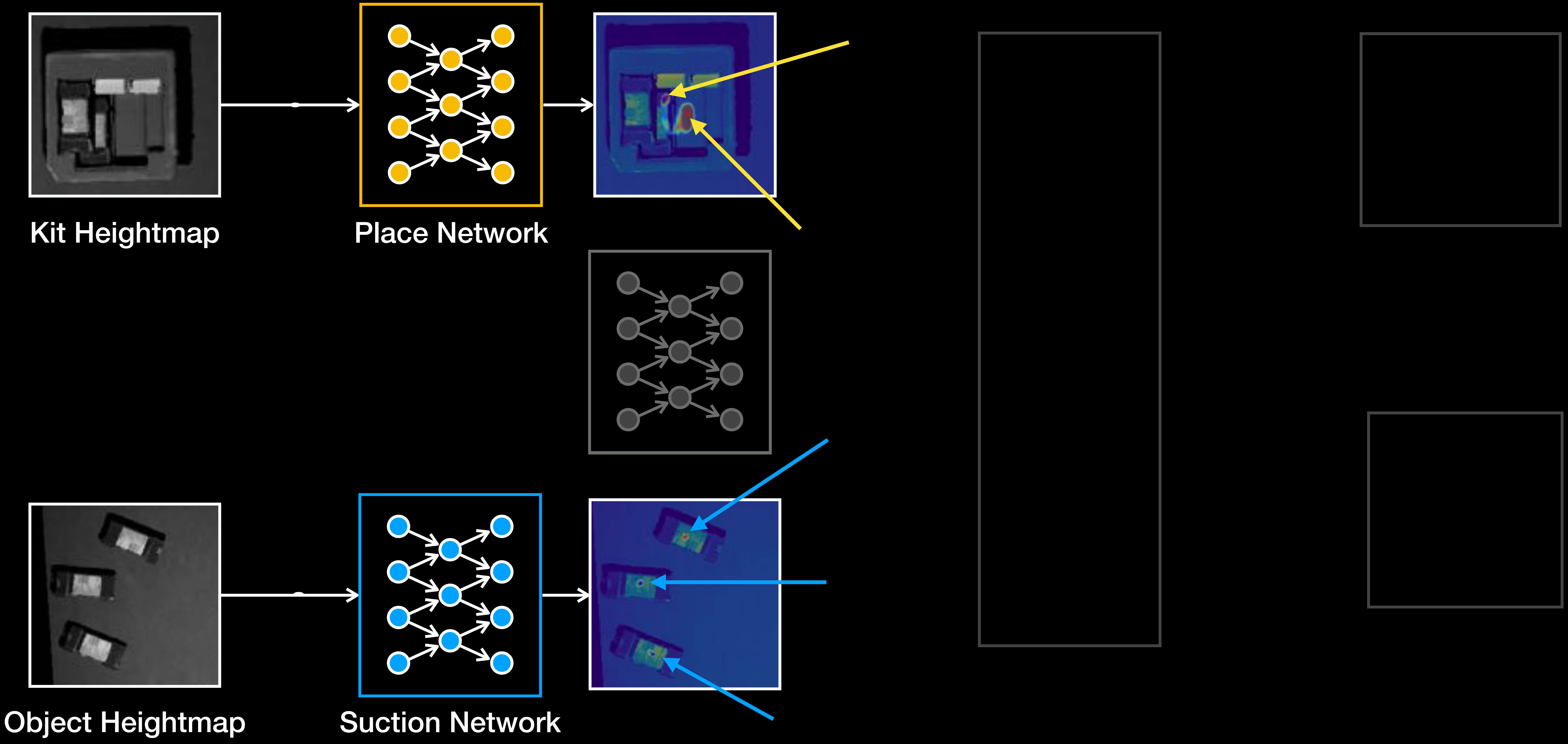

corresponding pick and place candidates

# Overview of Form2Fit



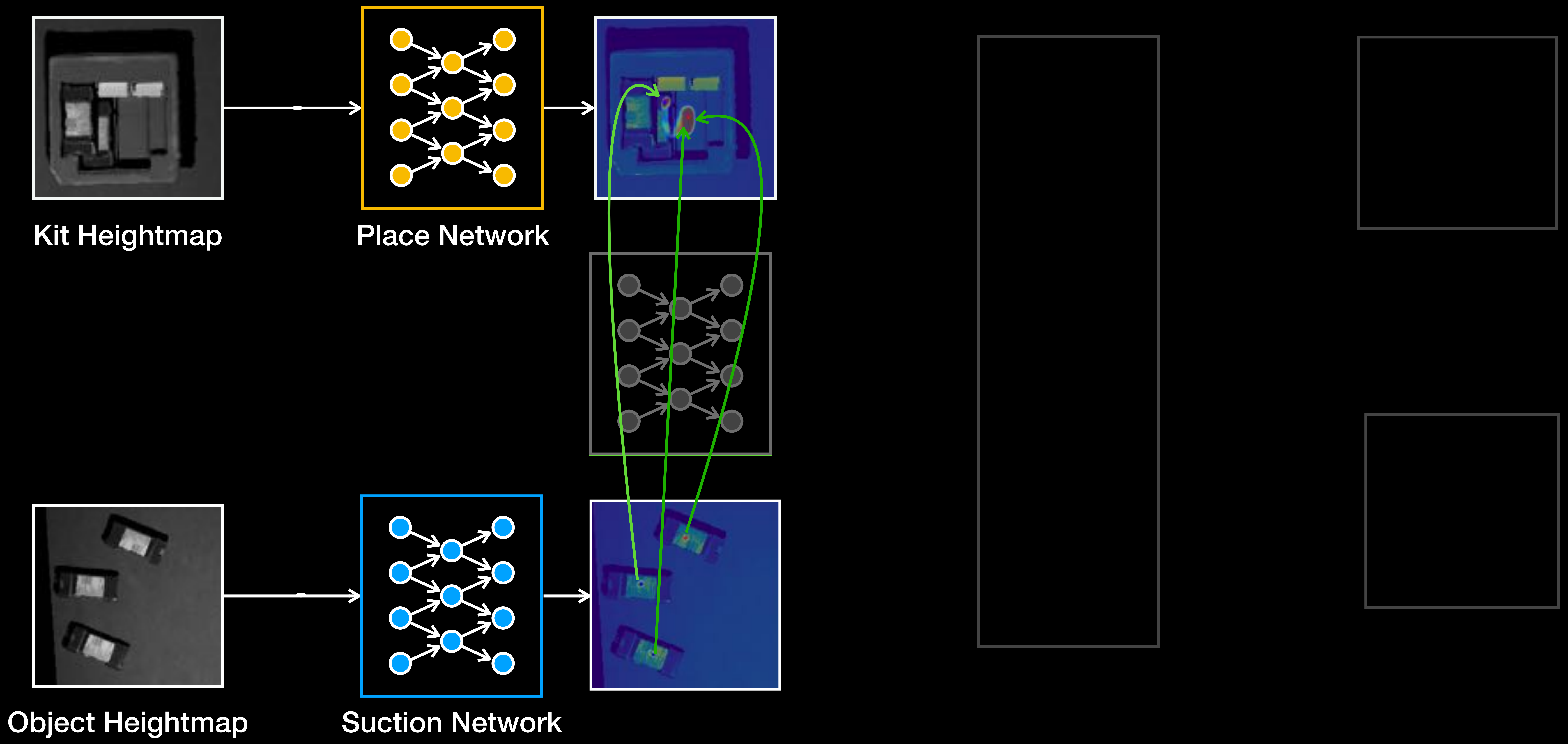

corresponding pick and place candidates

# Overview of Form2Fit



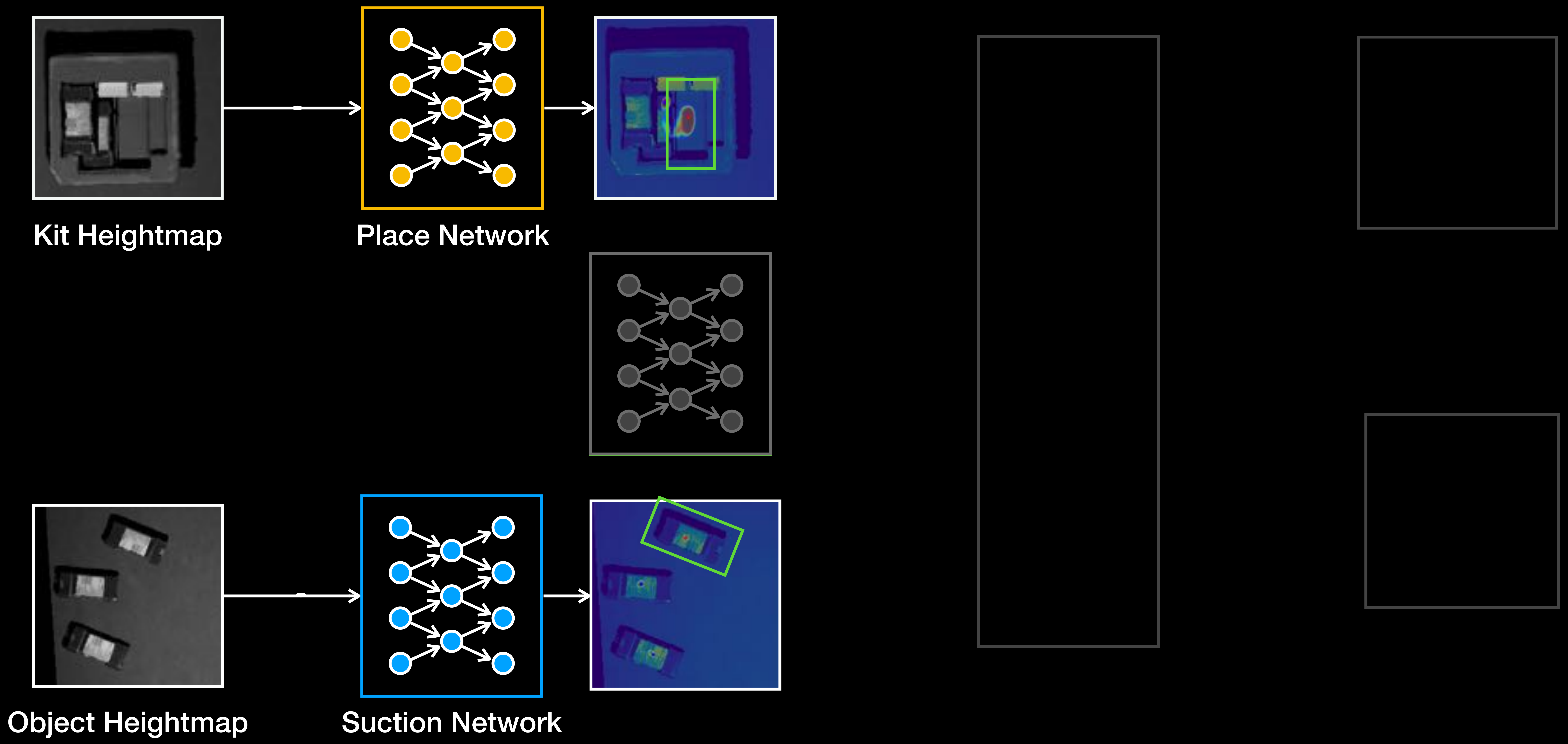

corresponding pick and place candidates

# Overview of Form2Fit



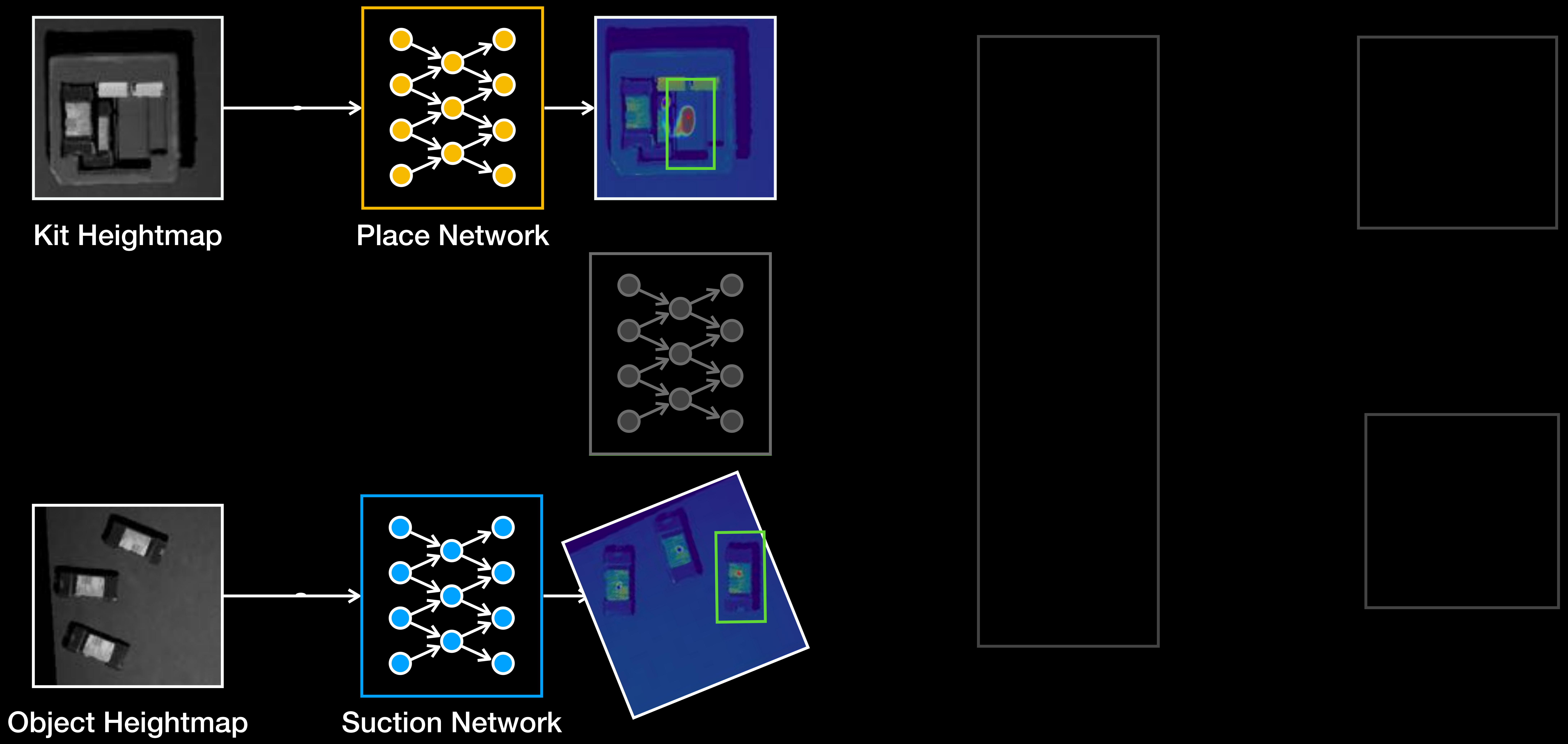

corresponding pick and place candidates

# Overview of Form2Fit



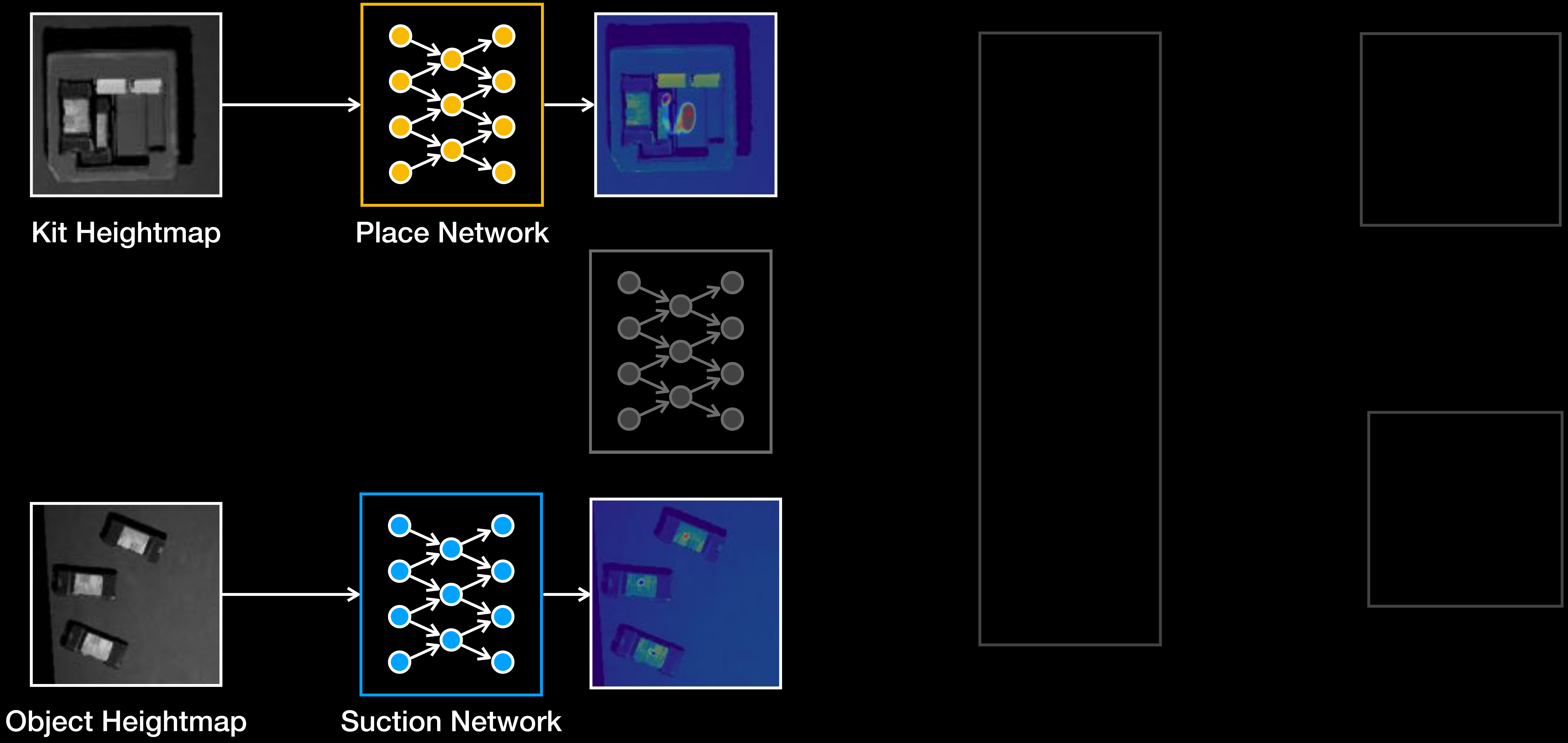

corresponding pick and place candidates

# Overview of Form2Fit



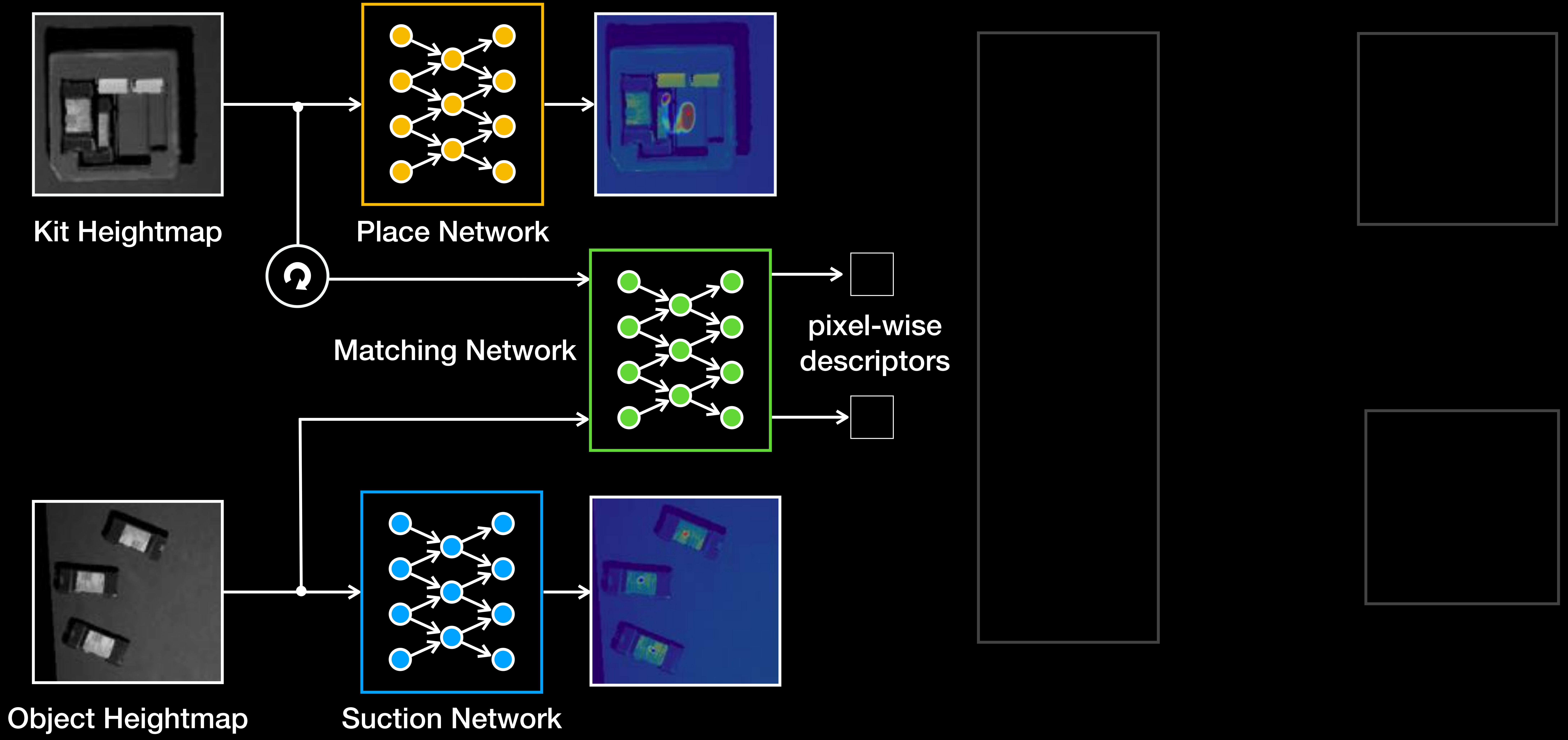

corresponding pick and place candidates

# Overview of Form2Fit



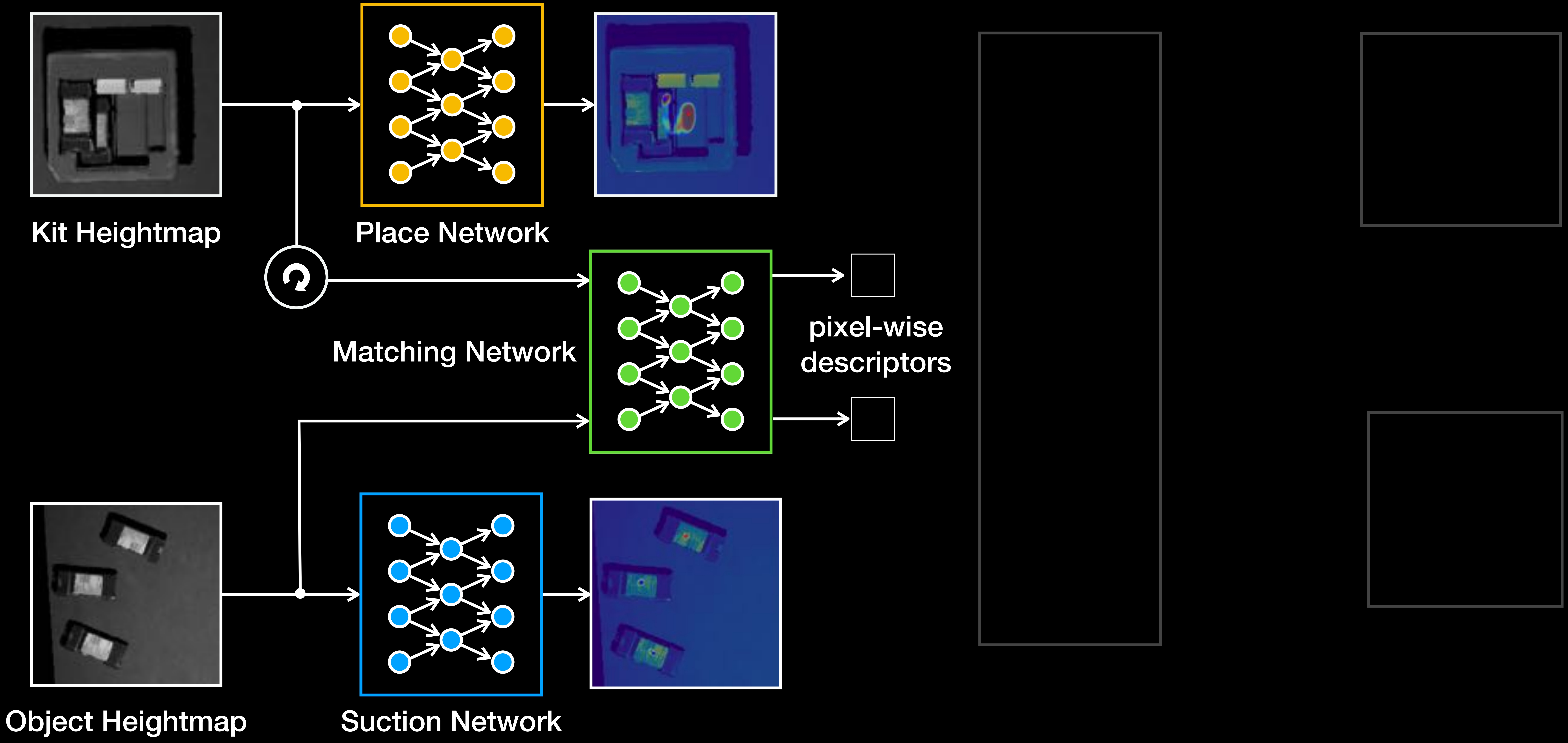

matching network ingests heightmaps and outputs descriptor maps

# Overview of Form2Fit



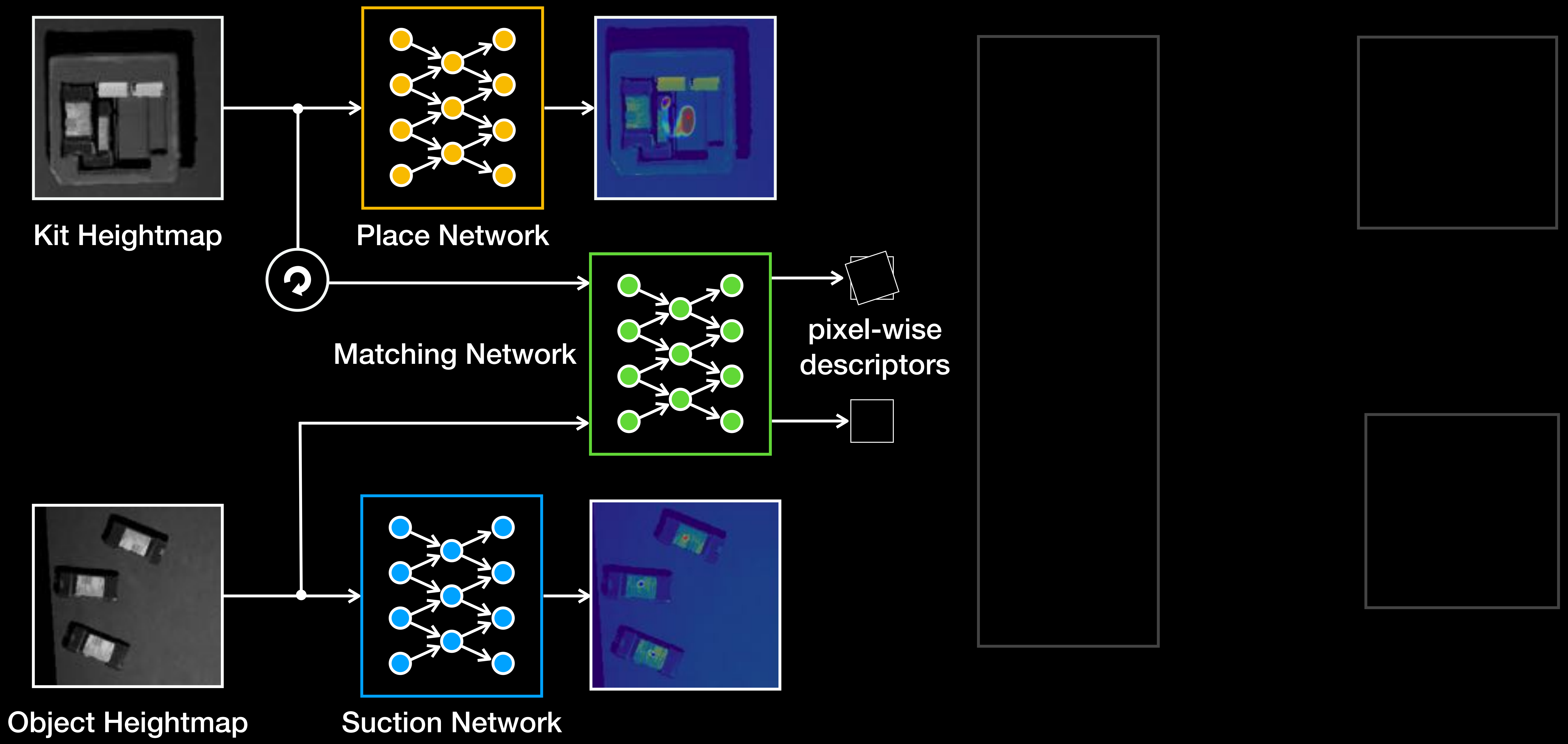

matching network ingests heightmaps and outputs descriptor maps

# Overview of Form2Fit



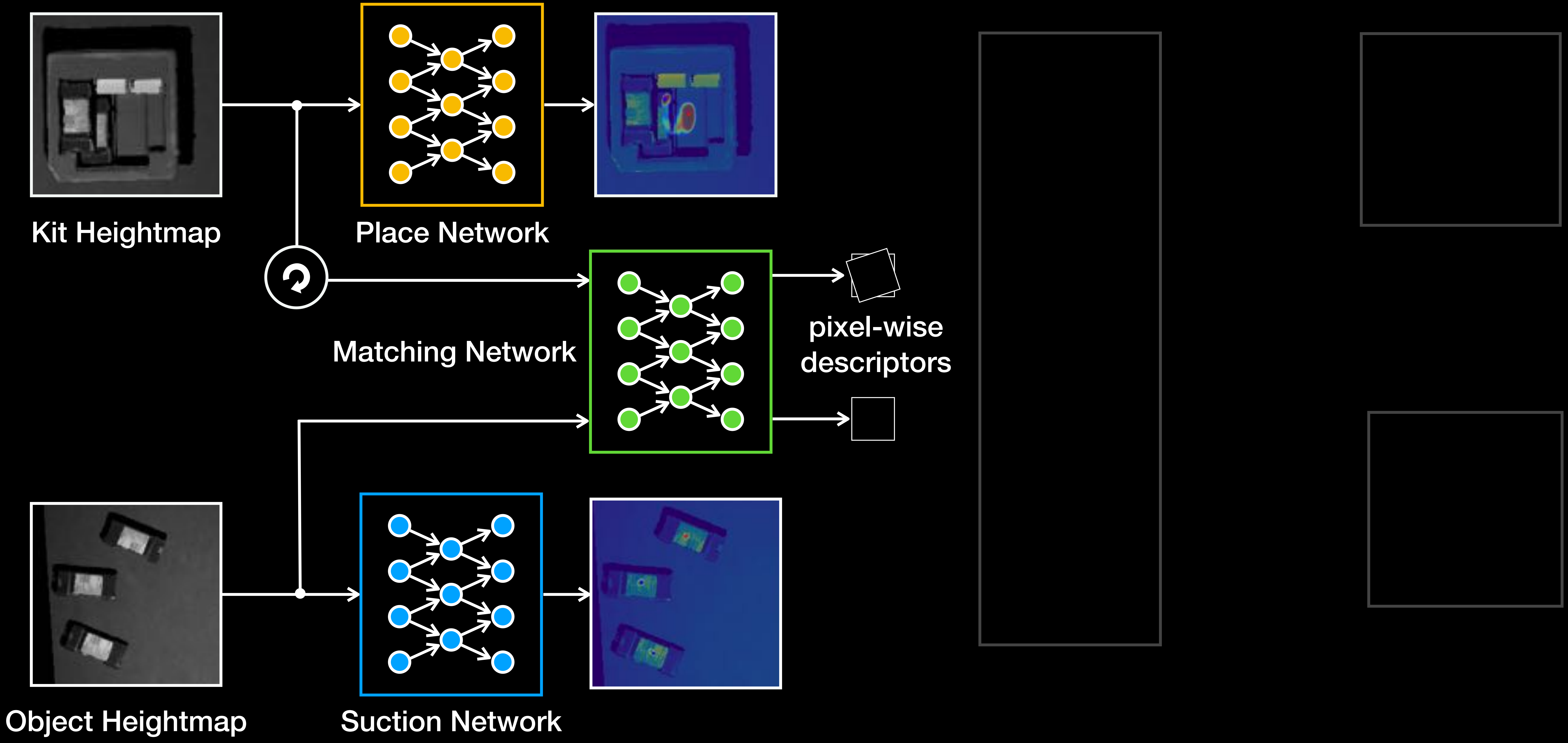

closer descriptor distances indicate better object-to-placement correspondences

# Overview of Form2Fit



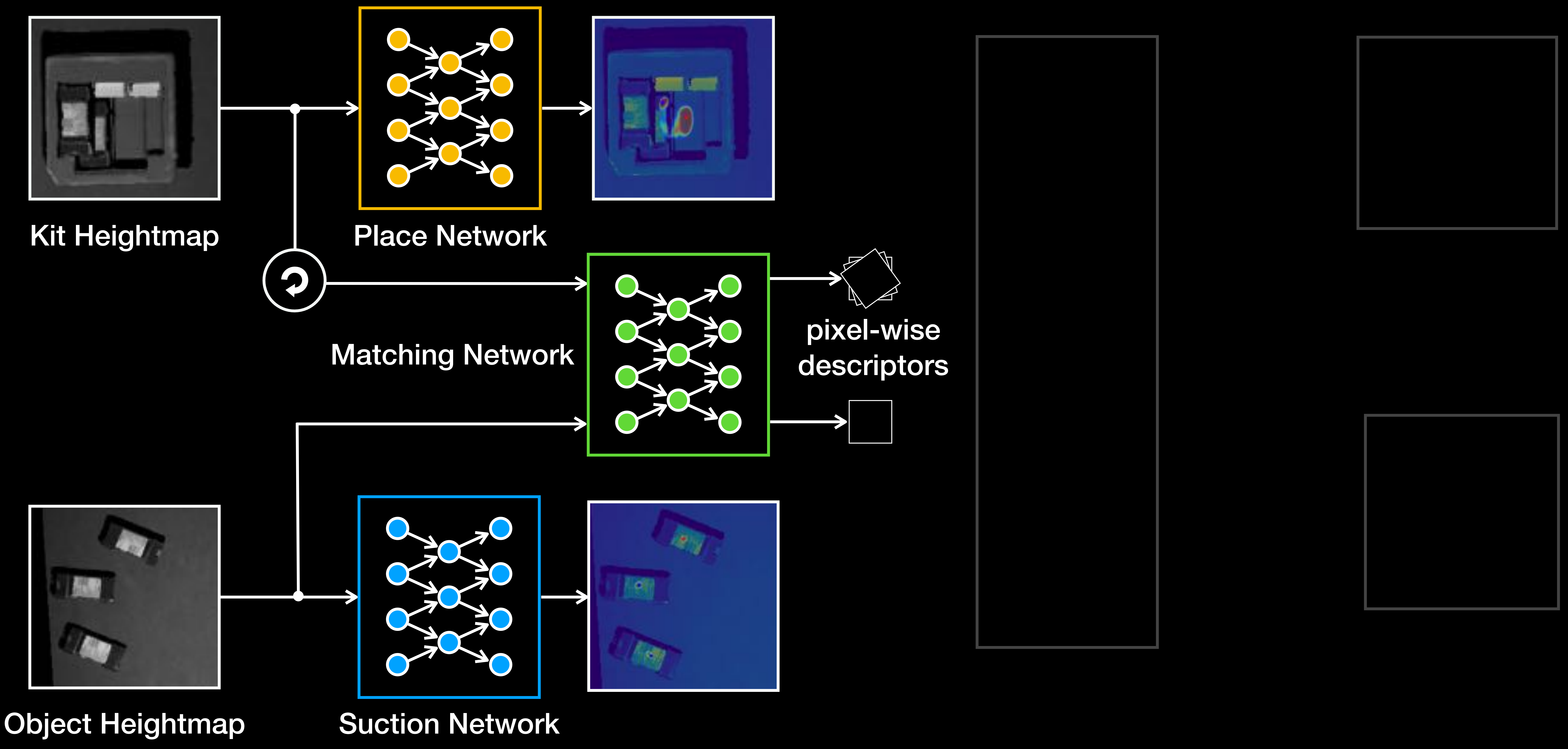

closer descriptor distances indicate better object-to-placement correspondences

# Overview of Form2Fit



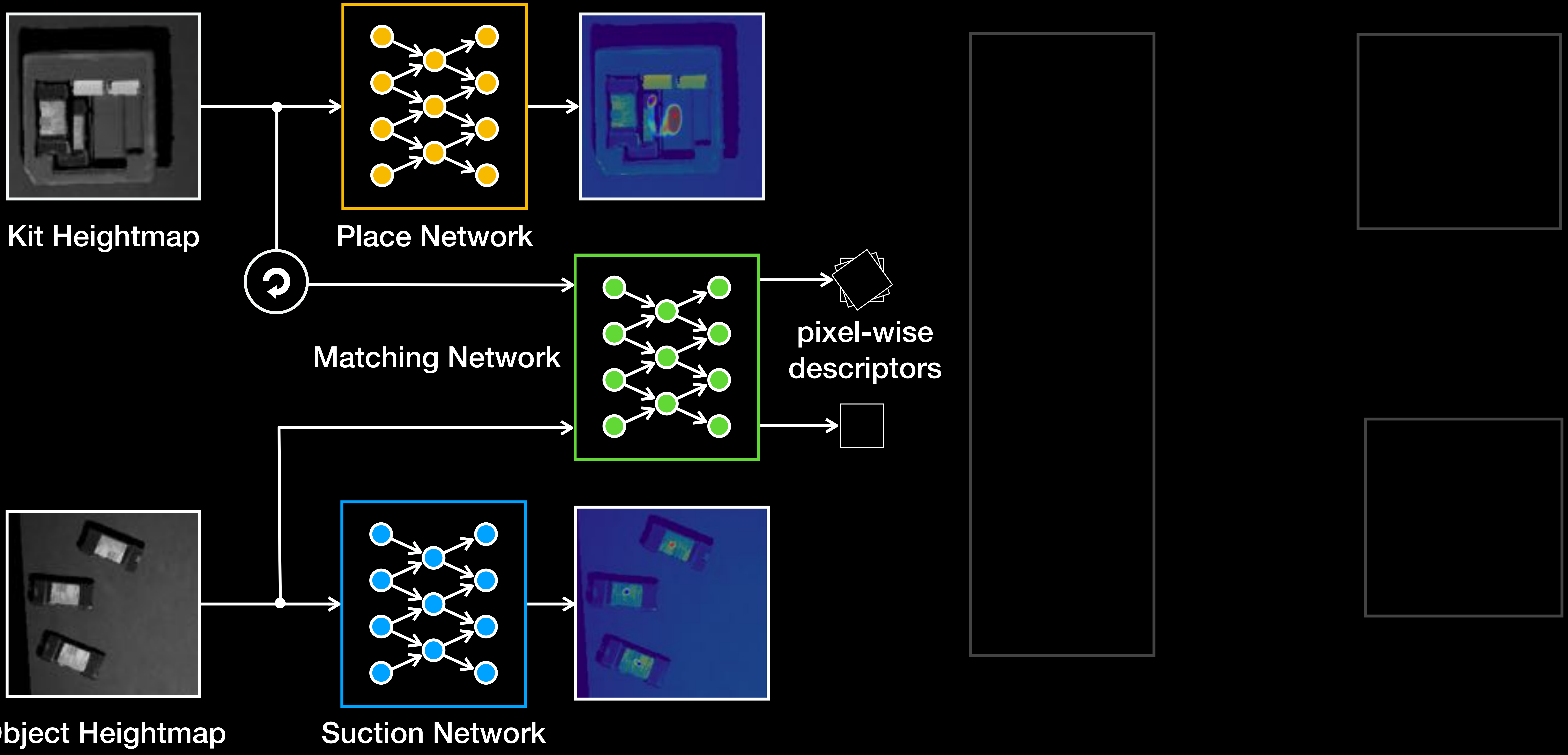

closer descriptor distances indicate better object-to-placement correspondences

# Overview of Form2Fit



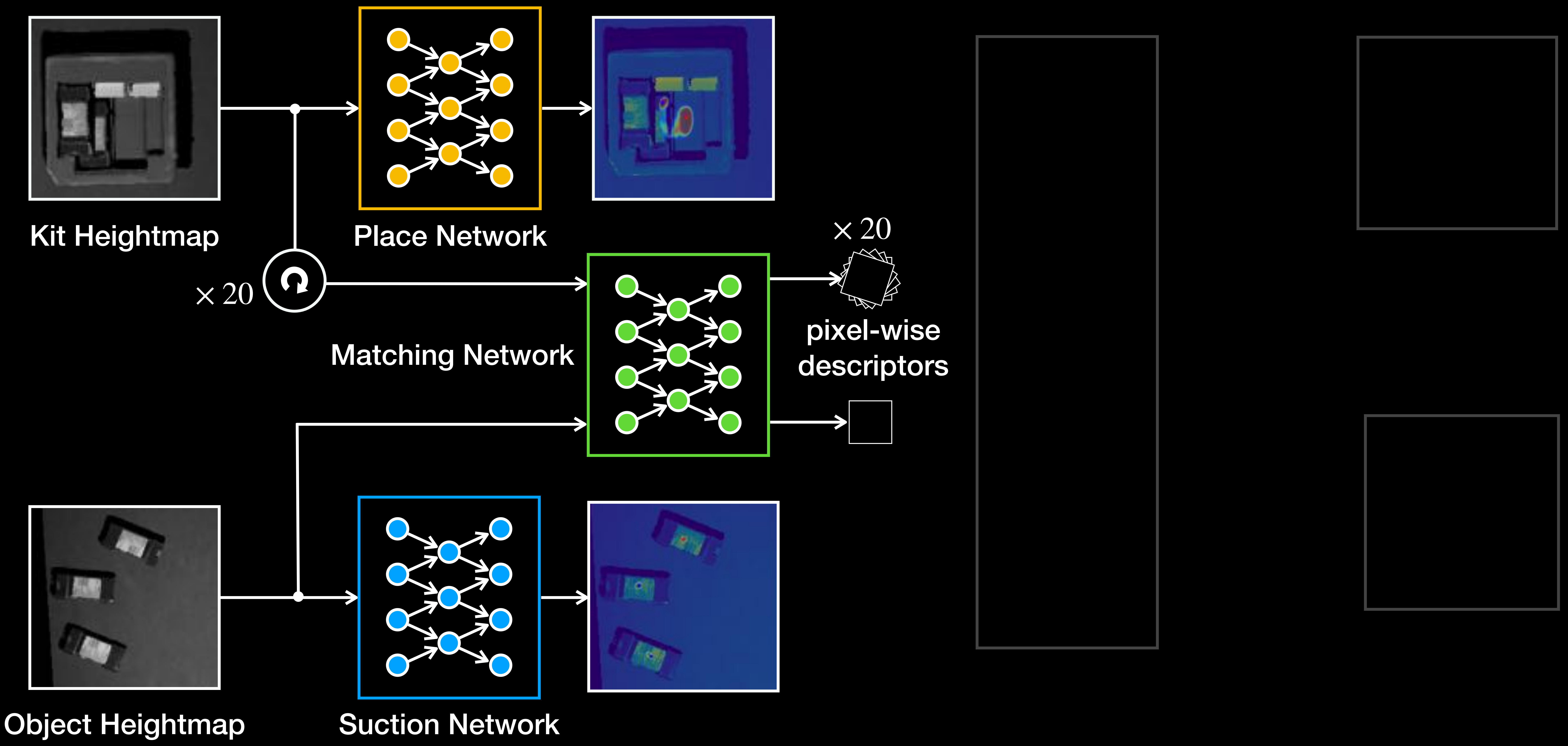

closer descriptor distances indicate better object-to-placement correspondences

# Overview of Form2Fit



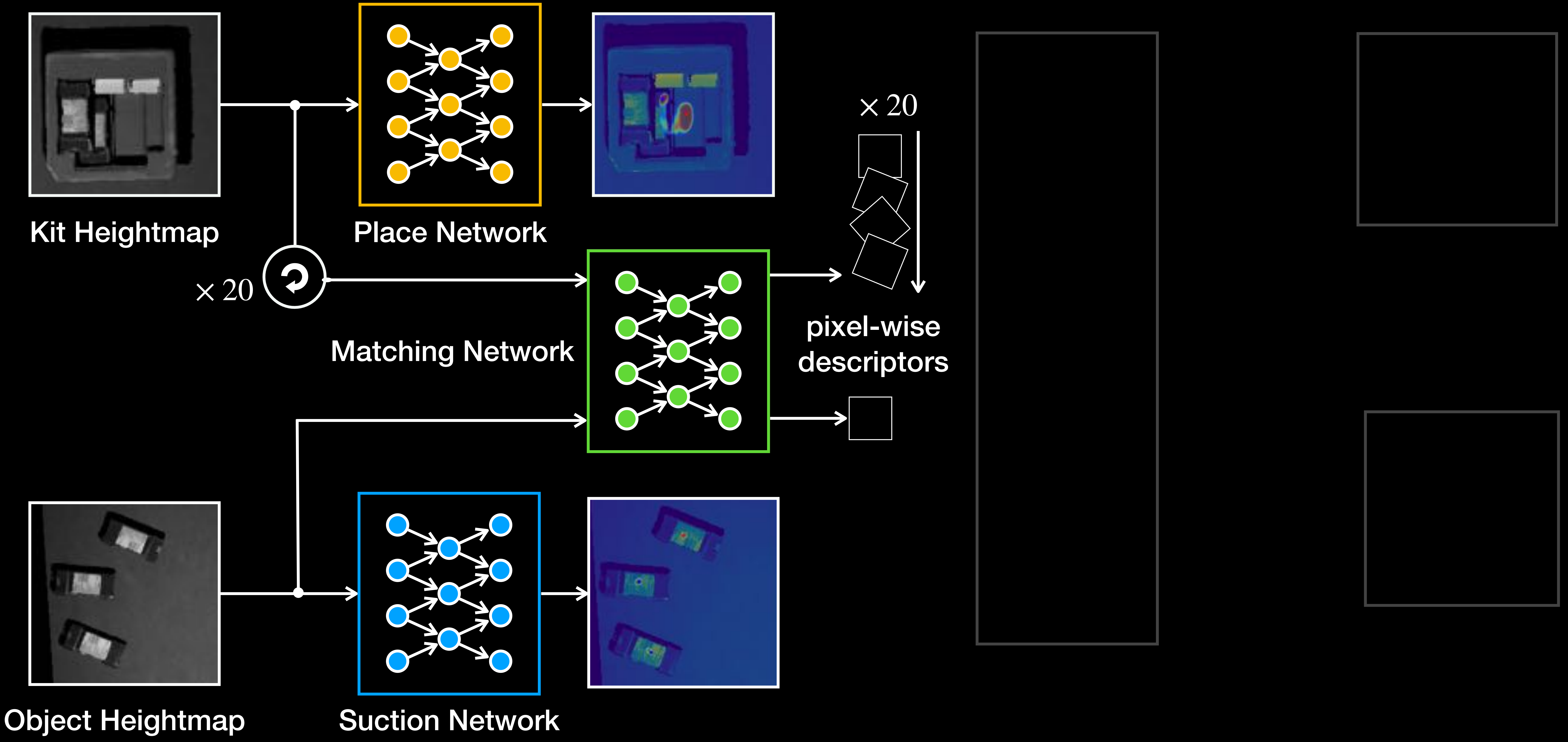

closer descriptor distances indicate better object-to-placement correspondences

# Overview of Form2Fit



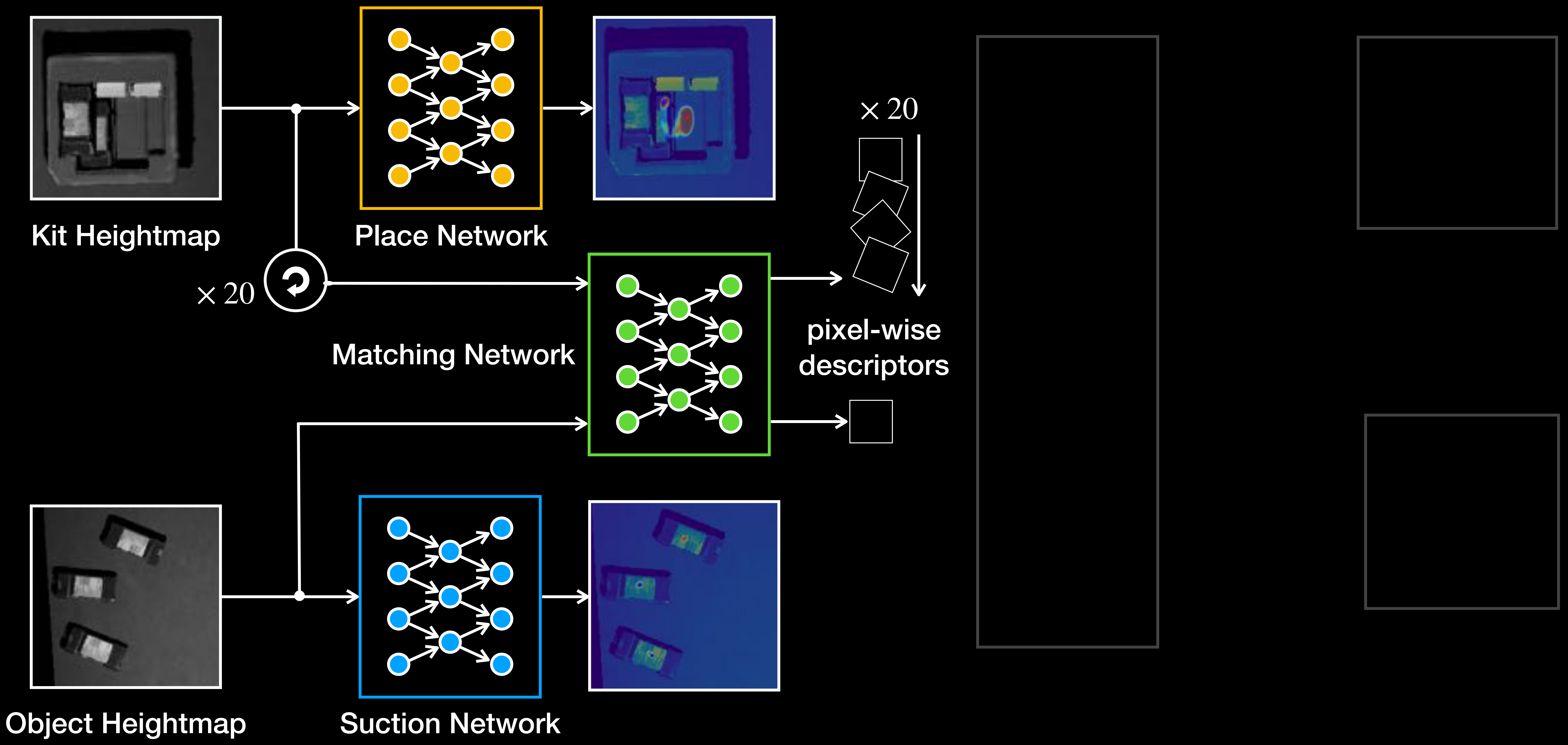

closer descriptor distances indicate better object-to-placement correspondences

# Overview of Form2Fit



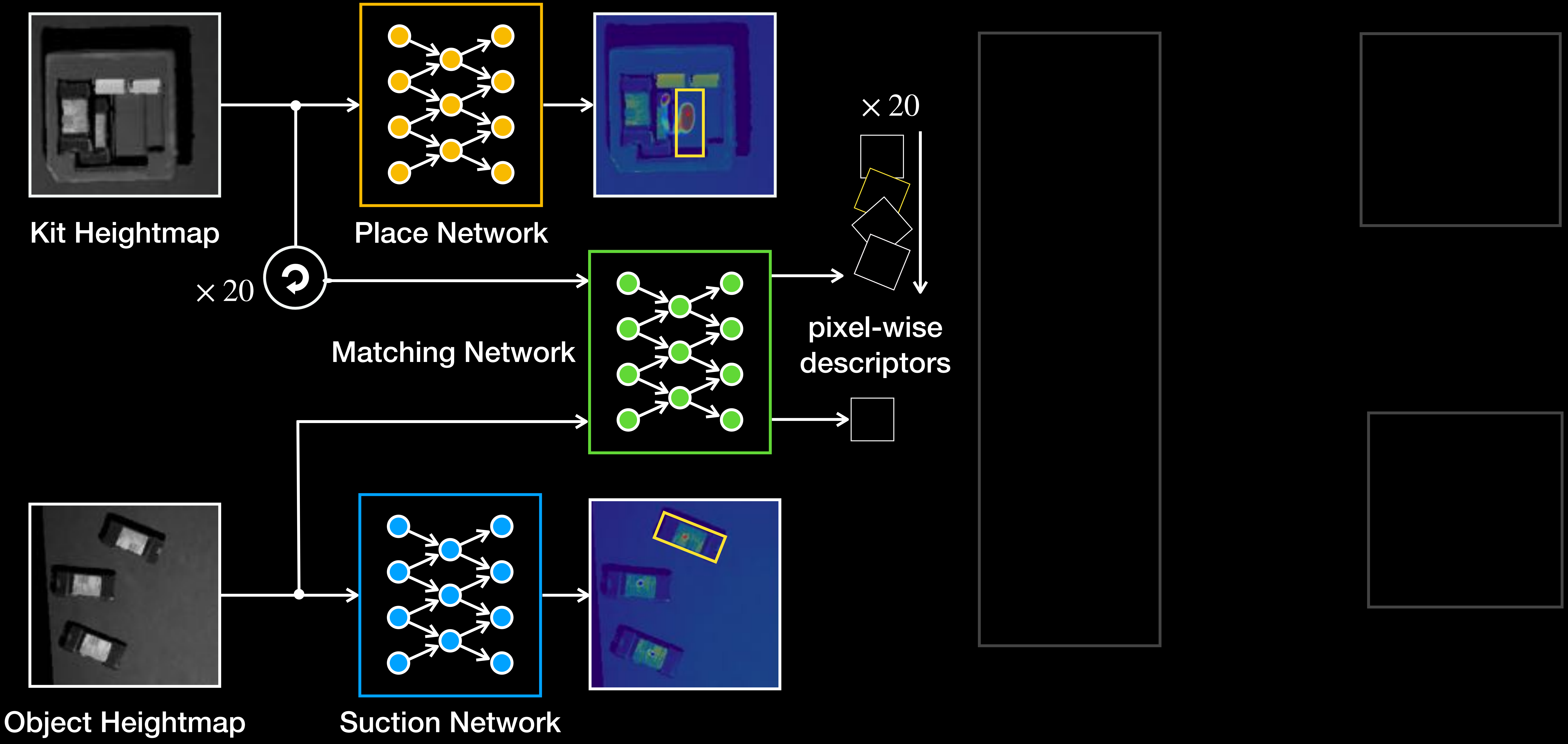

descriptors are rotation-sensitive

# Overview of Form2Fit



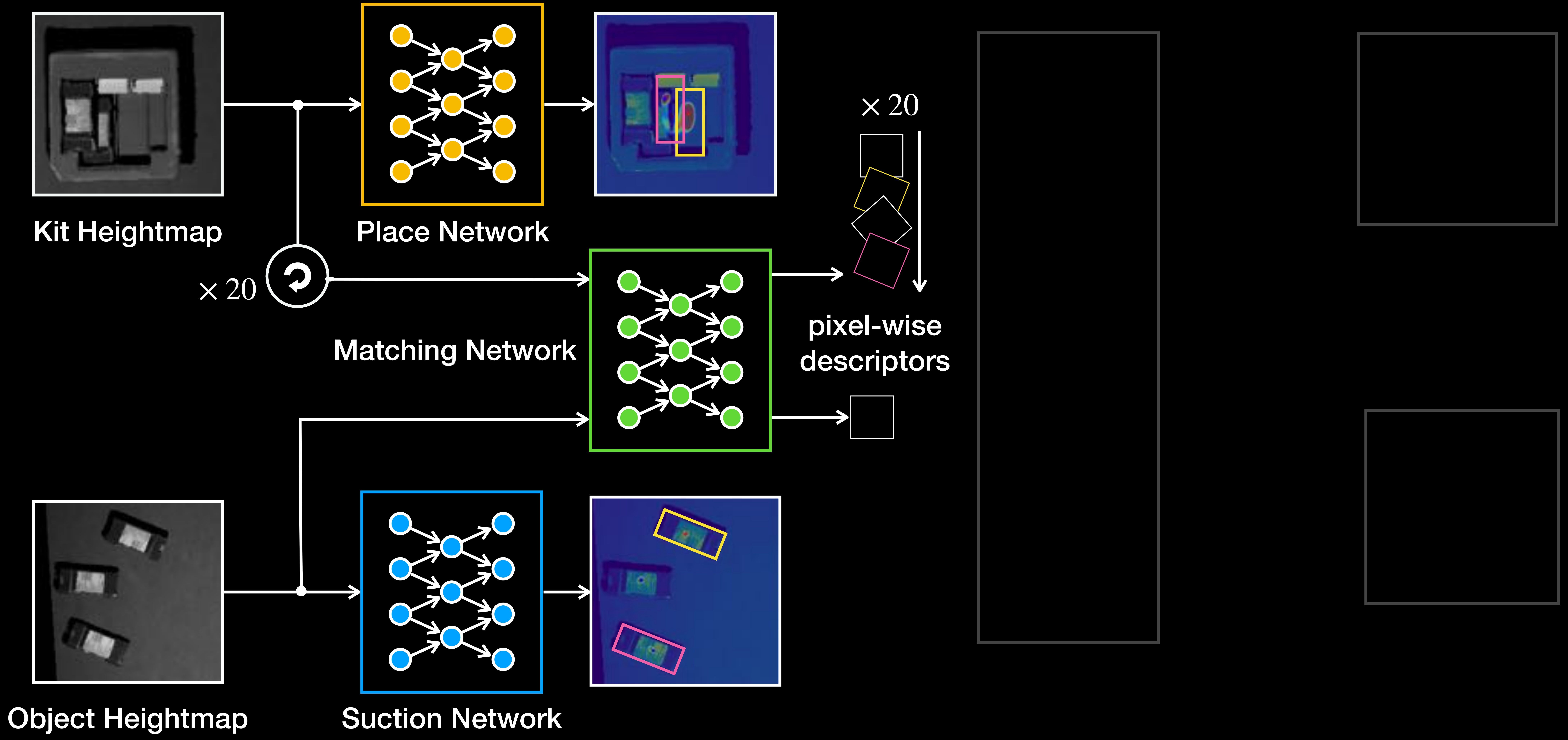

descriptors are rotation-sensitive

# Overview of Form2Fit



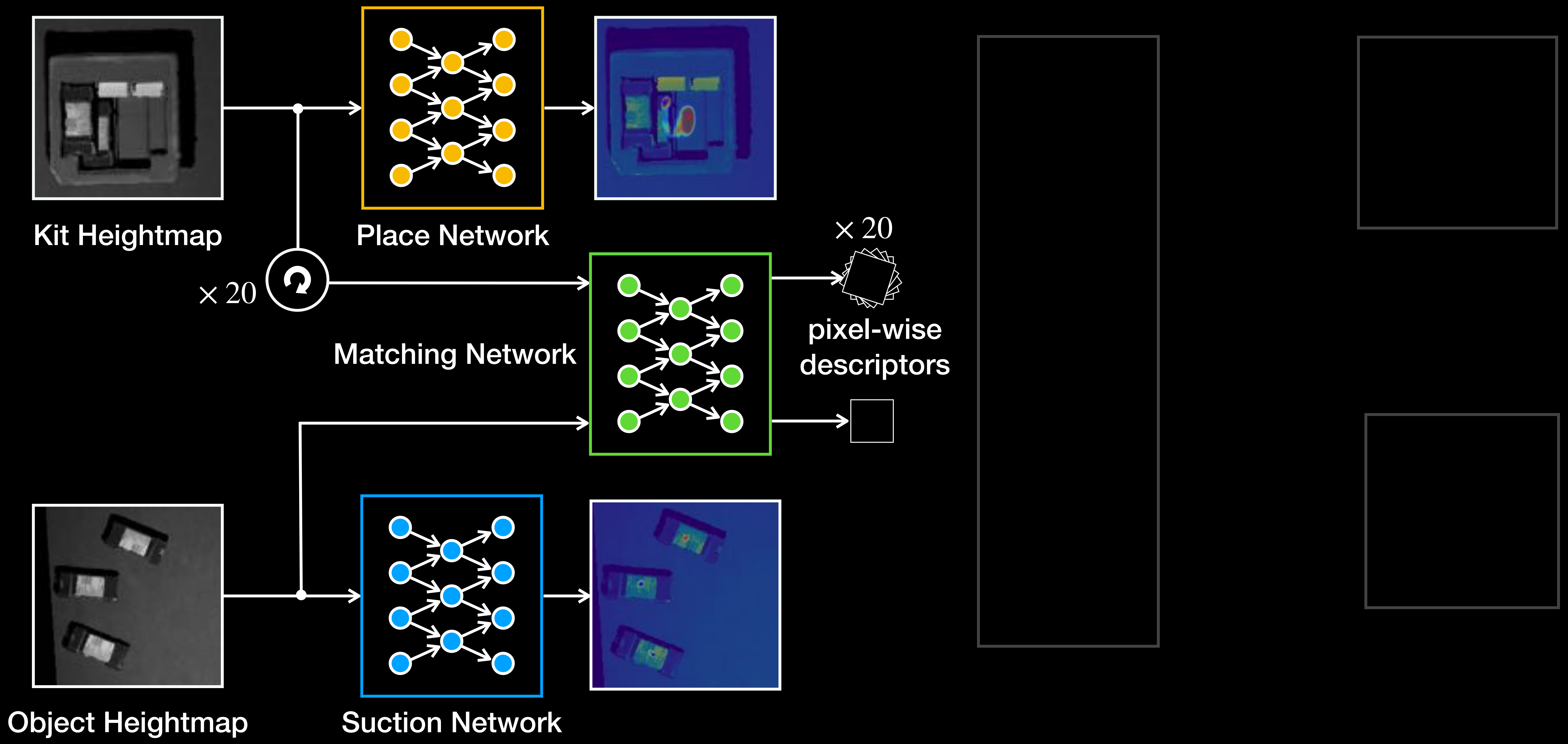

descriptors are rotation-sensitive

# Overview of Form2Fit



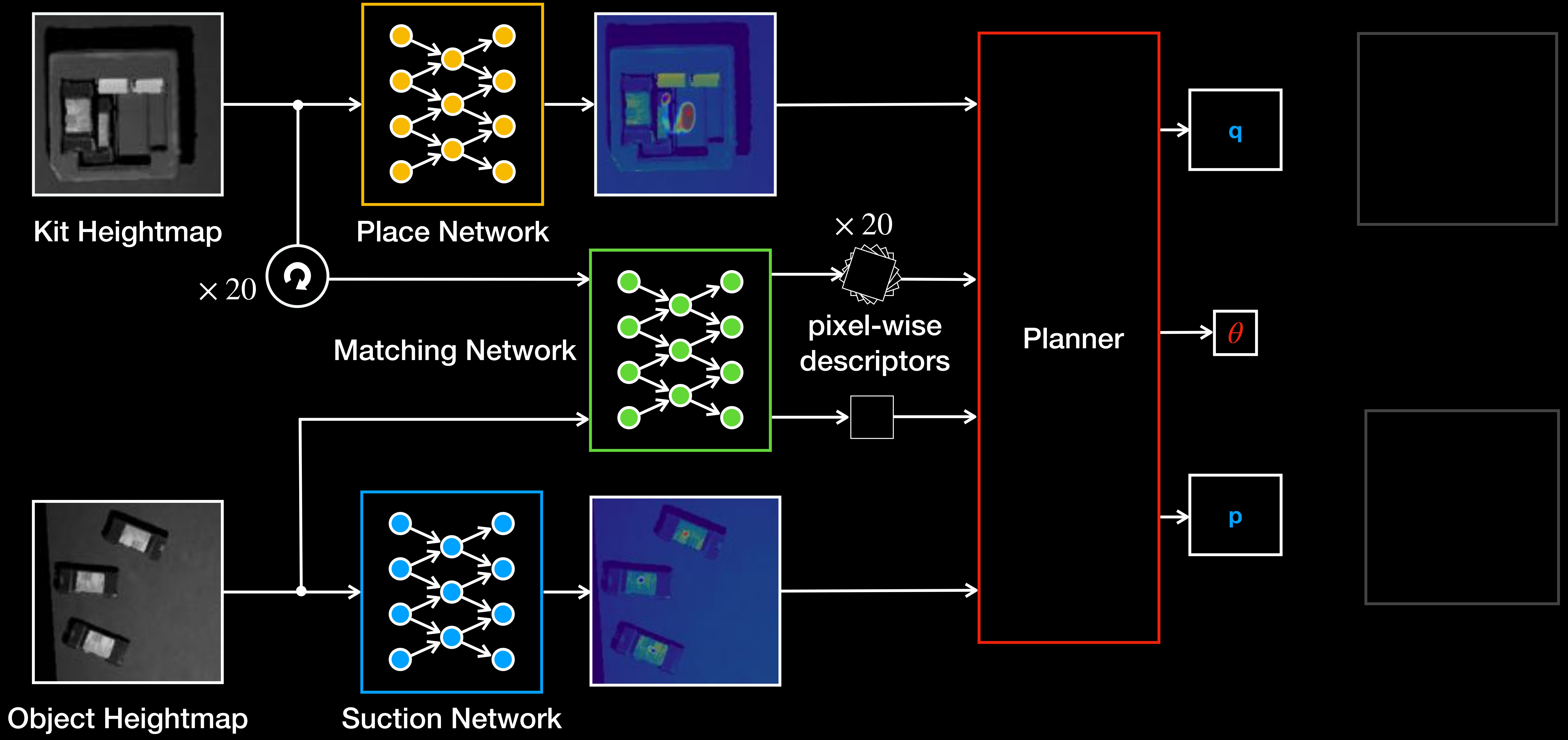

descriptors are rotation-sensitive

# Overview of Form2Fit



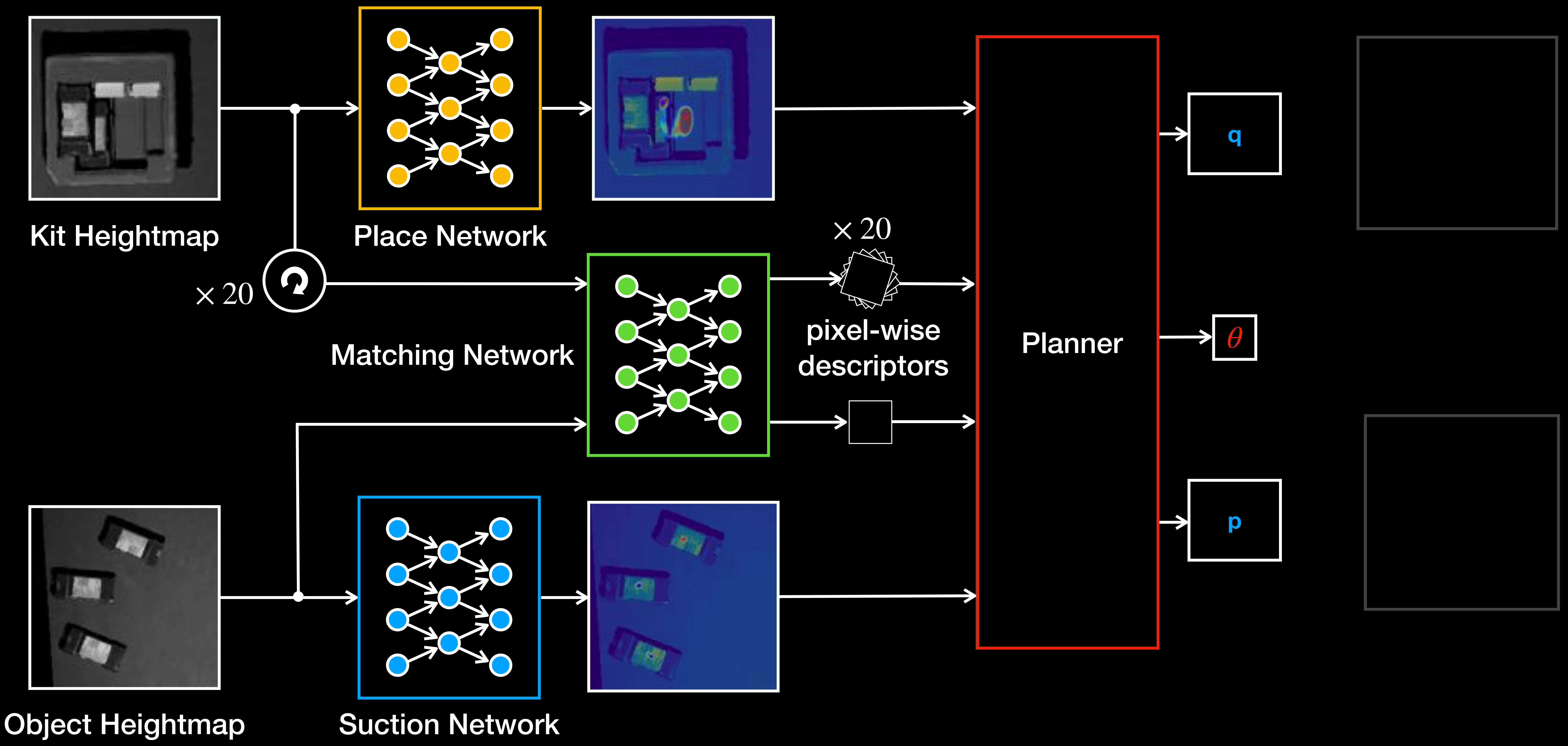

planner integrates information to produce suction/place poses & end-effector rotation

# Overview of Form2Fit



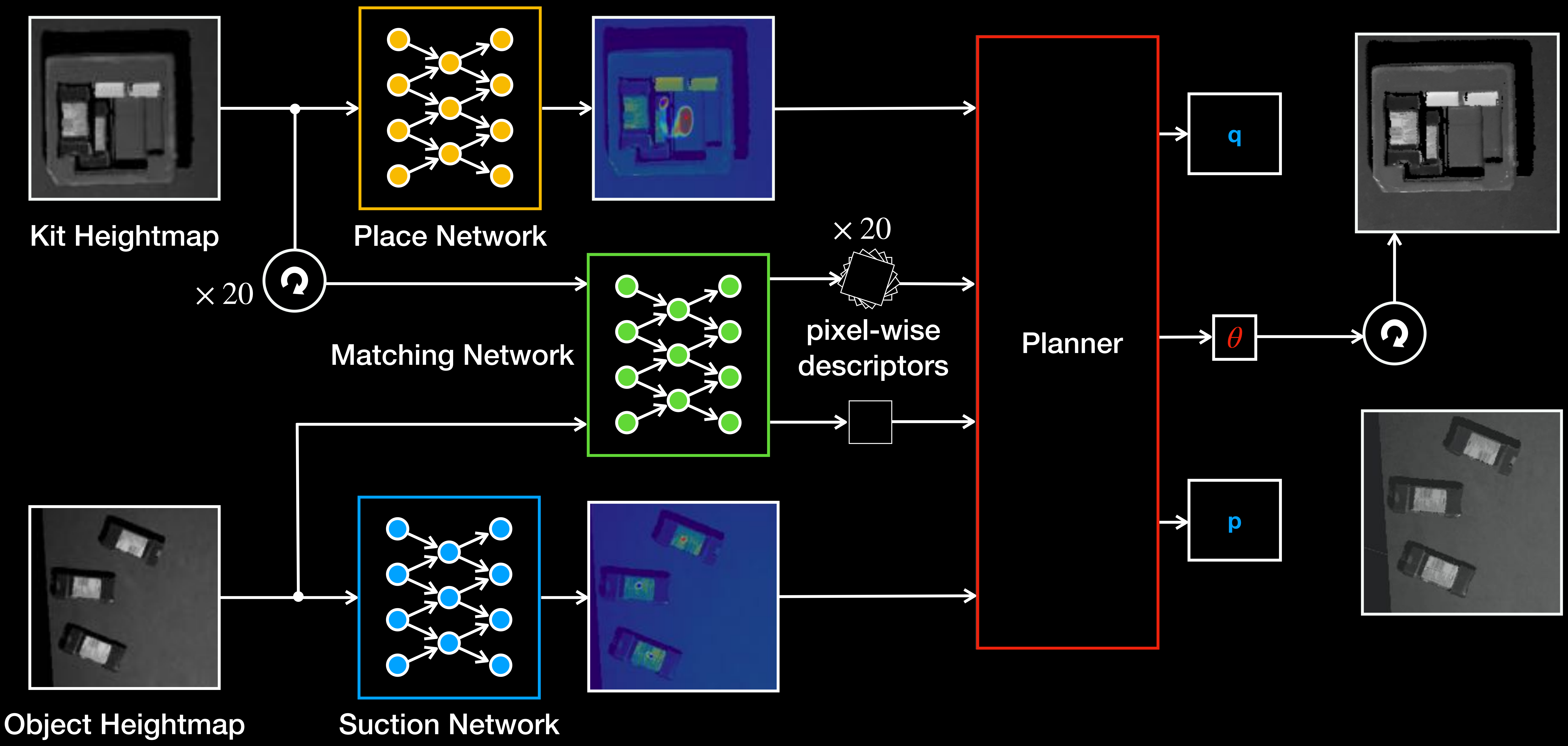

planner integrates information to produce suction/place poses & end-effector rotation

# Overview of Form2Fit



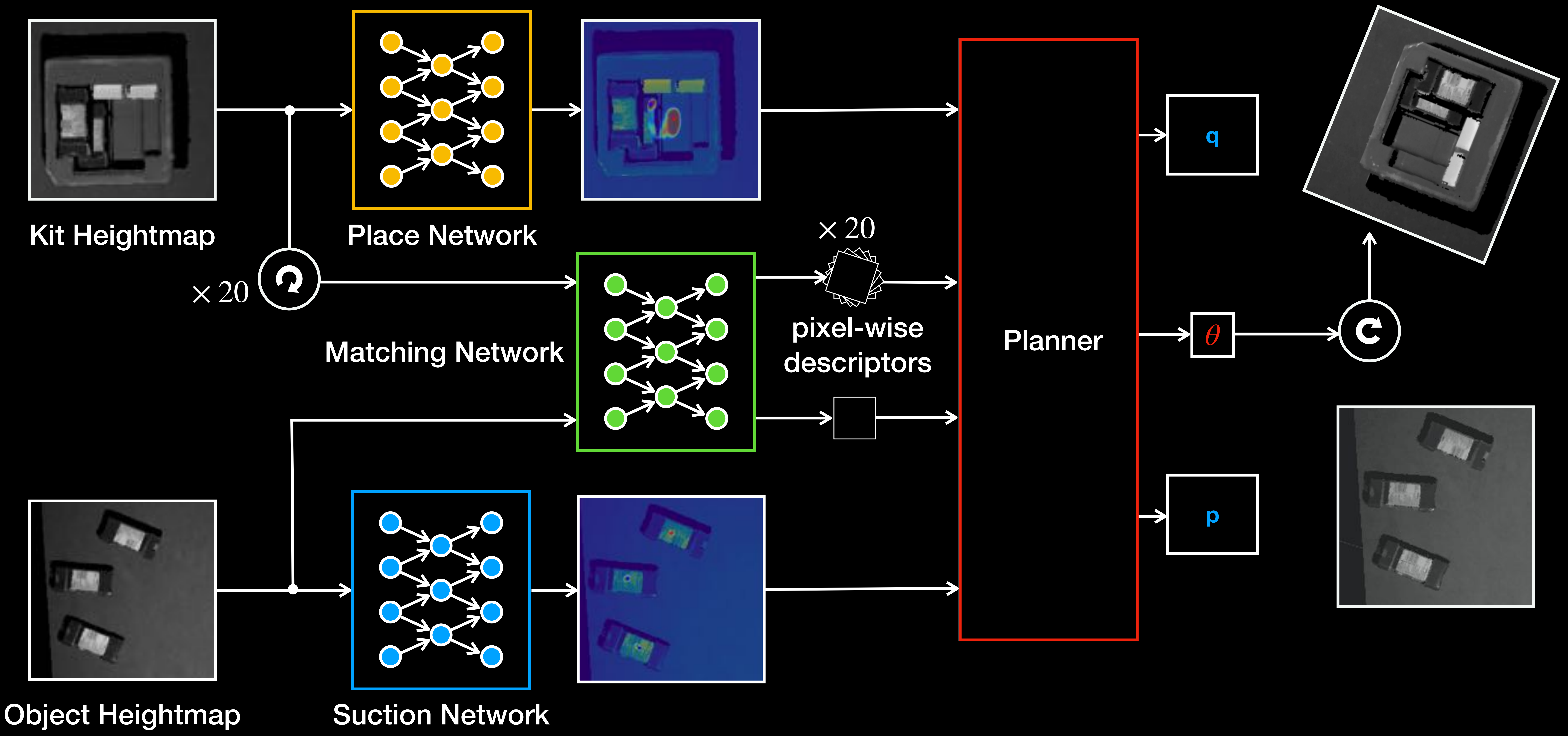

planner integrates information to produce suction/place poses & end-effector rotation

# Overview of Form2Fit



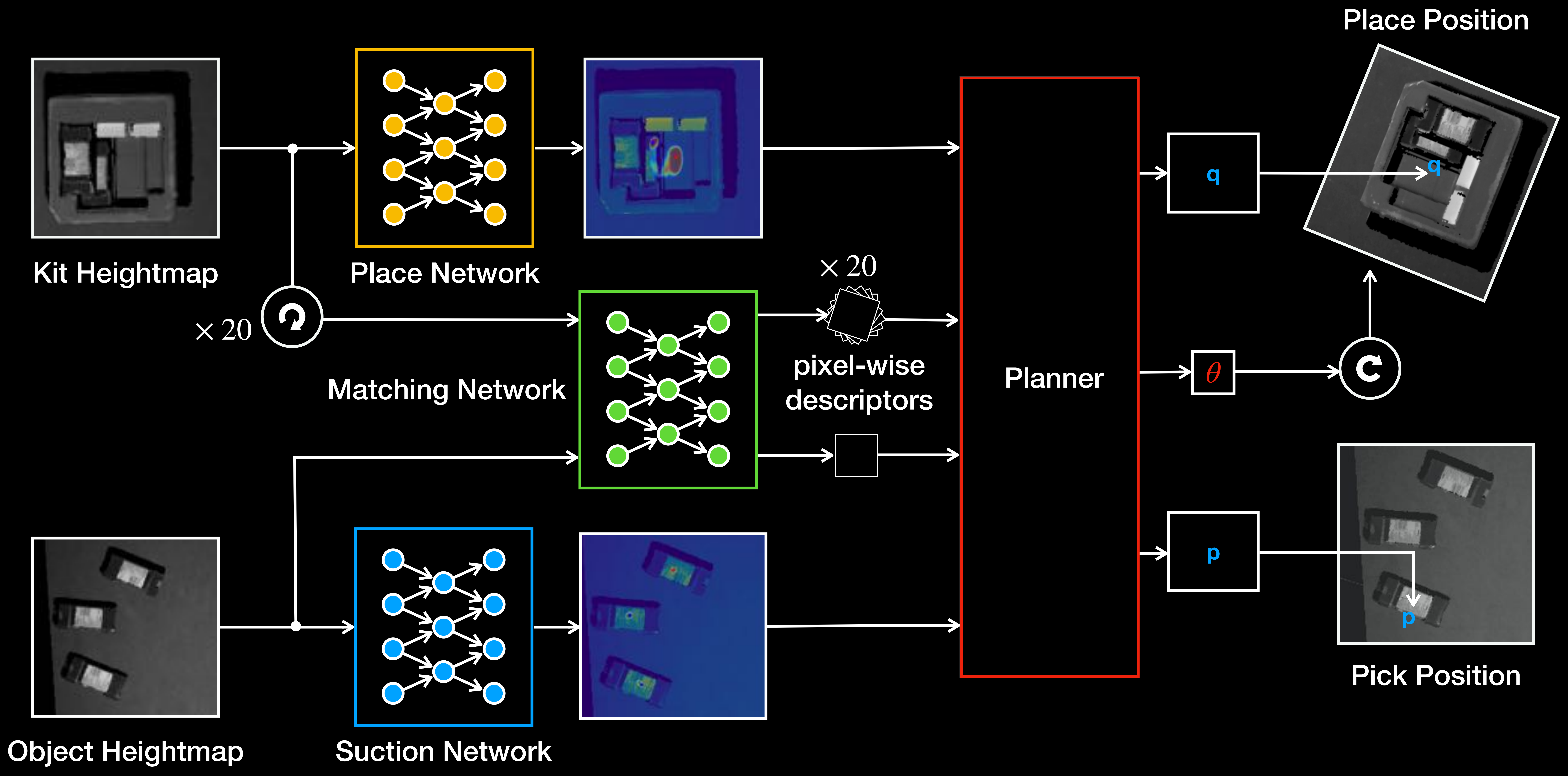

planner integrates information to produce suction/place poses & end-effector rotation

# Overview of Form2Fit



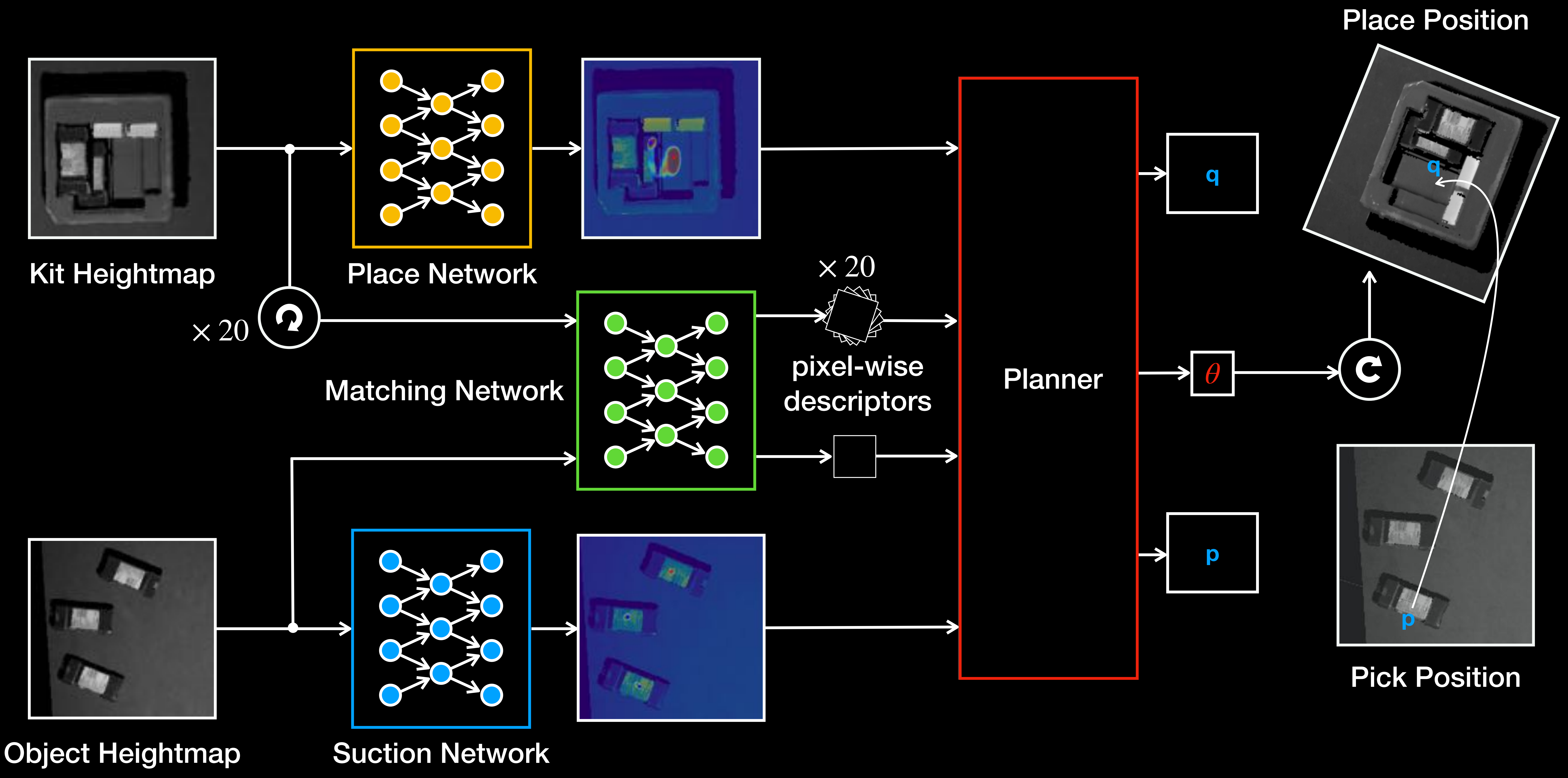

planner integrates information to produce suction/place poses & end-effector rotation

# Overview of Form2Fit




planner integrates information to produce suction/place poses & end-effector rotation

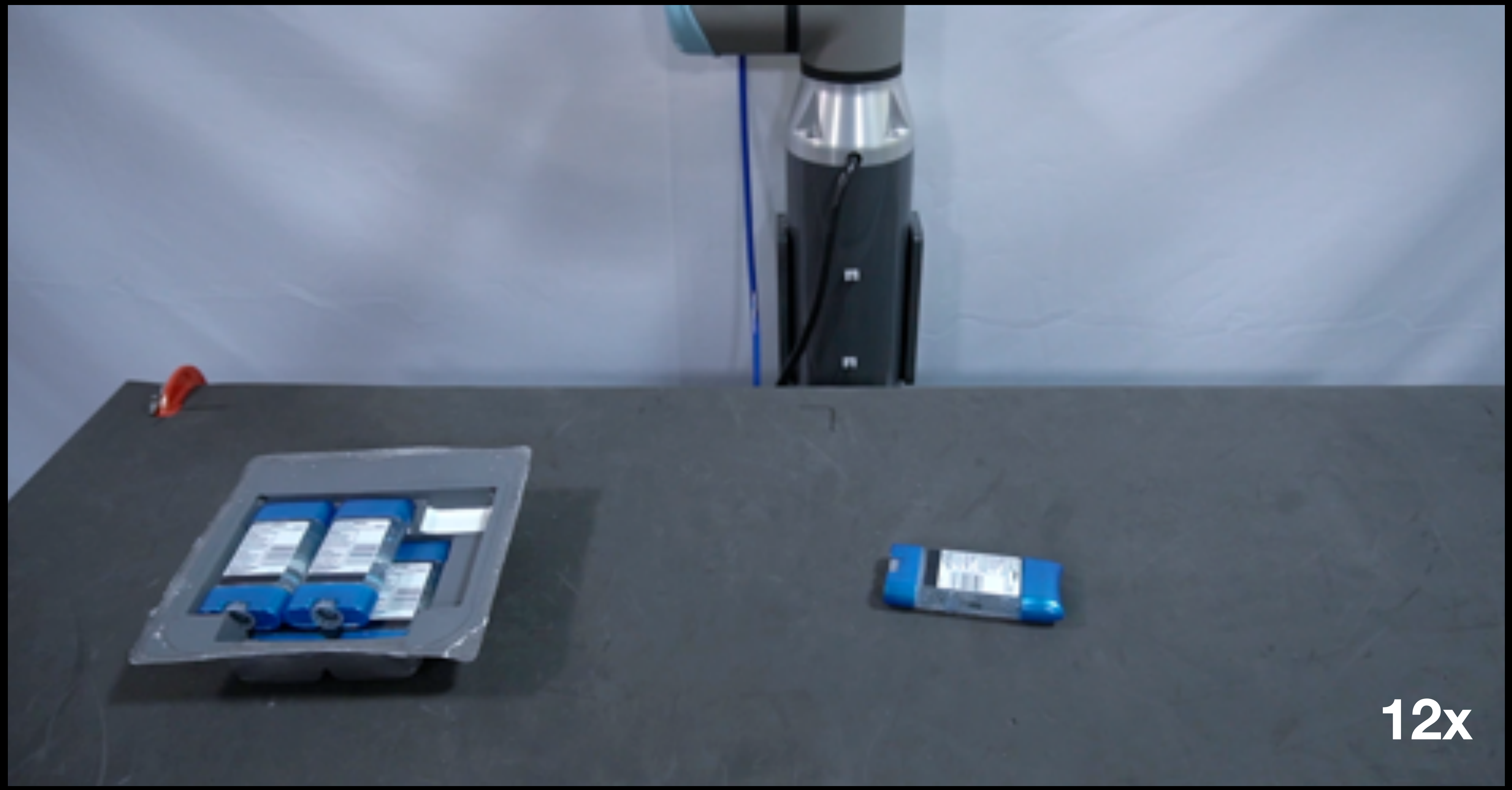
# Overview of Form2Fit



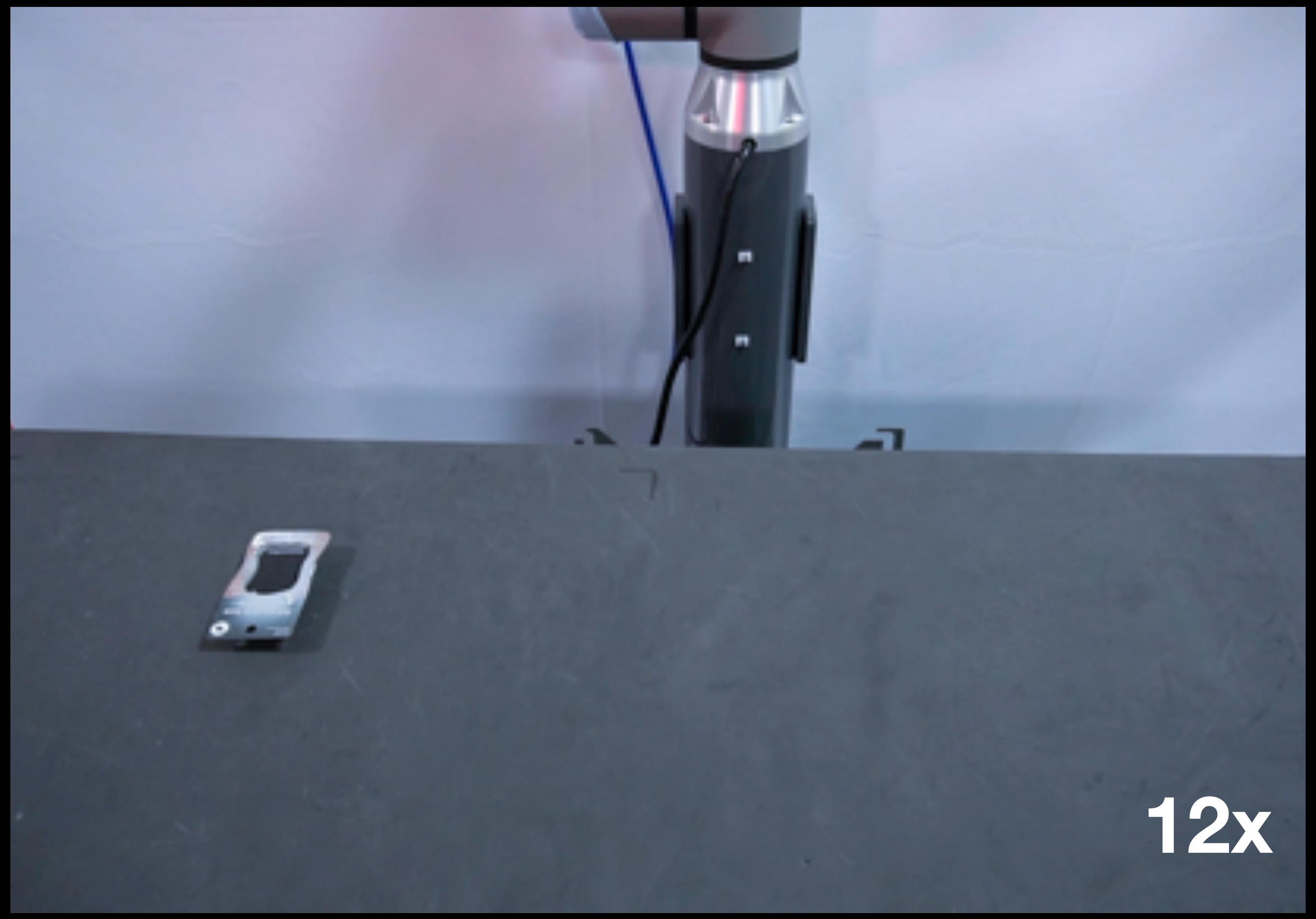

planner integrates information to produce suction/place poses & end-effector rotation

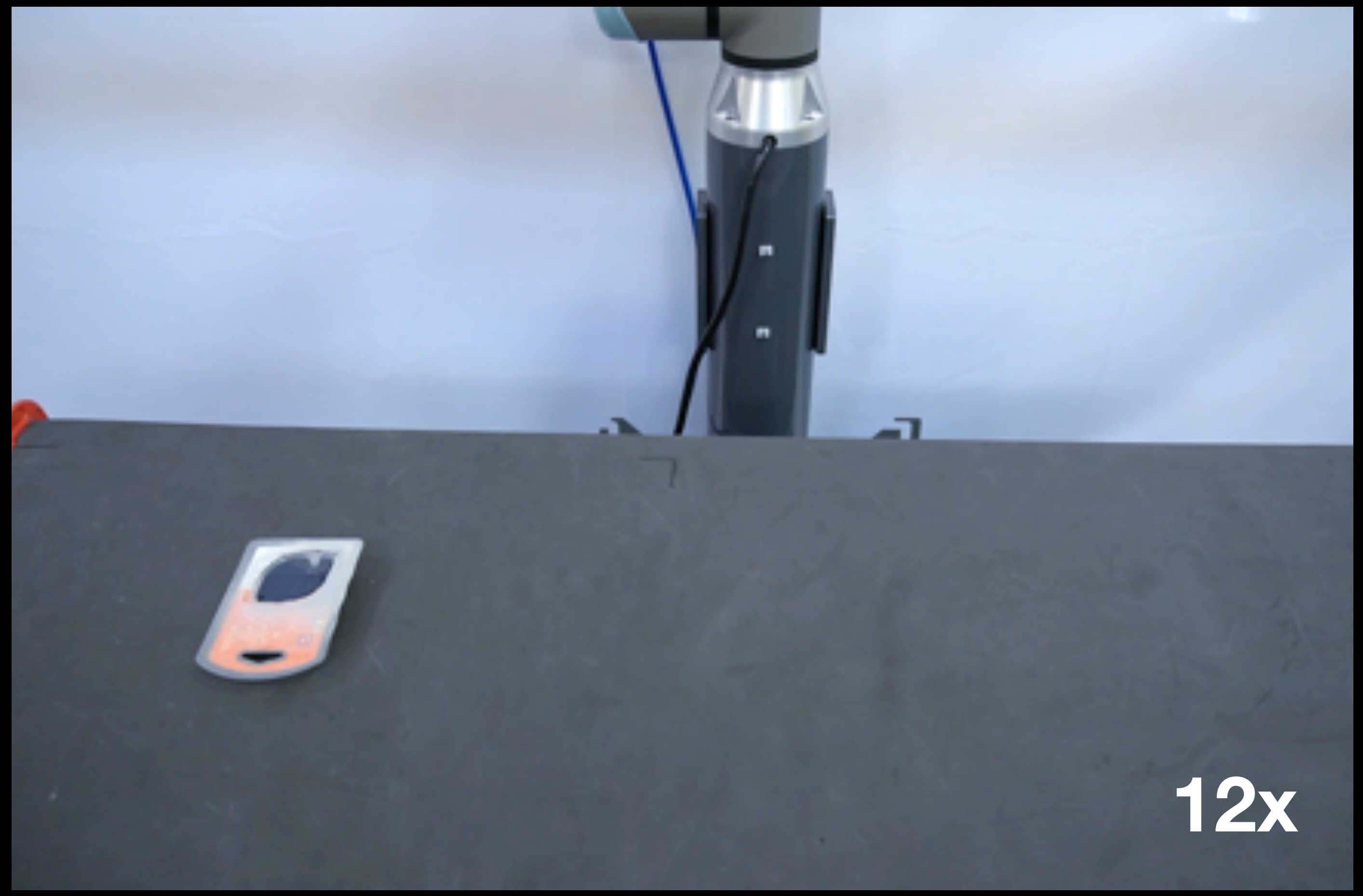
# Data Collection

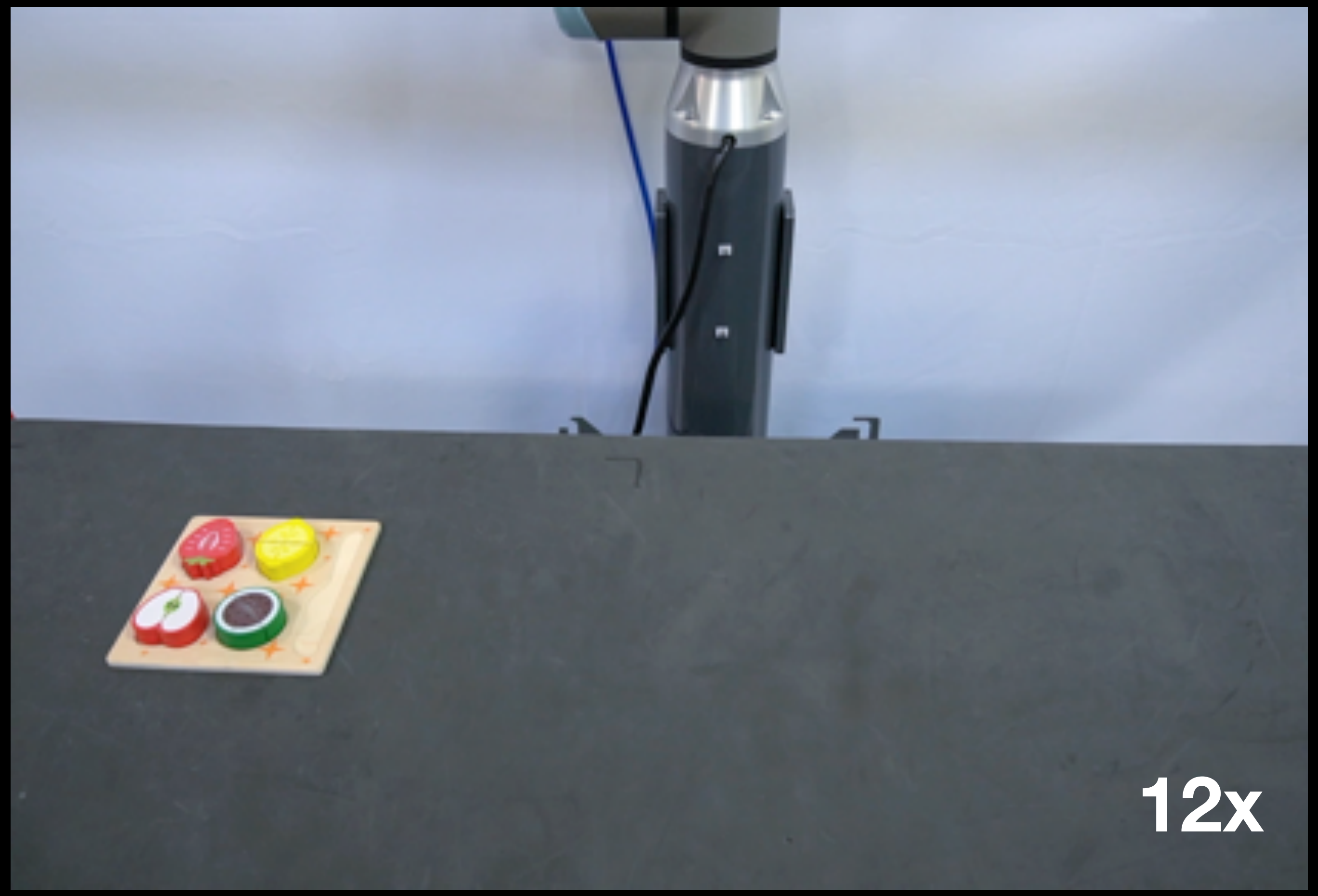
# Data Collection





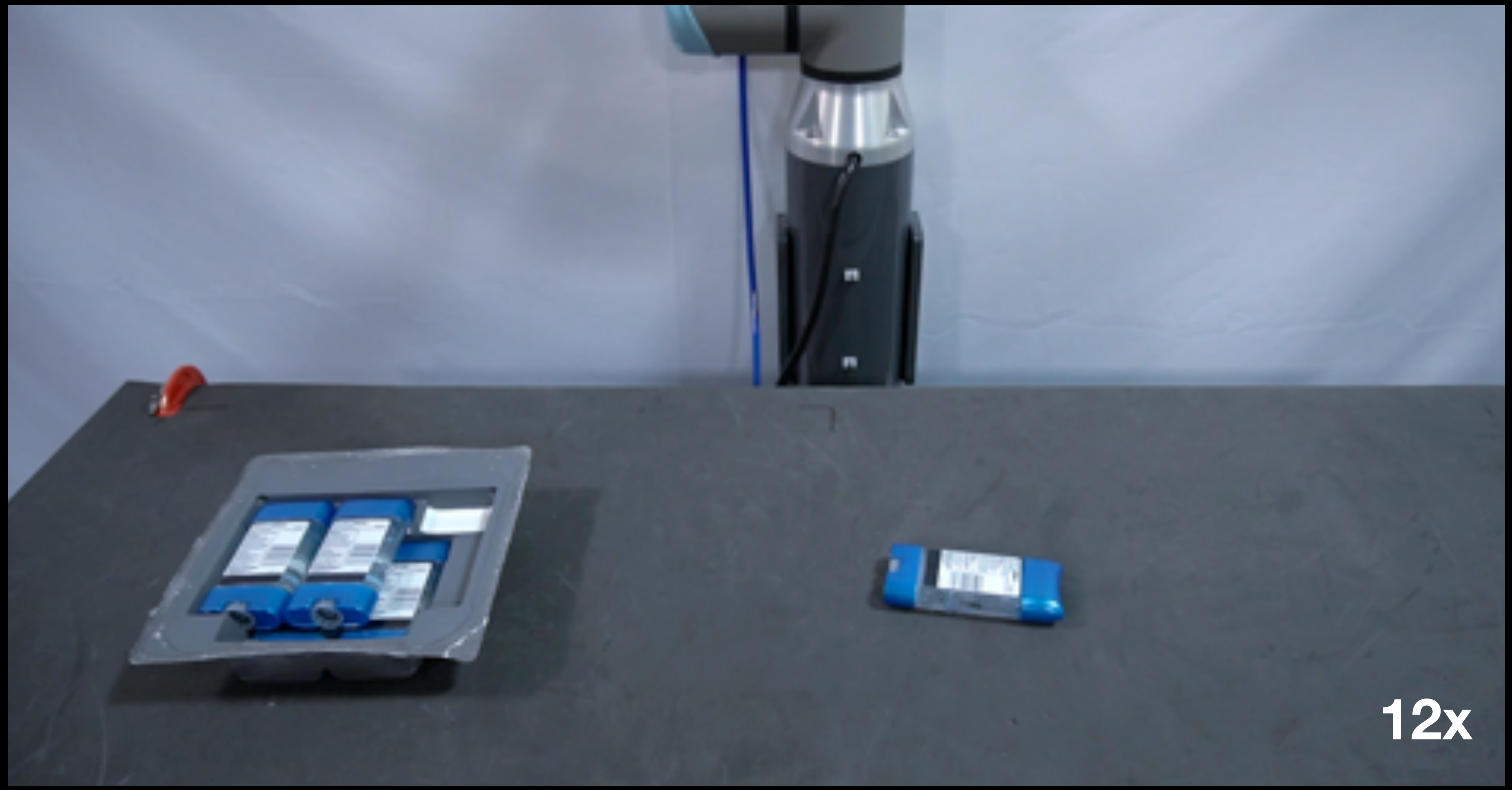


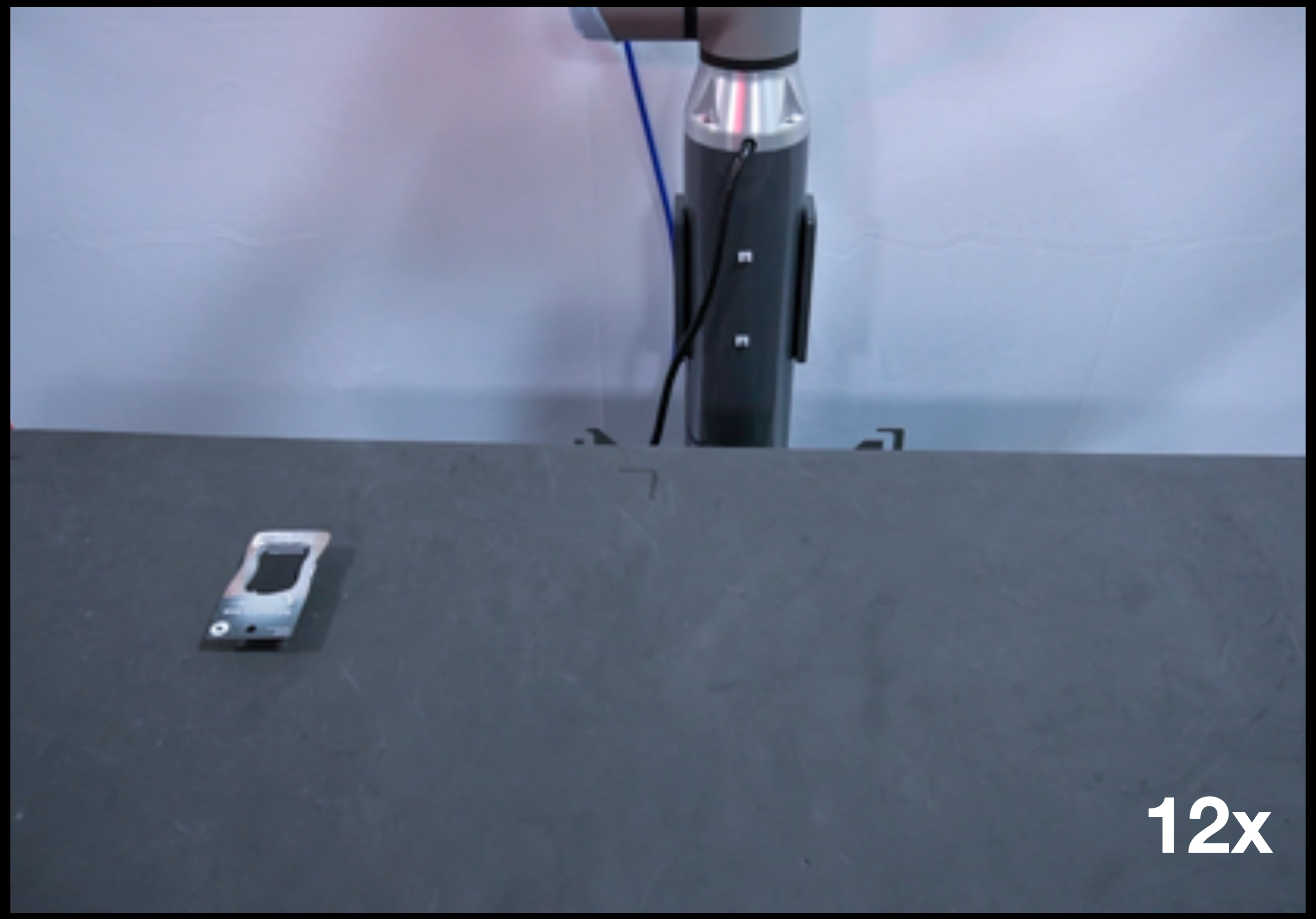



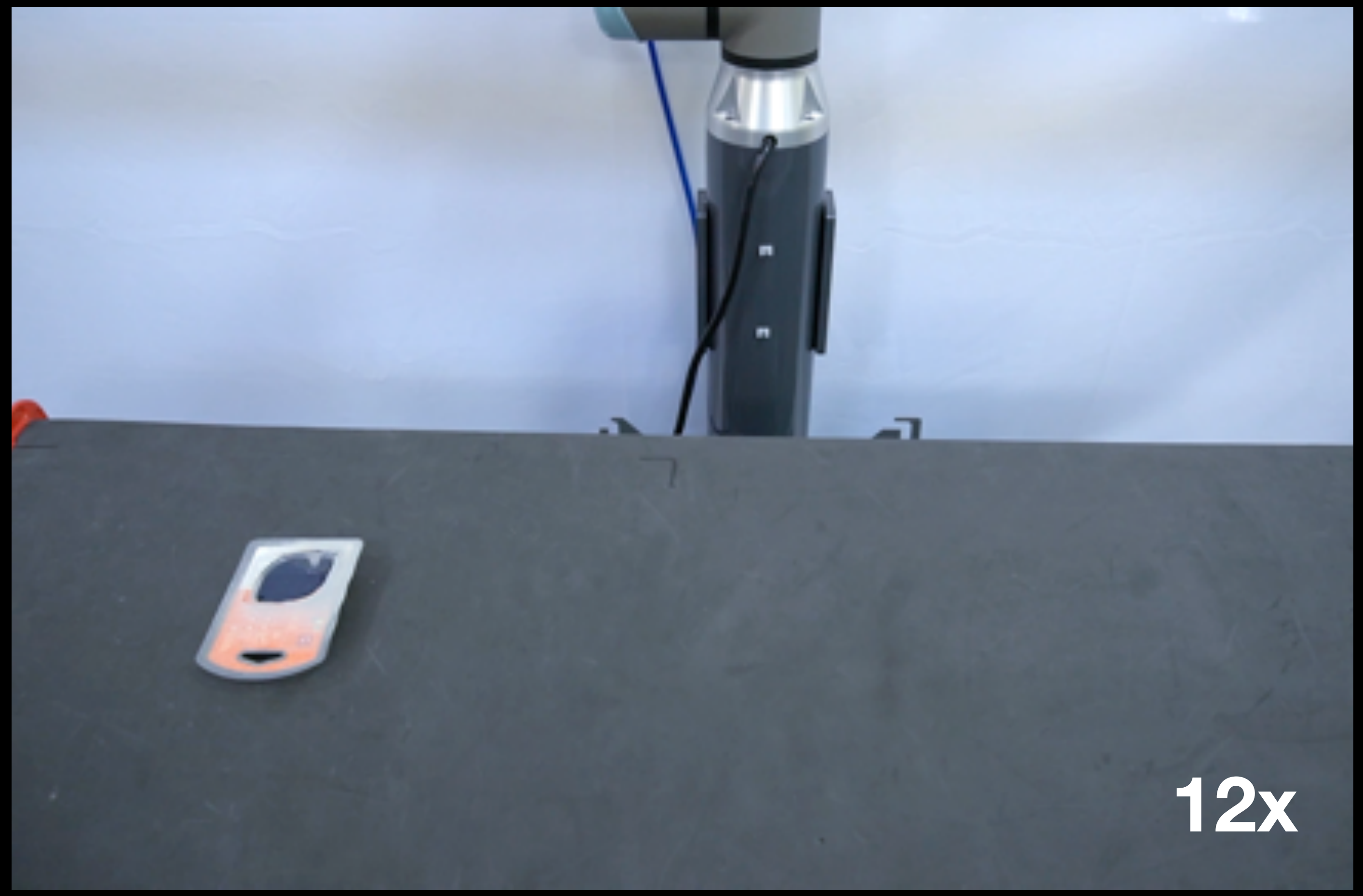

500 disassembly sequence (~ 8 to 10 hours) for each kit

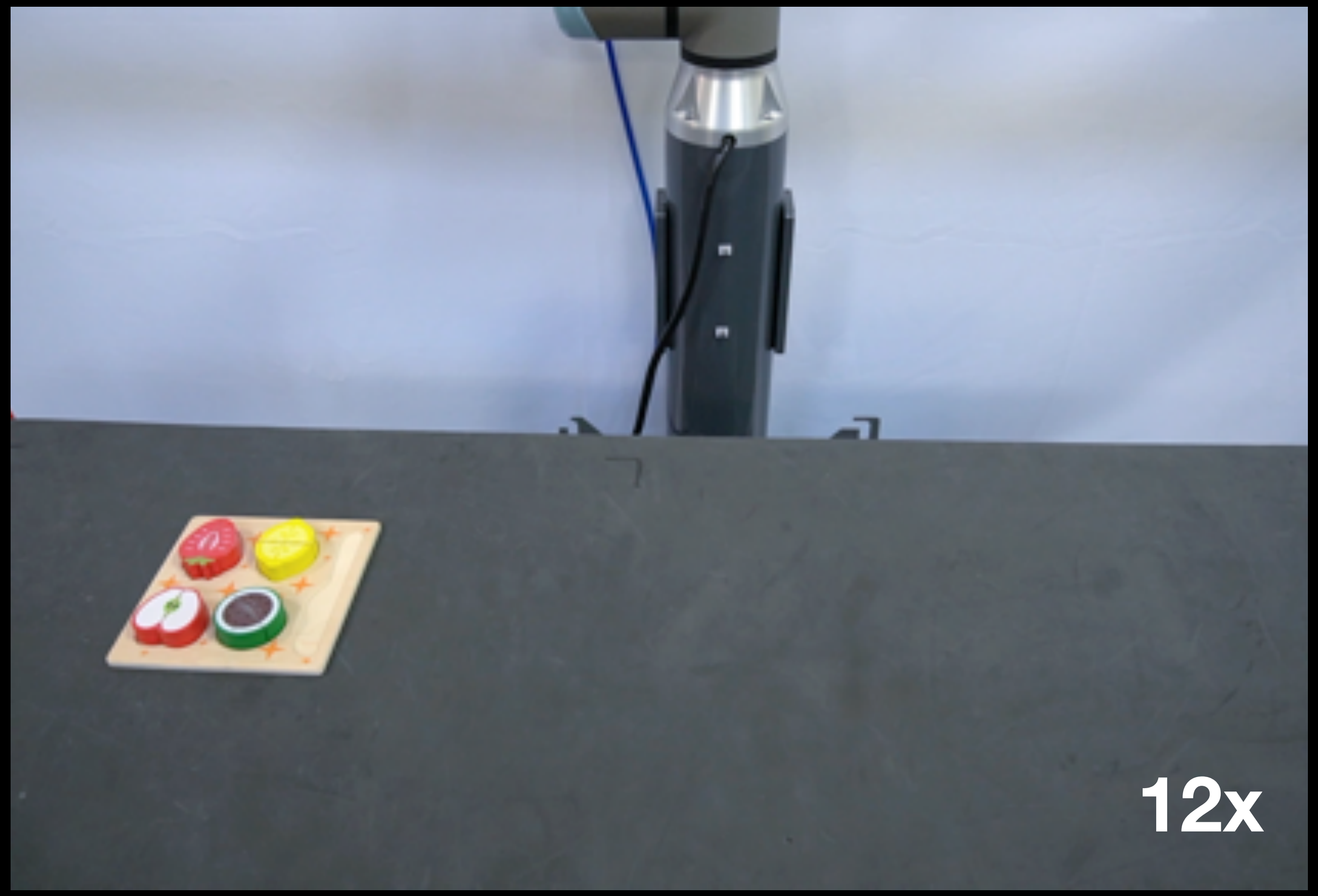
# Data Collection





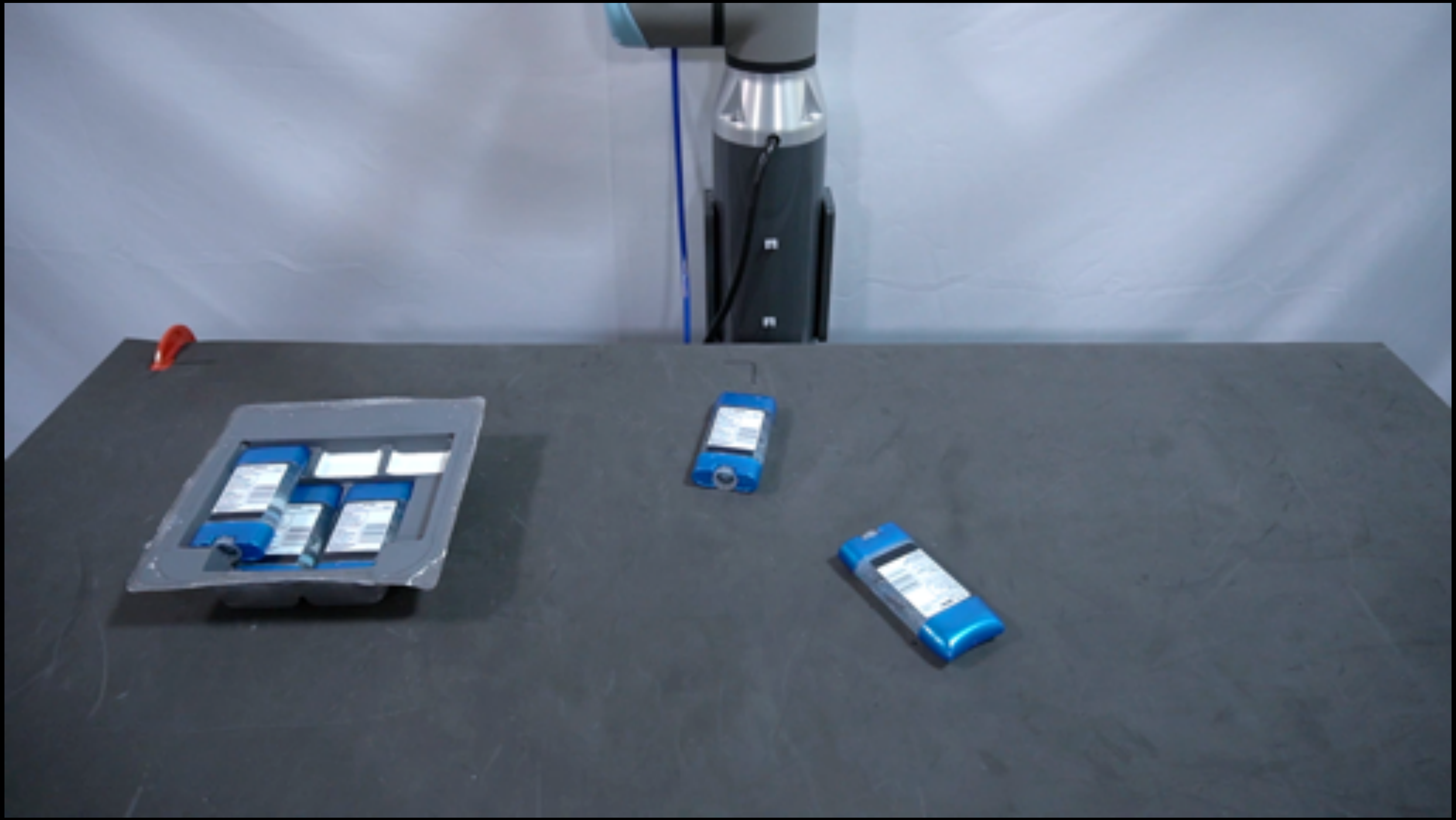


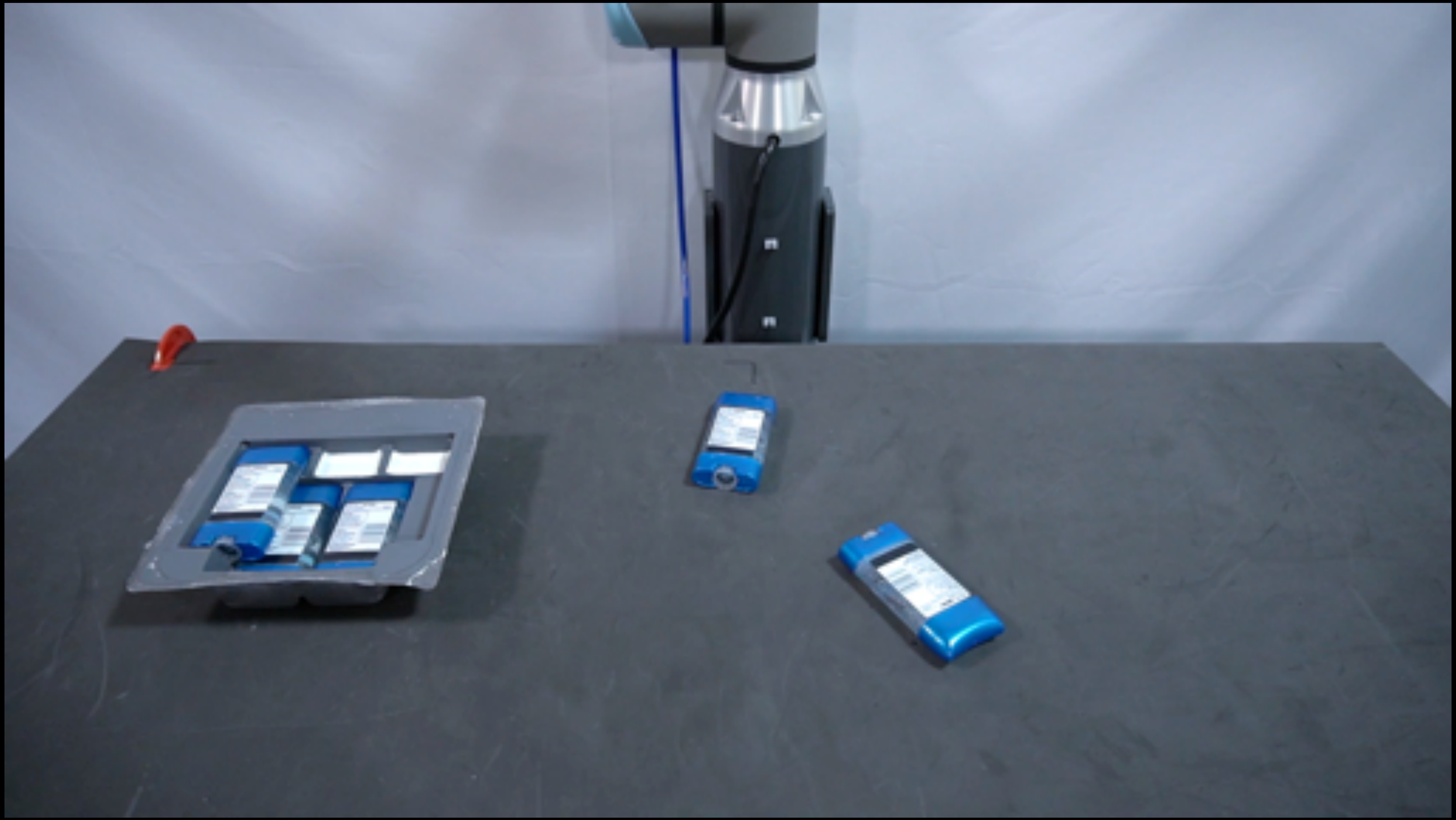




500 disassembly sequence (~ 8 to 10 hours) for each kit

# Data Collection from Disassembly

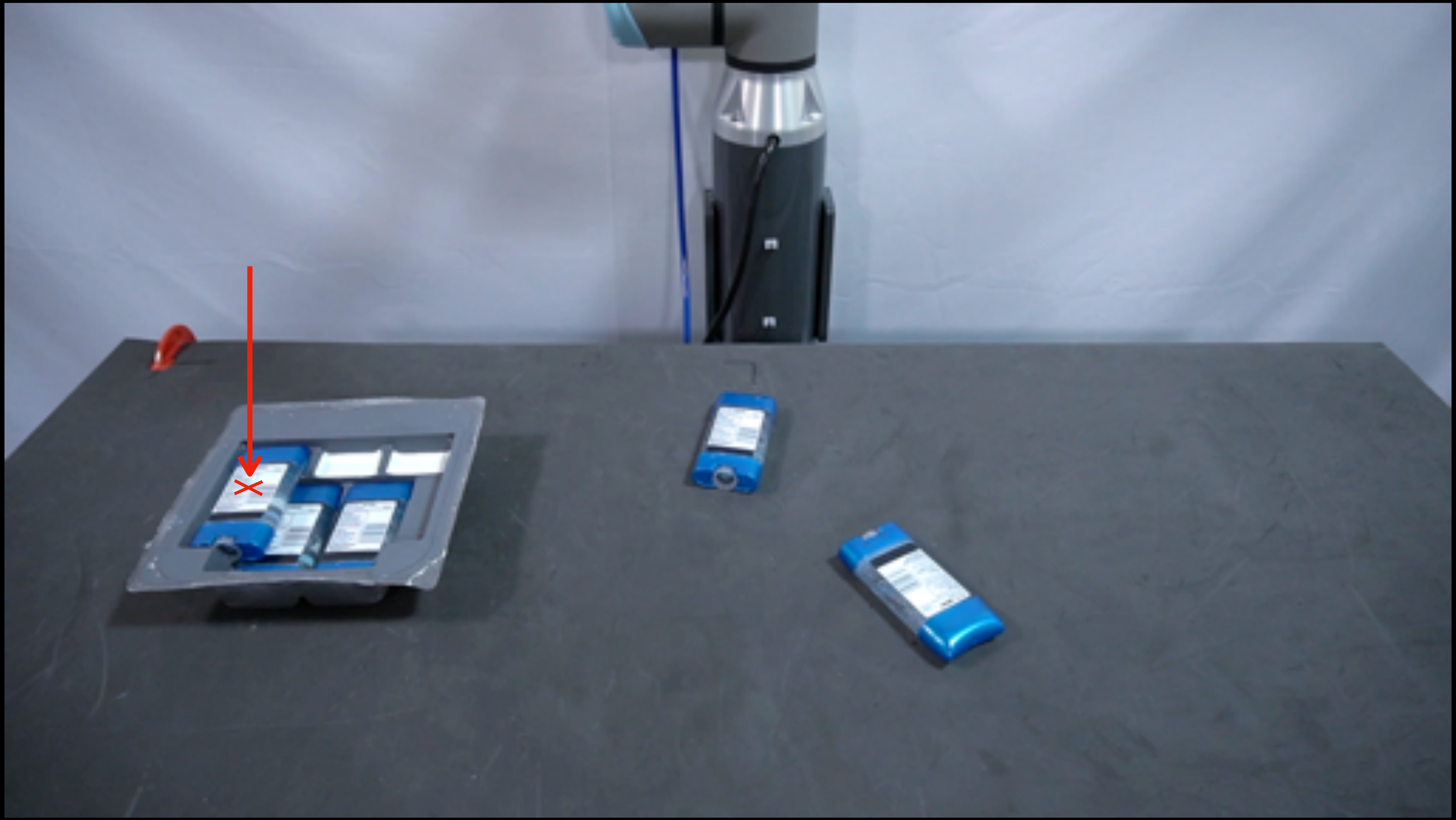


# Data Collection from Disassembly



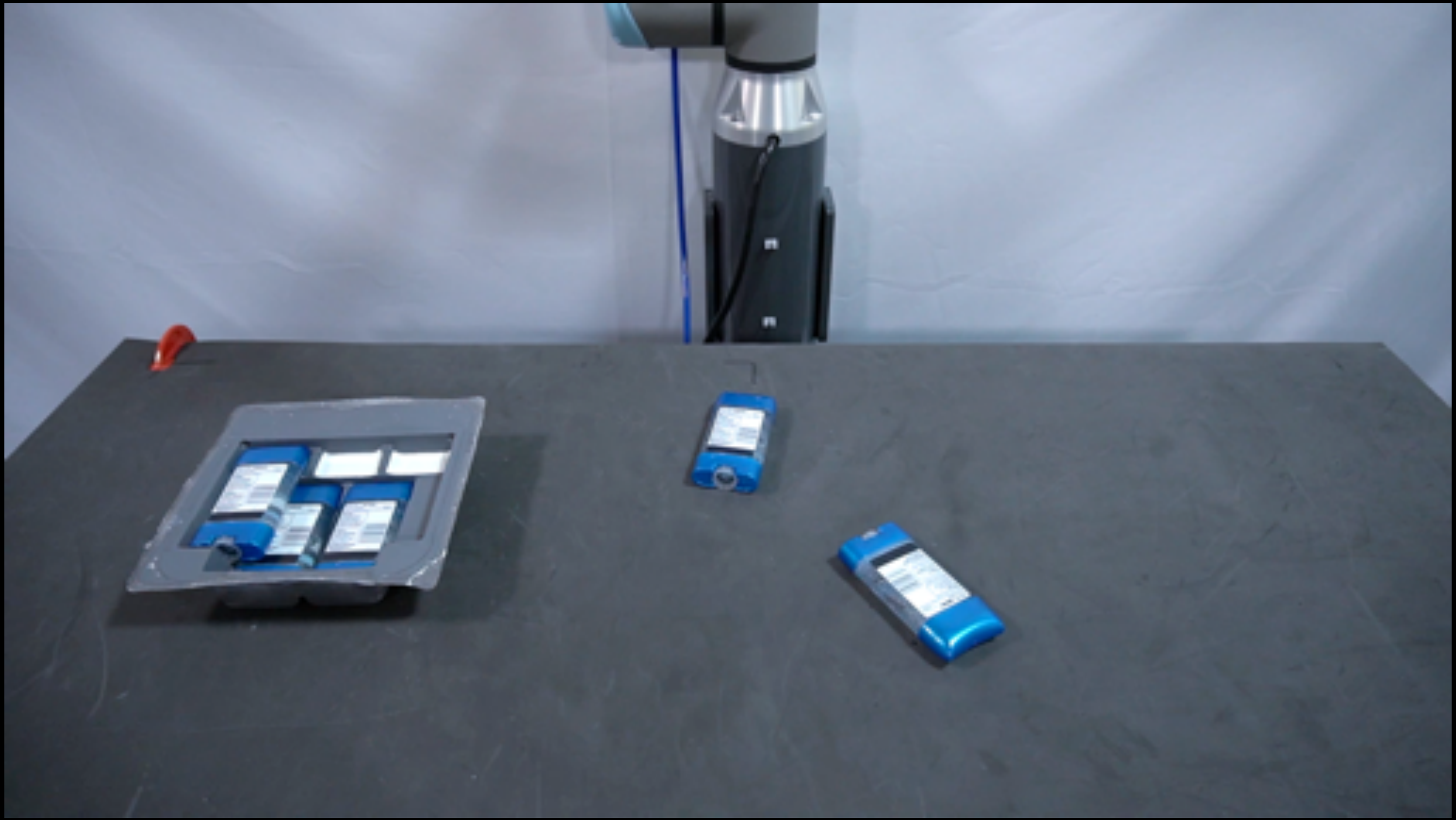
suction network predicts a suction candidate

# Data Collection from Disassembly



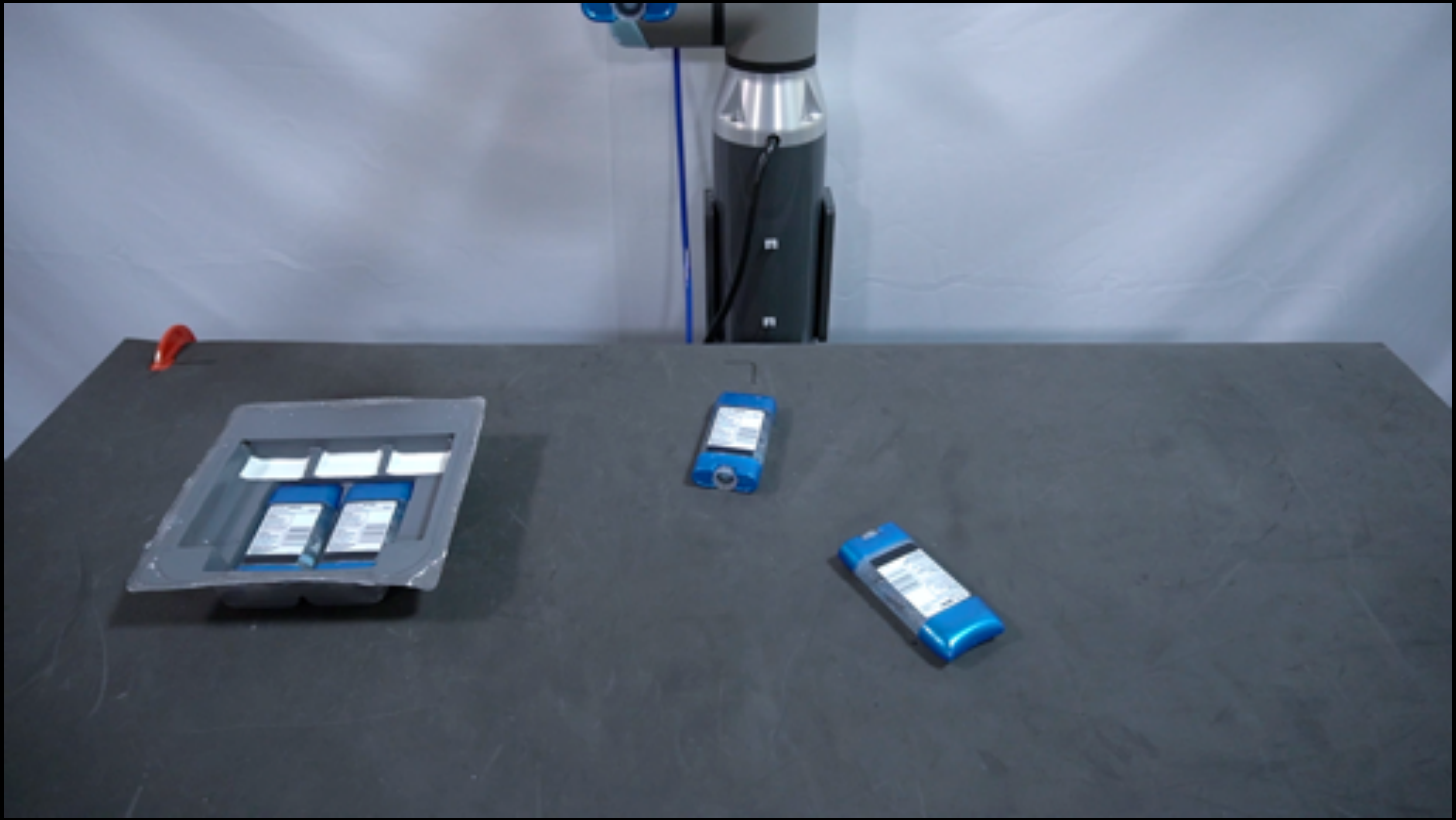
suction network predicts a suction candidate

# Data Collection from Disassembly

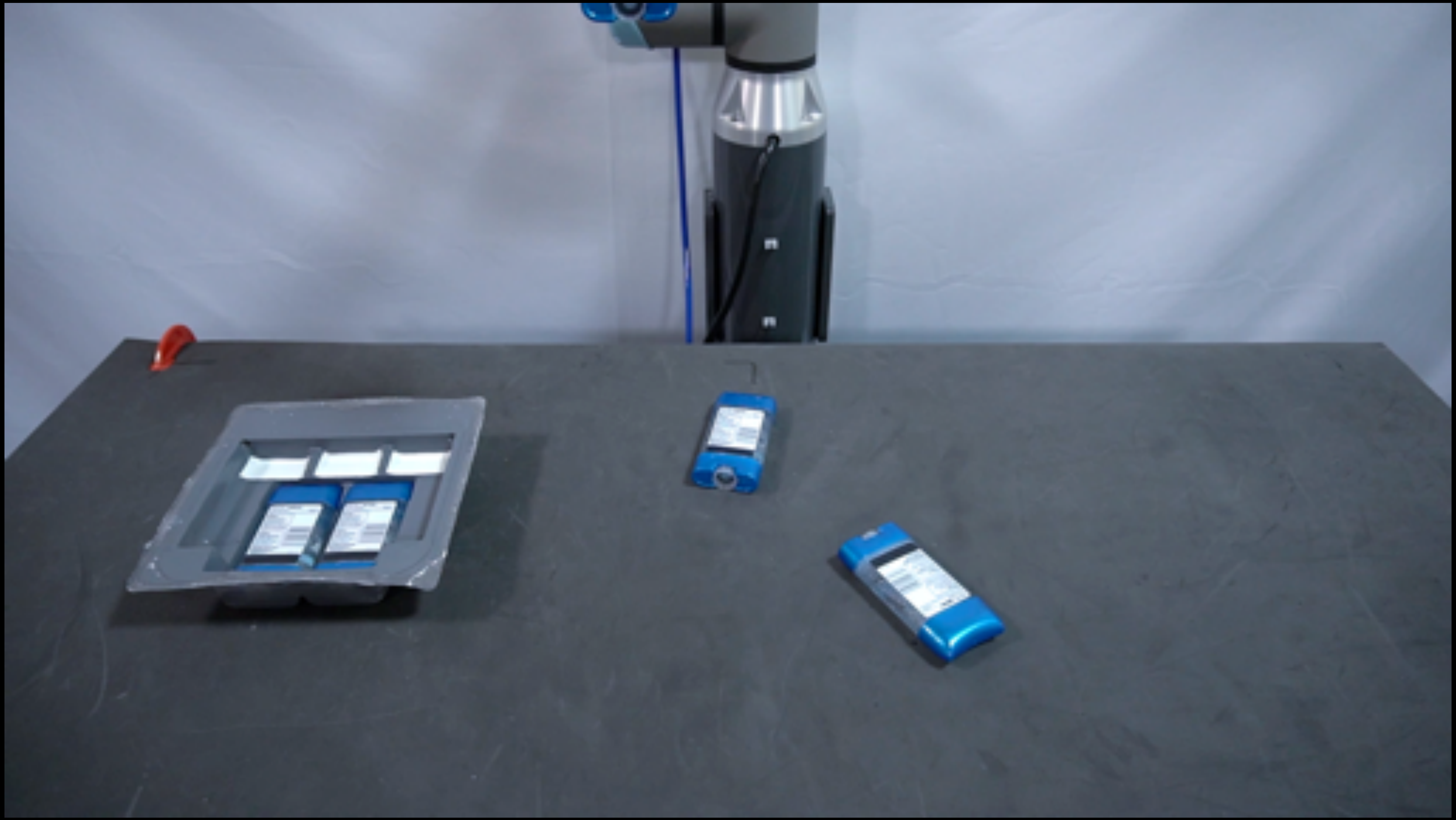


suction network predicts a suction candidate

# Data Collection from Disassembly

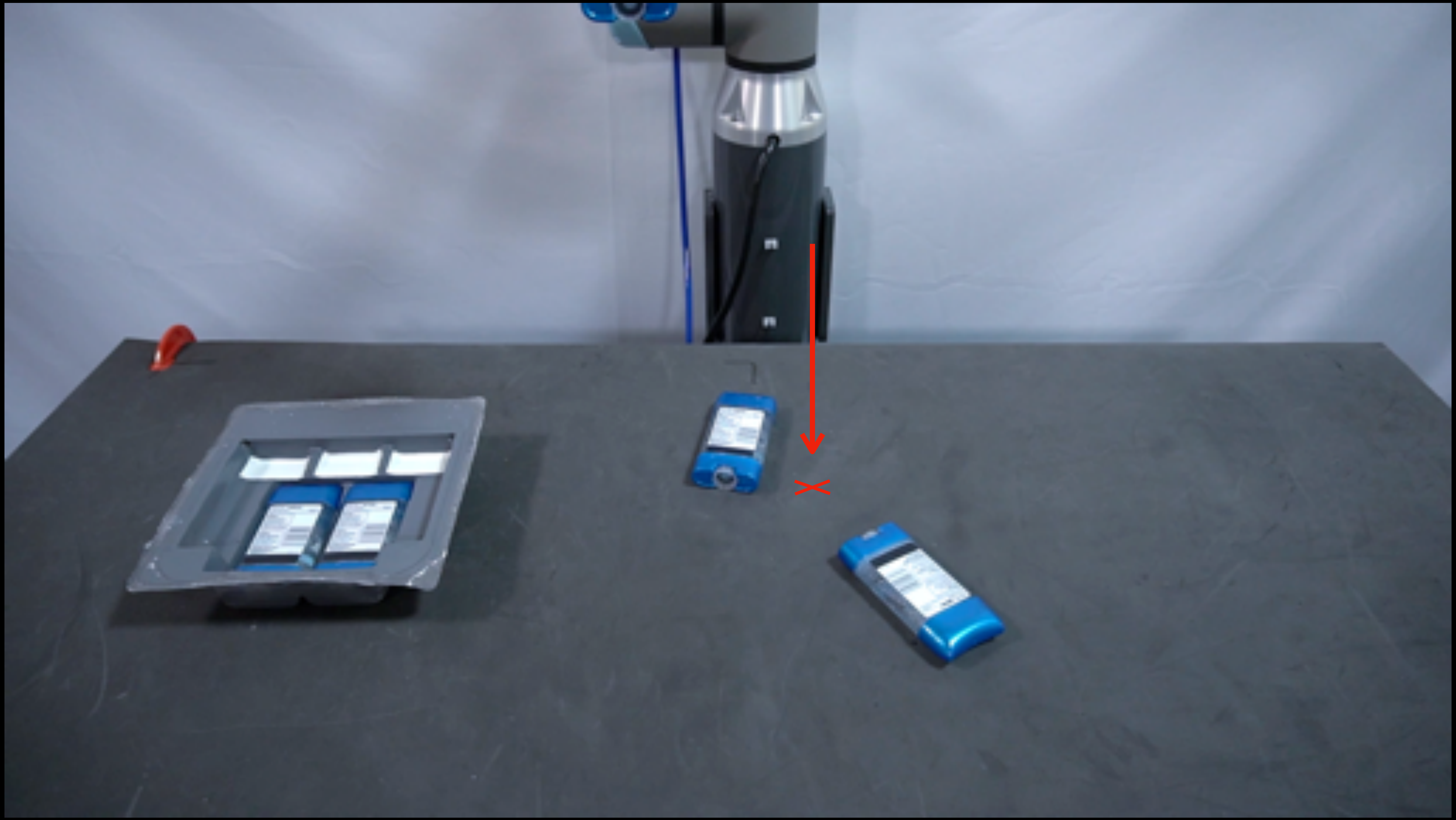


# Data Collection from Disassembly



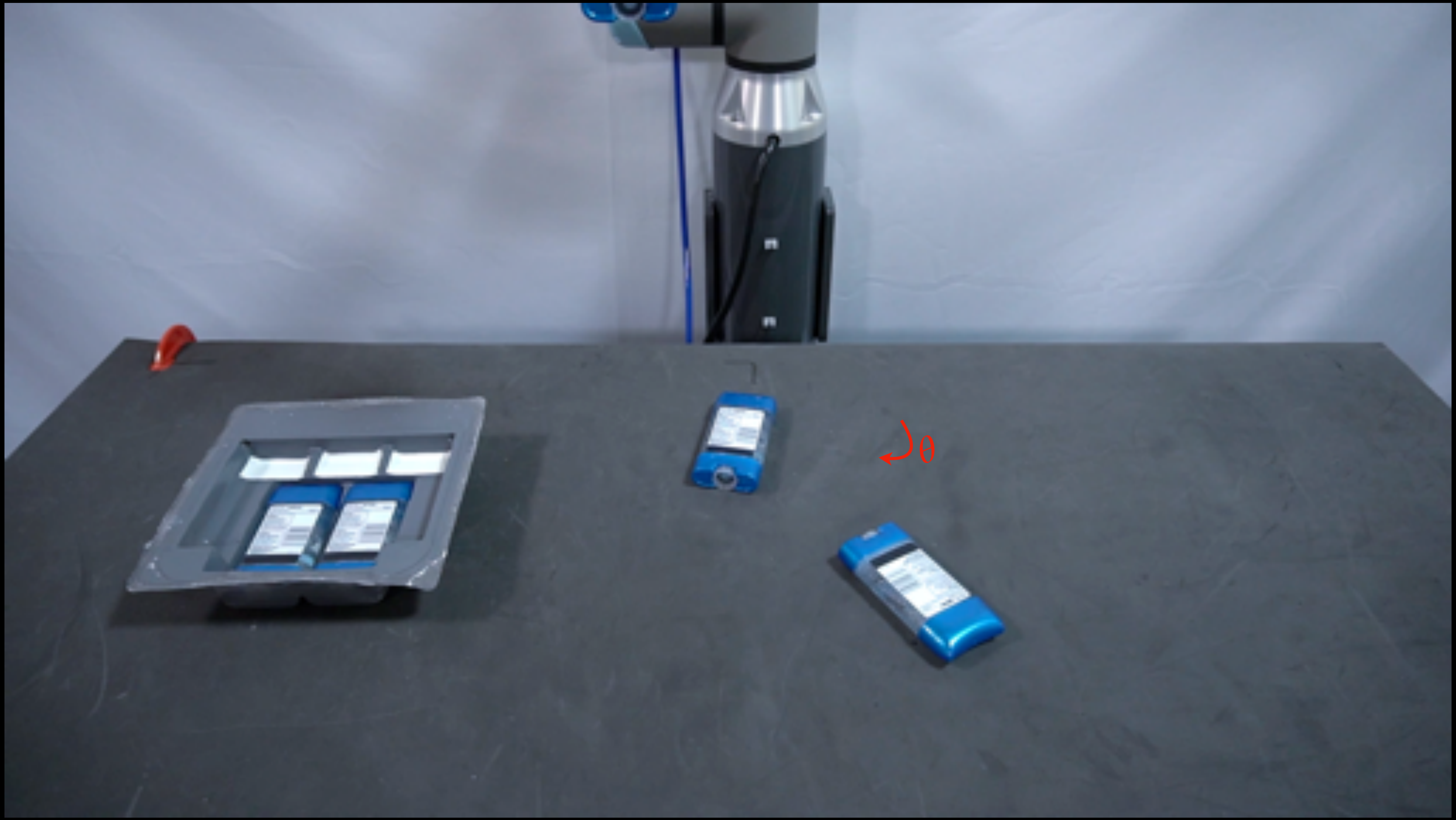
place pose randomly generated $(q, \theta)$

# Data Collection from Disassembly



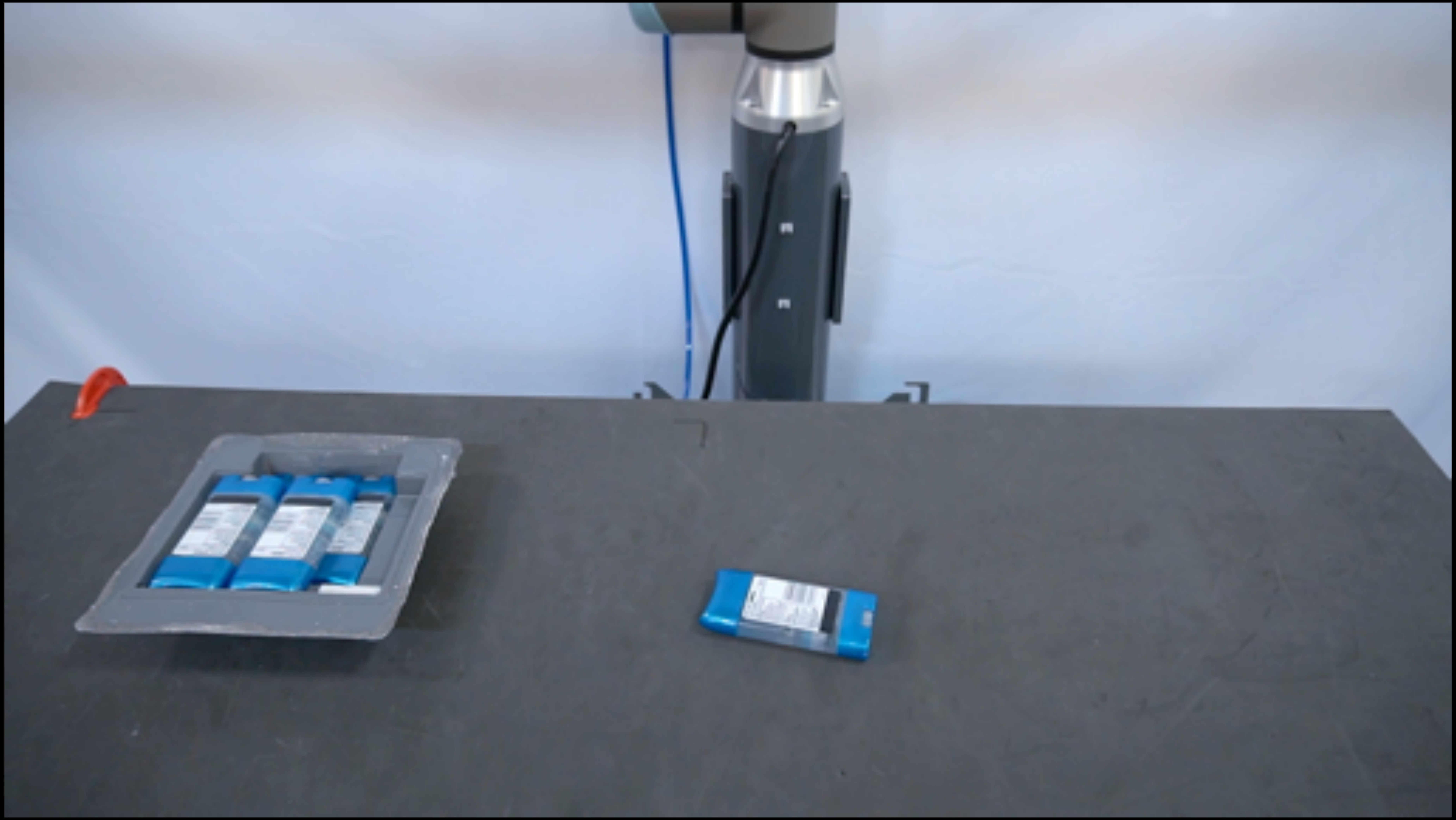
place pose randomly generated $(q, \theta)$

# Data Collection from Disassembly



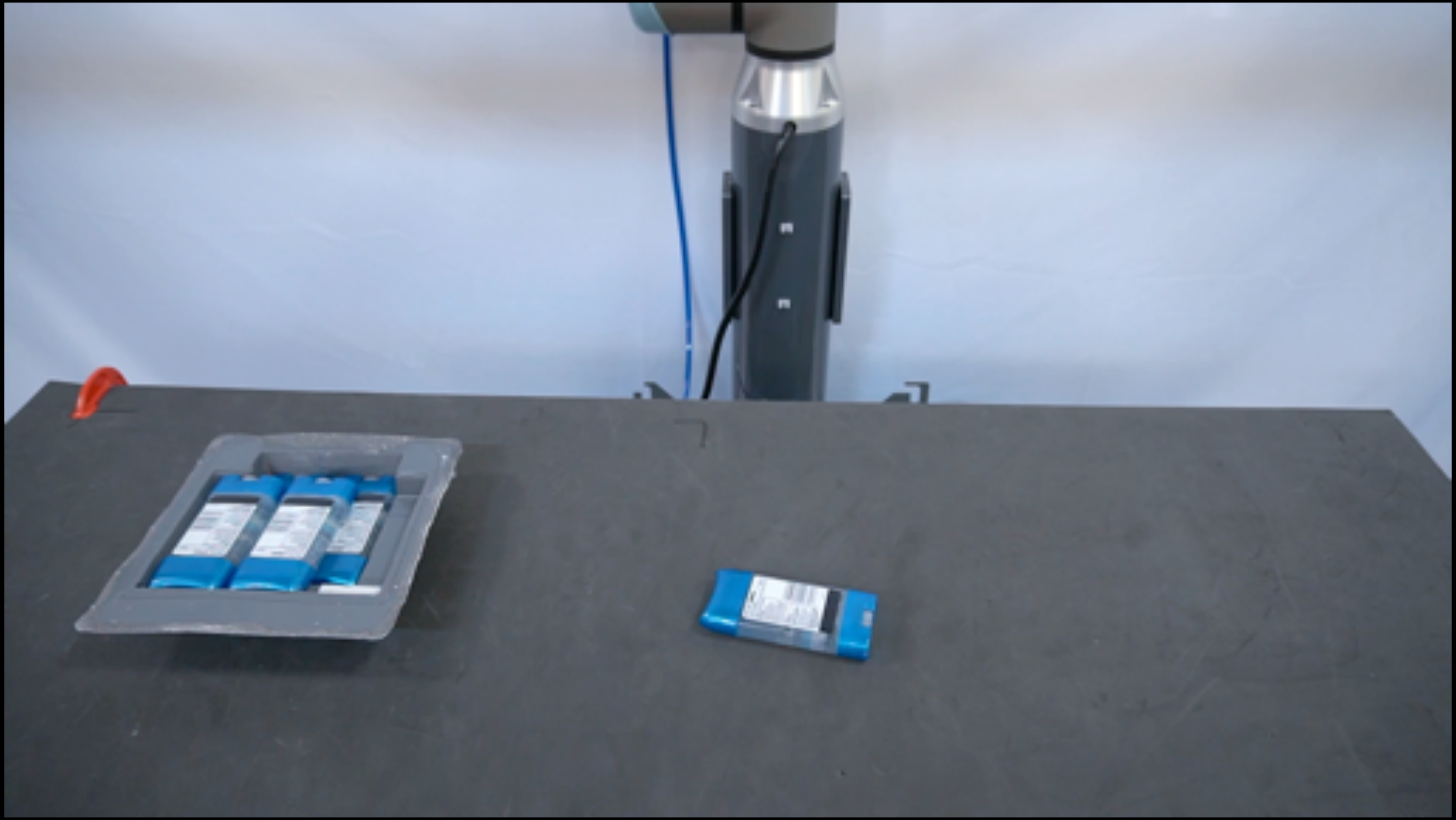

place pose randomly generated $(q, \theta)$

# Data Collection from Disassembly



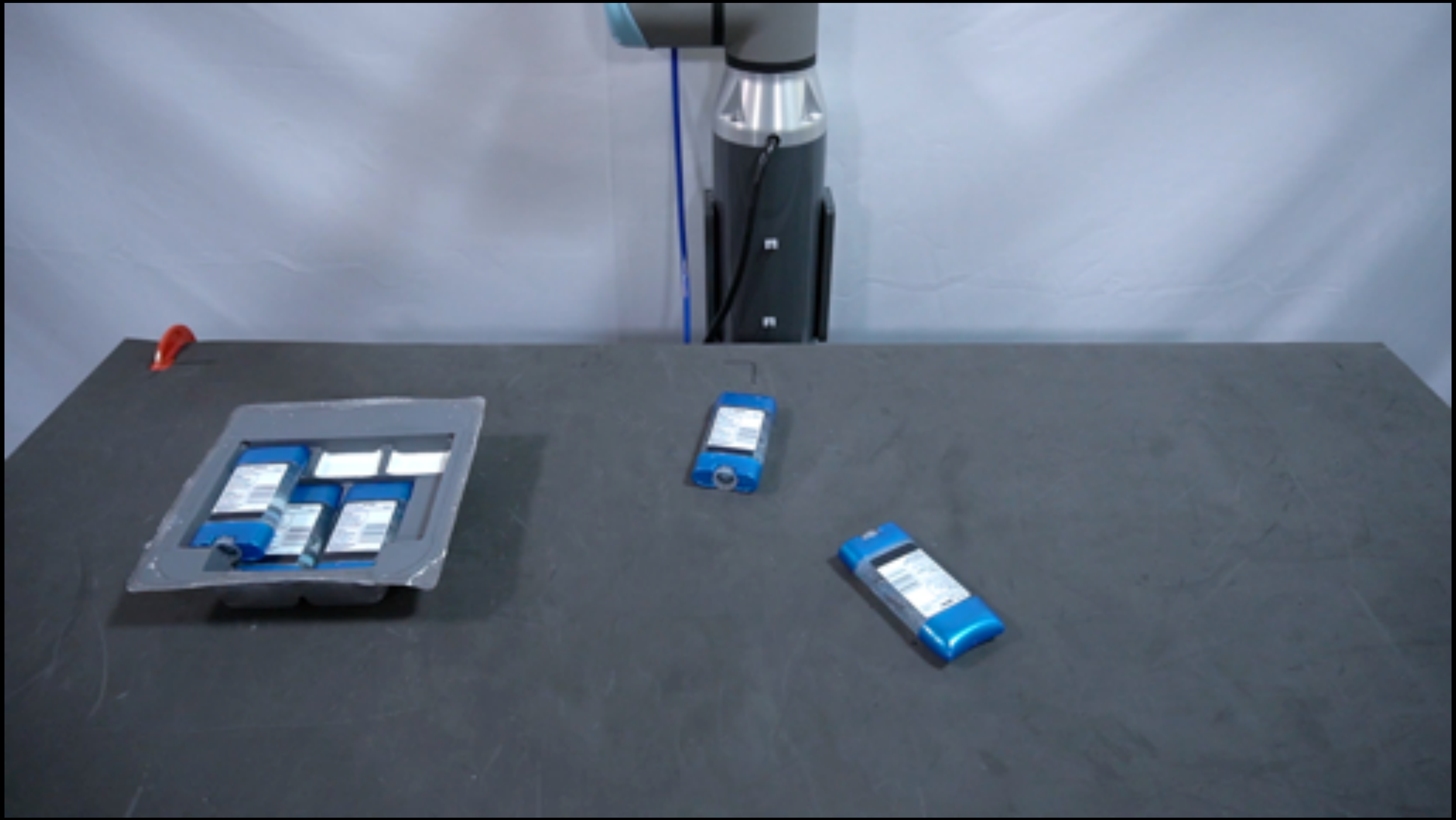
kit is secured to table to prevent accidental displacement from bad suction grasps

# Data Collection from Disassembly



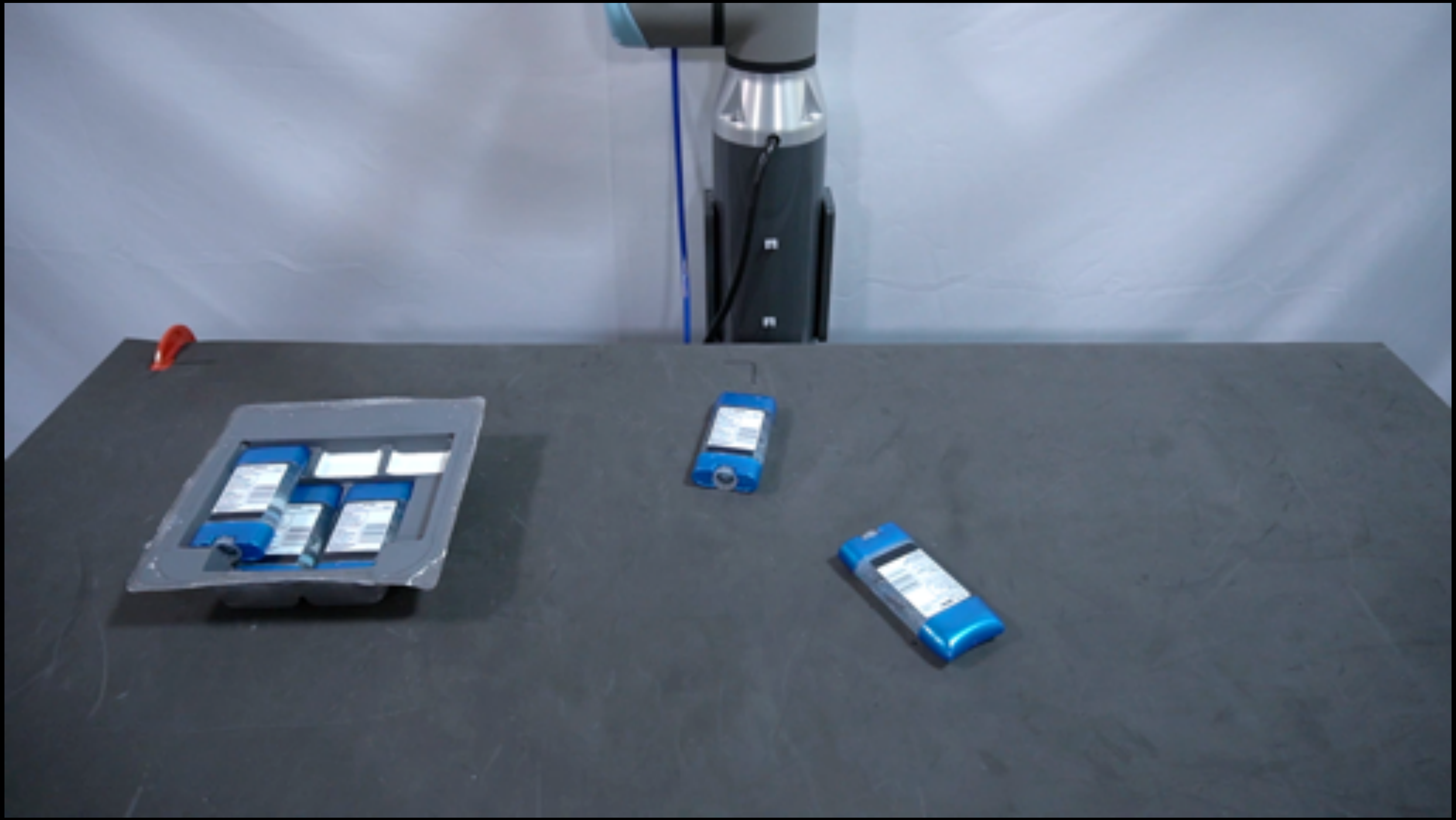
kit is secured to table to prevent accidental displacement from bad suction grasps

# Data Collection from Disassembly



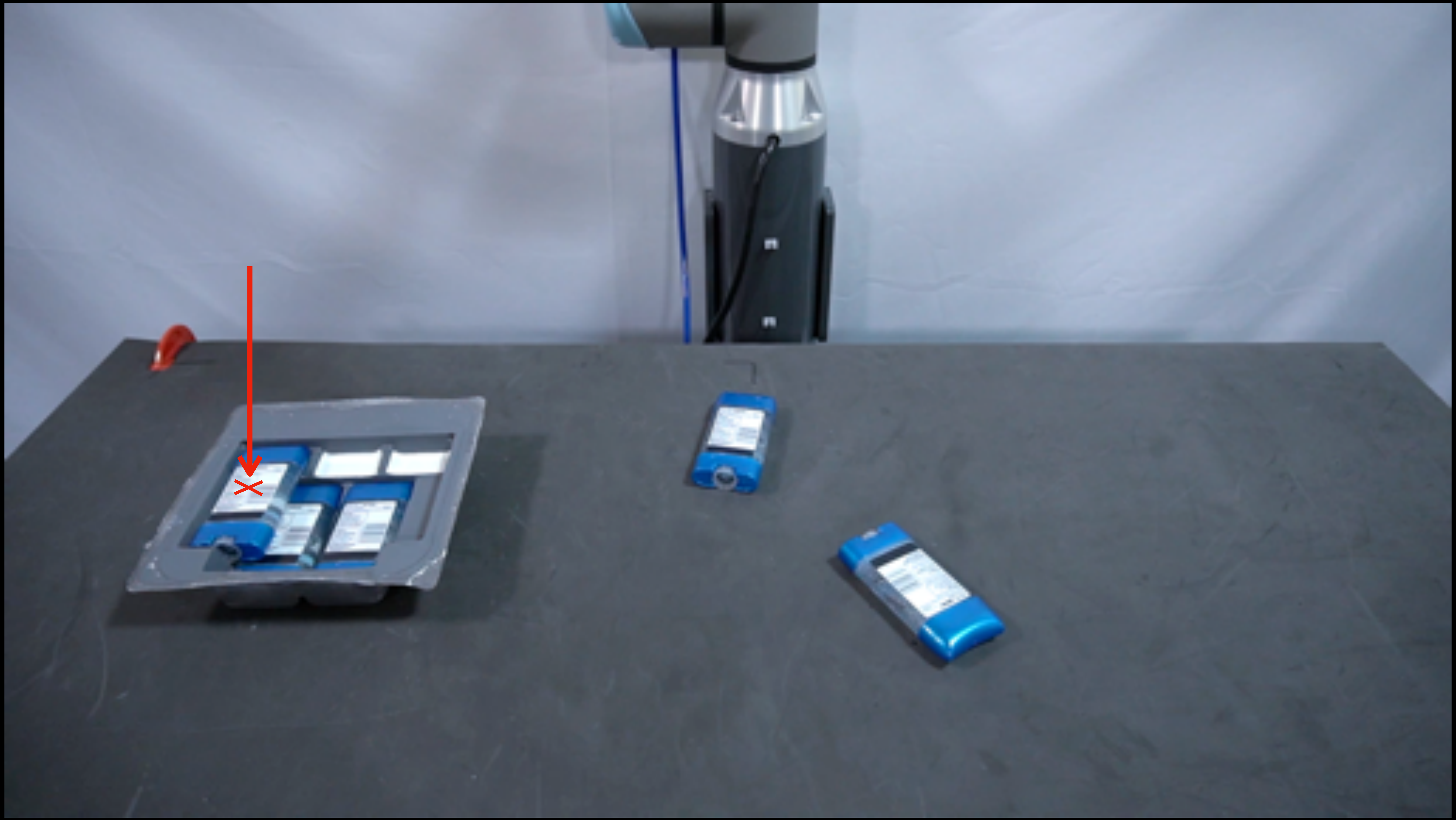
place point ground-truth obtained from suction

# Data Collection from Disassembly



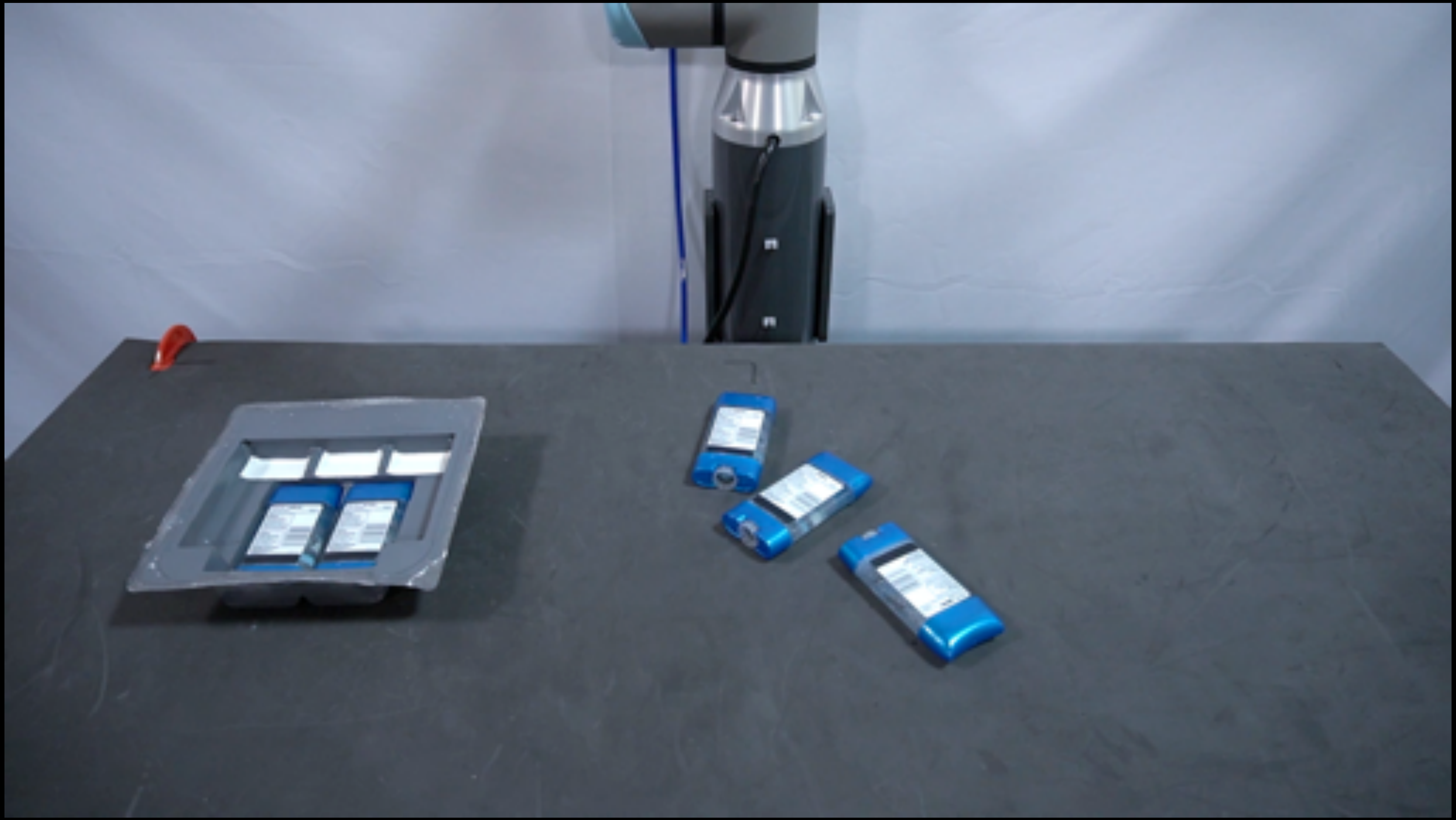
place point ground-truth obtained from suction

# Data Collection from Disassembly



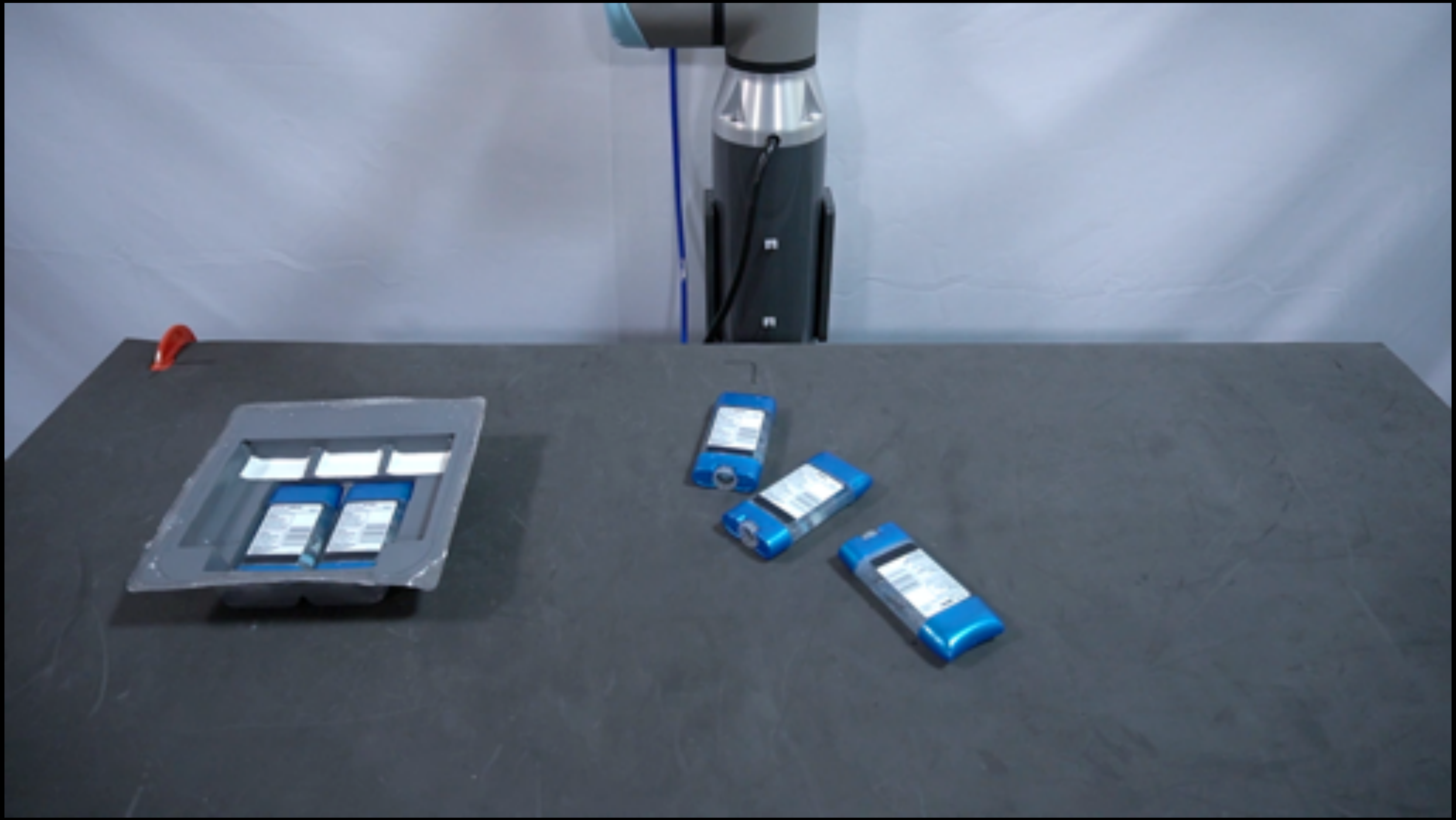
place point ground-truth obtained from suction

# Data Collection from Disassembly



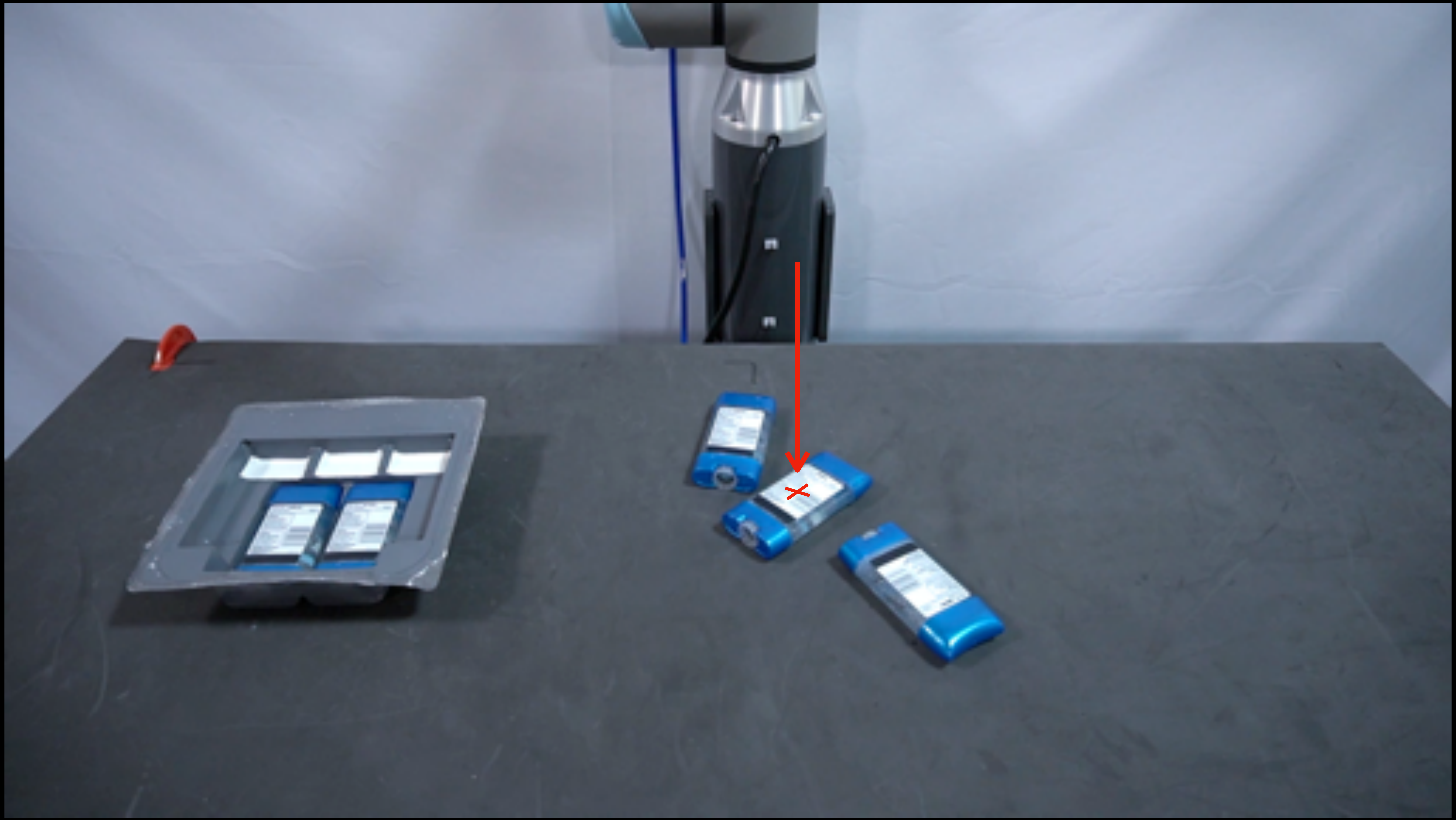
suction point ground-truth obtained from place

# Data Collection from Disassembly



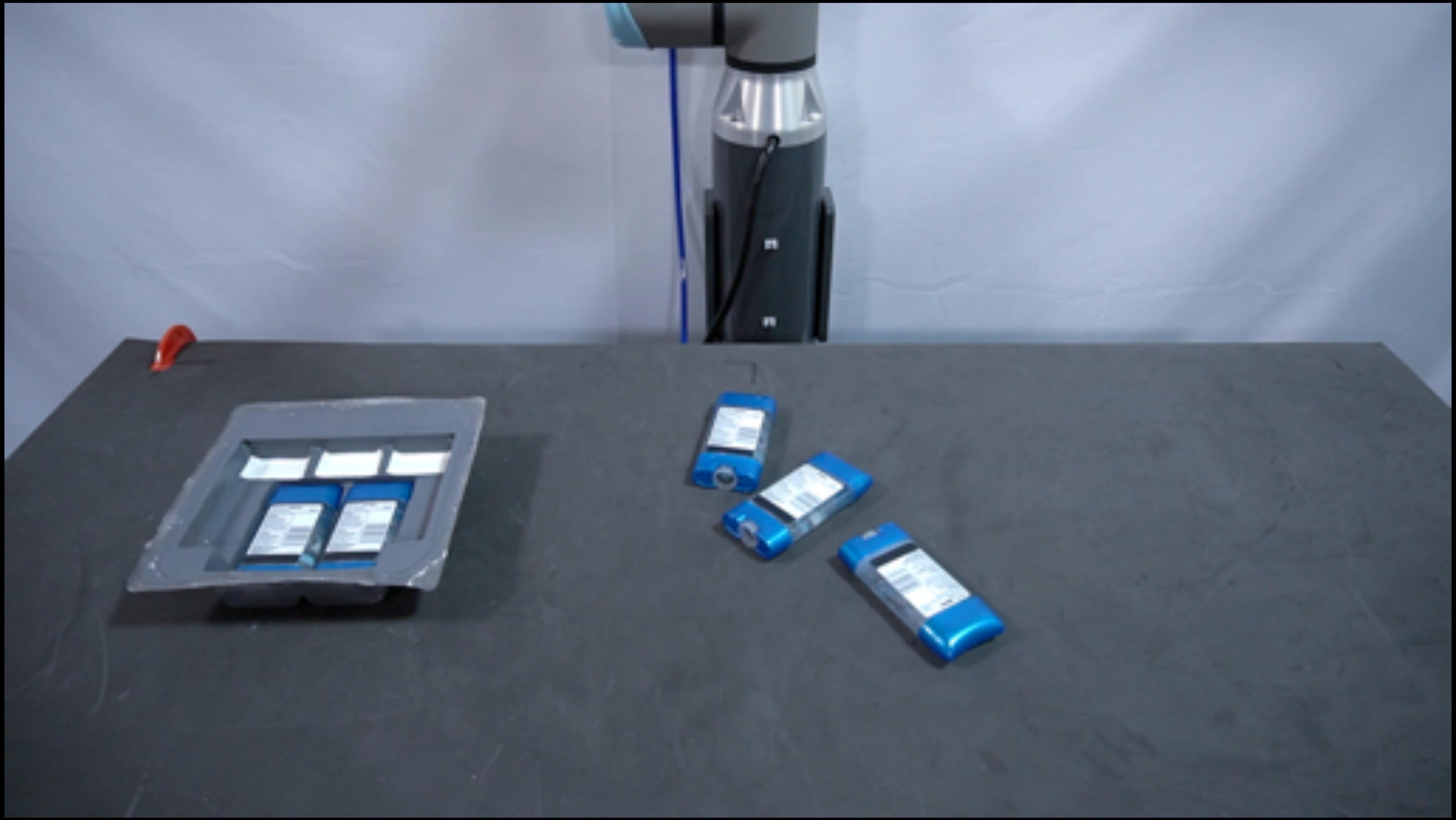
suction point ground-truth obtained from place

# Data Collection from Disassembly

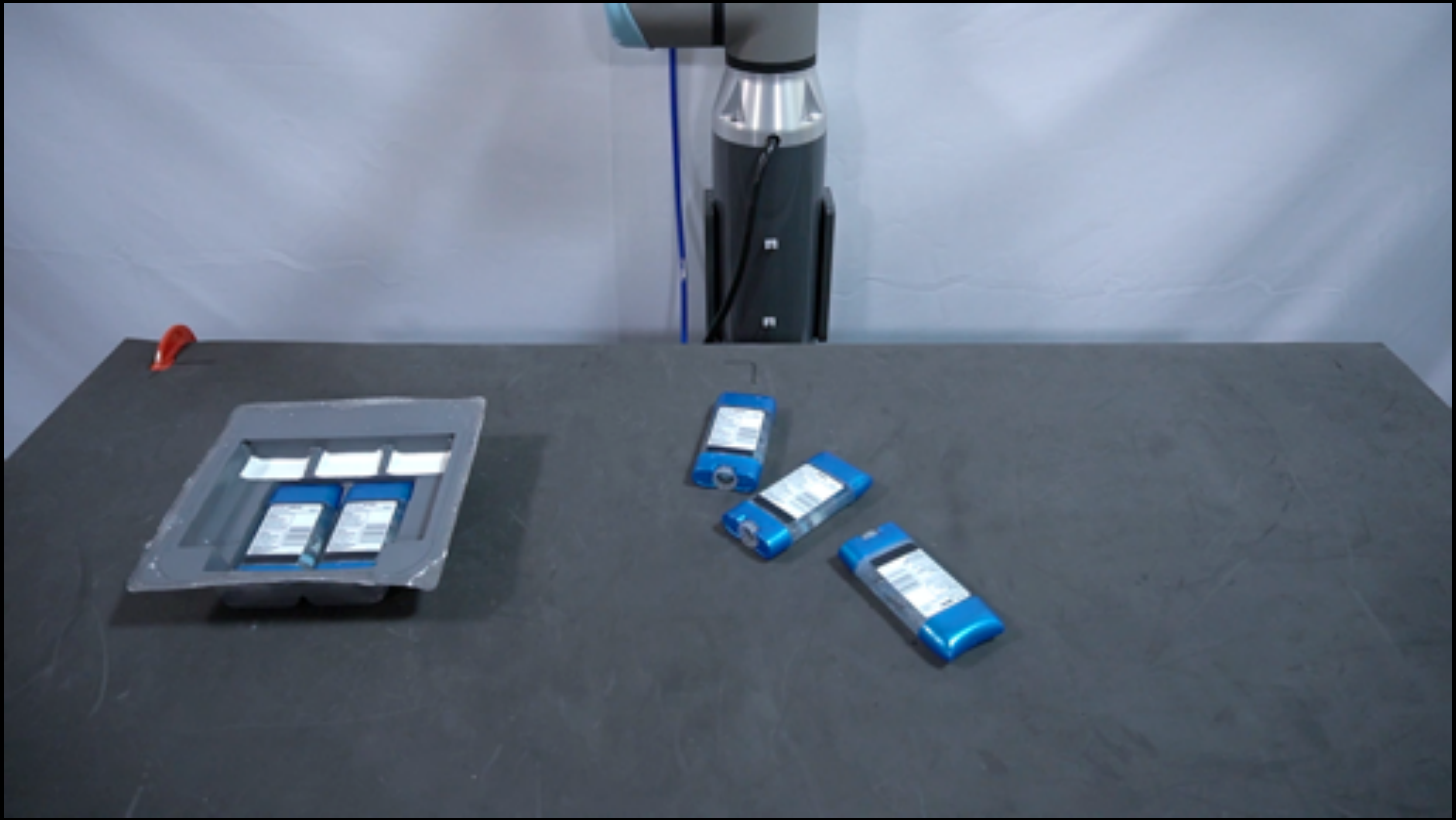


suction point ground-truth obtained from place

# Data Collection from Disassembly

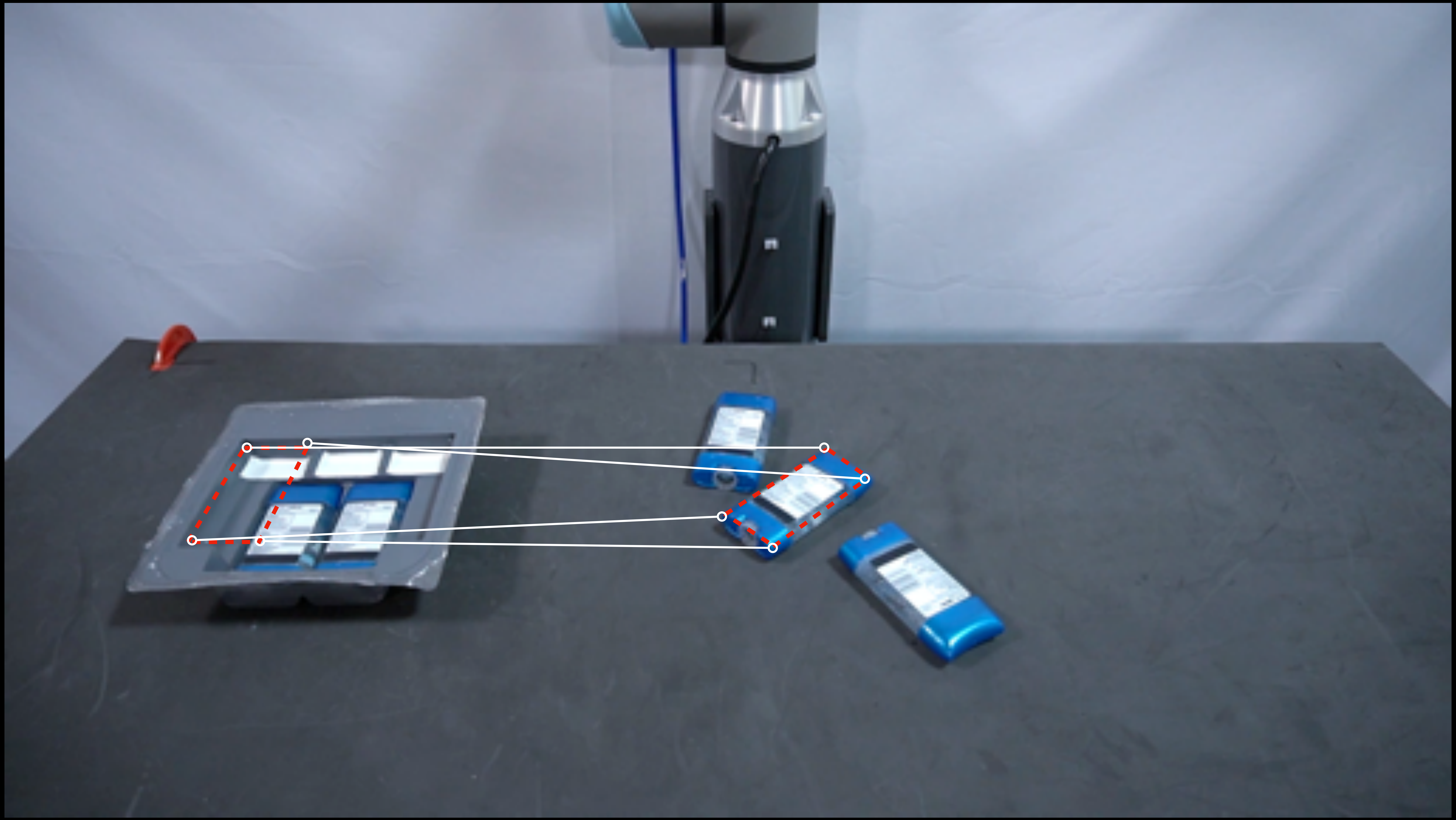


# Data Collection from Disassembly



dense correspondence ground-truth obtained from robot motion

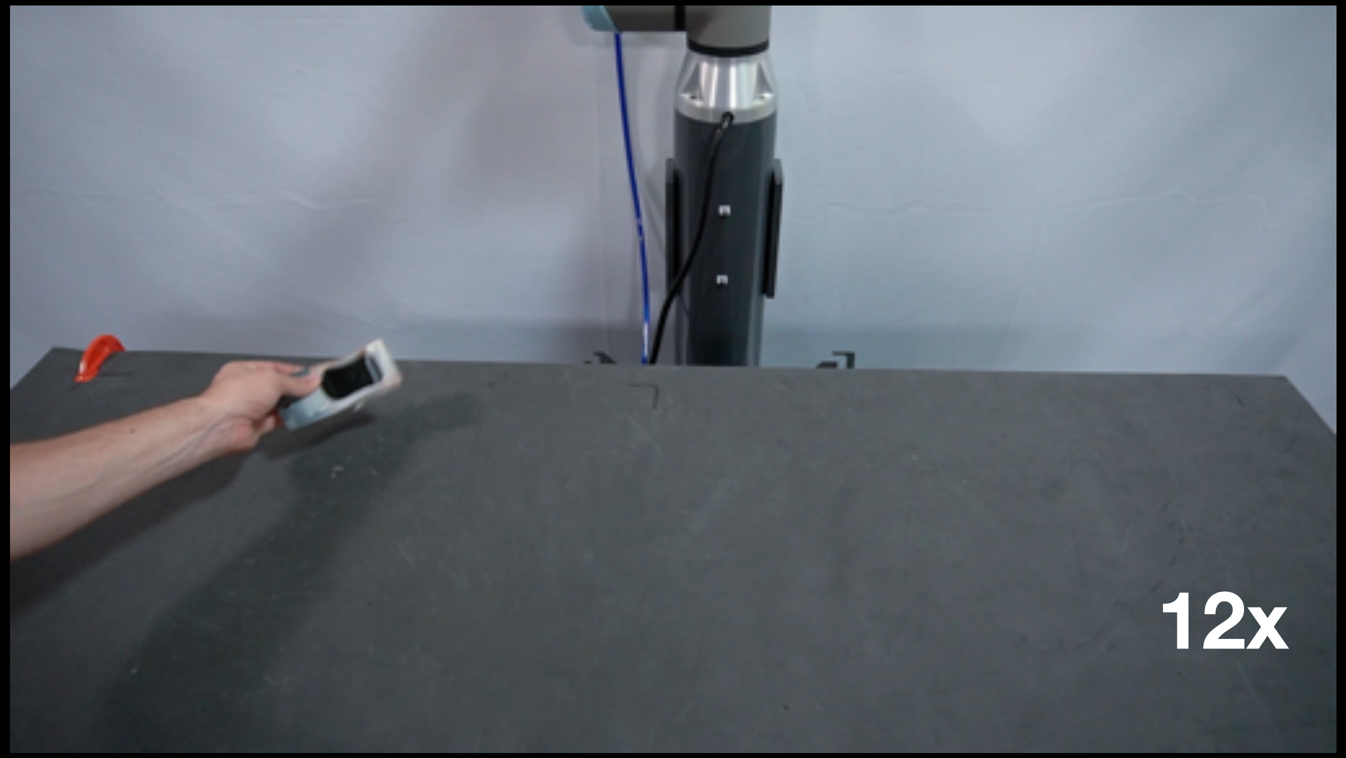
# Data Collection from Disassembly

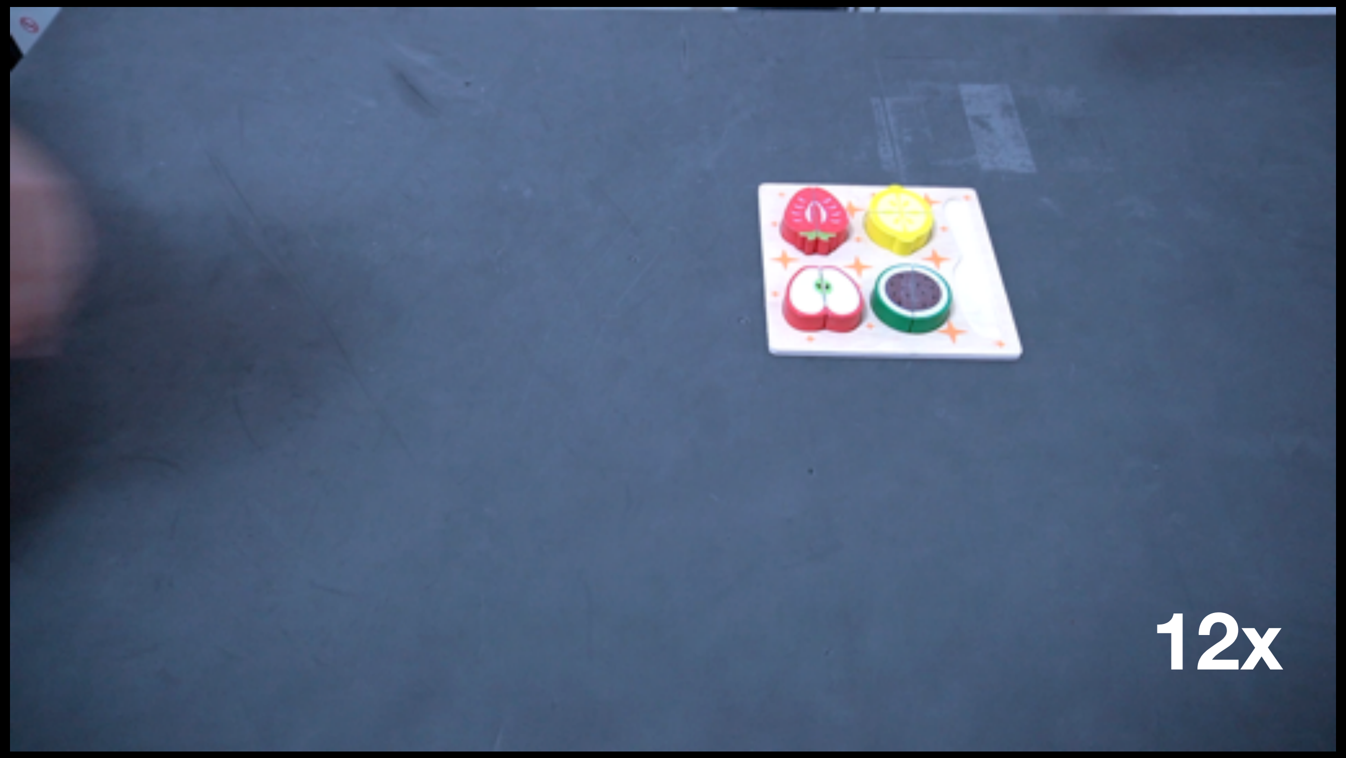


dense correspondence ground-truth obtained from robot motion

# Results

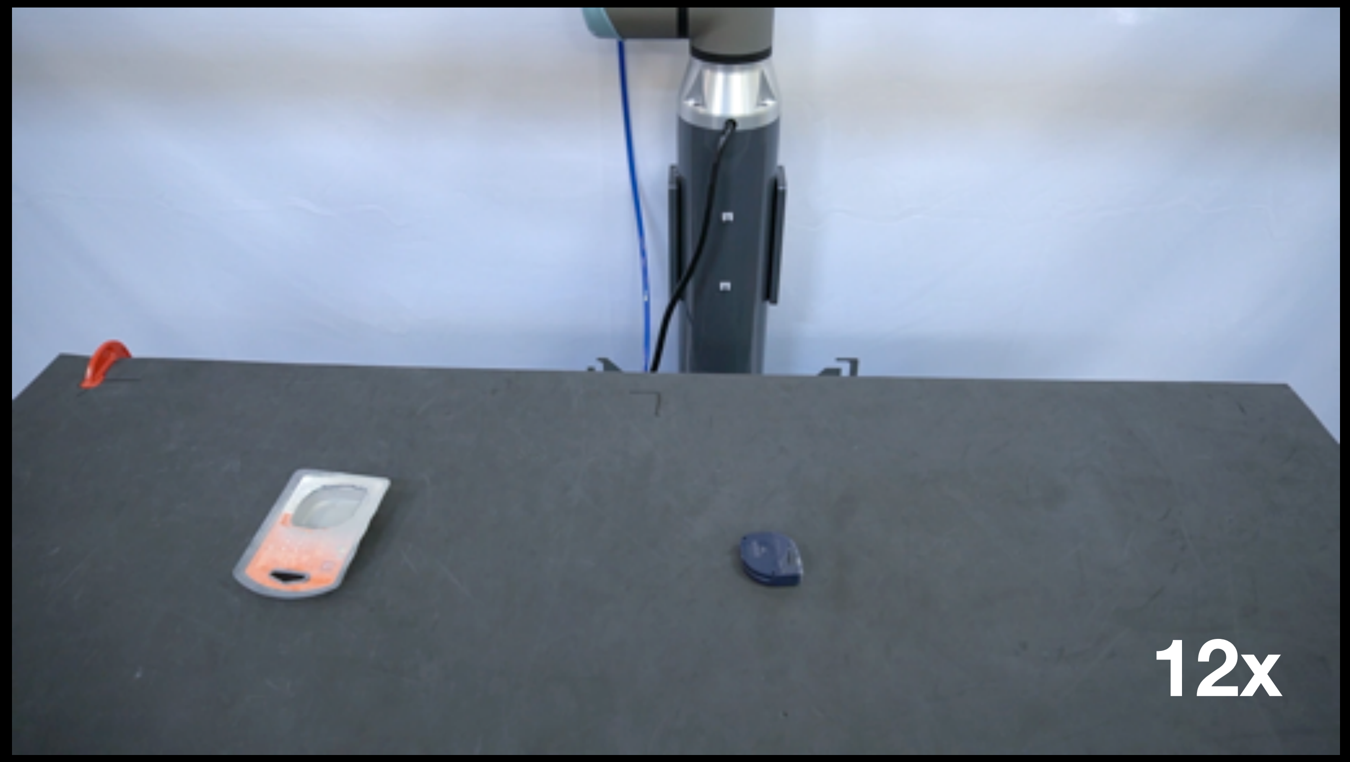
# Varying Initial Conditions

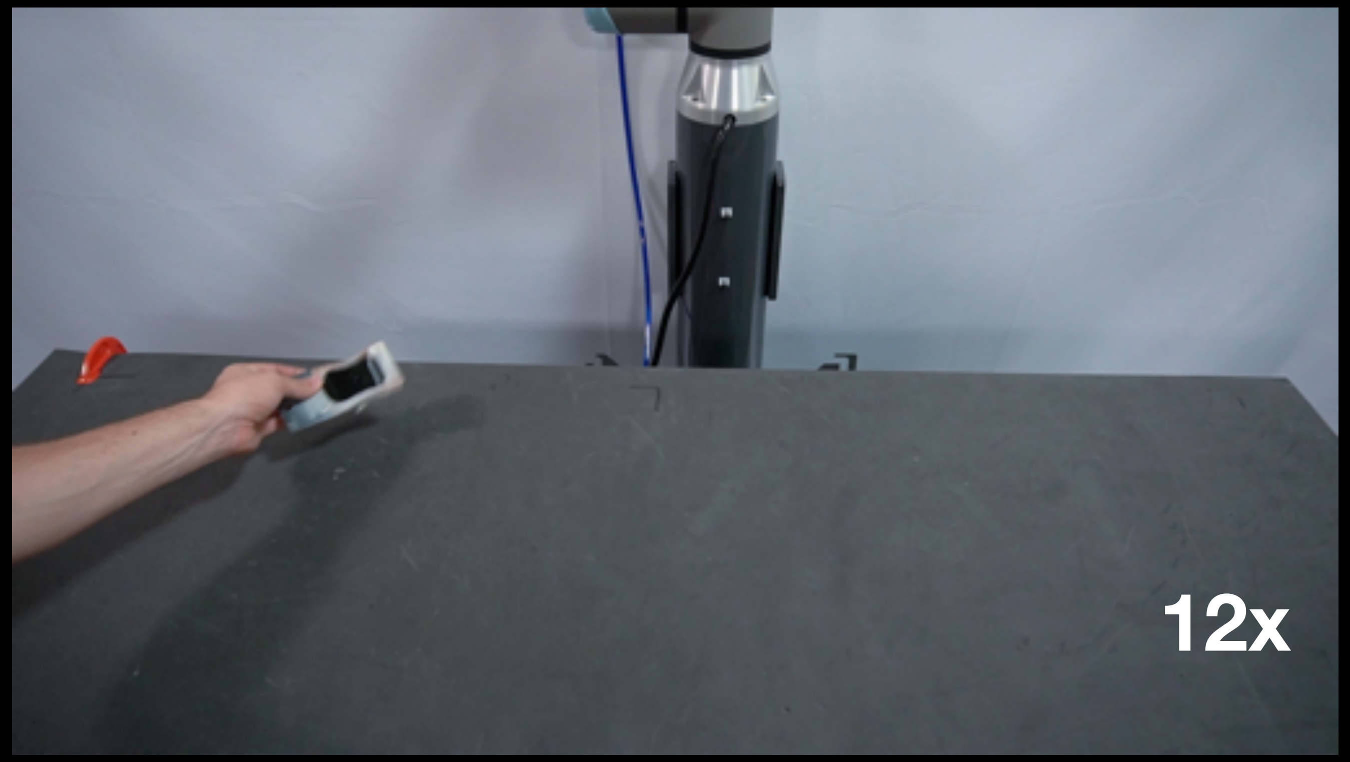





model trained and tested on each kit

# Varying Initial Conditions

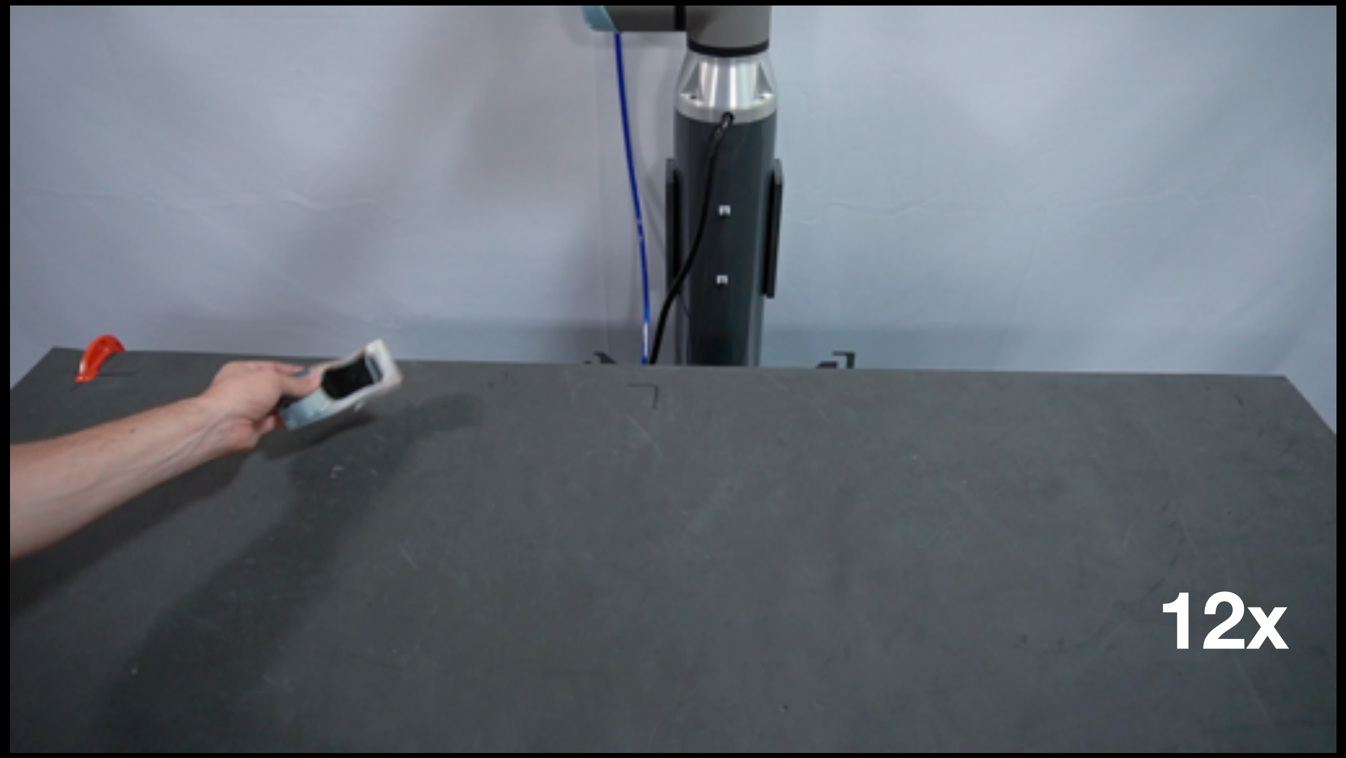



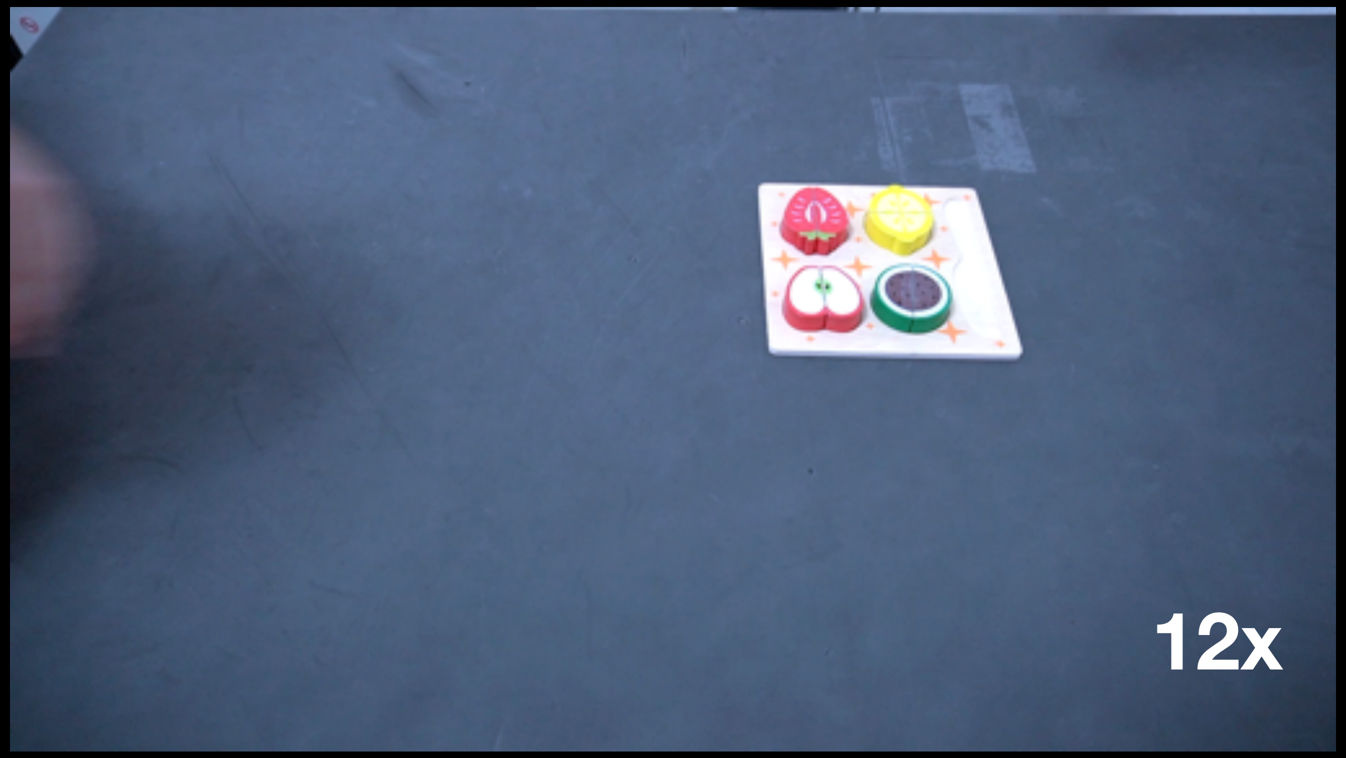






model trained and tested on each kit

# Varying Initial Conditions

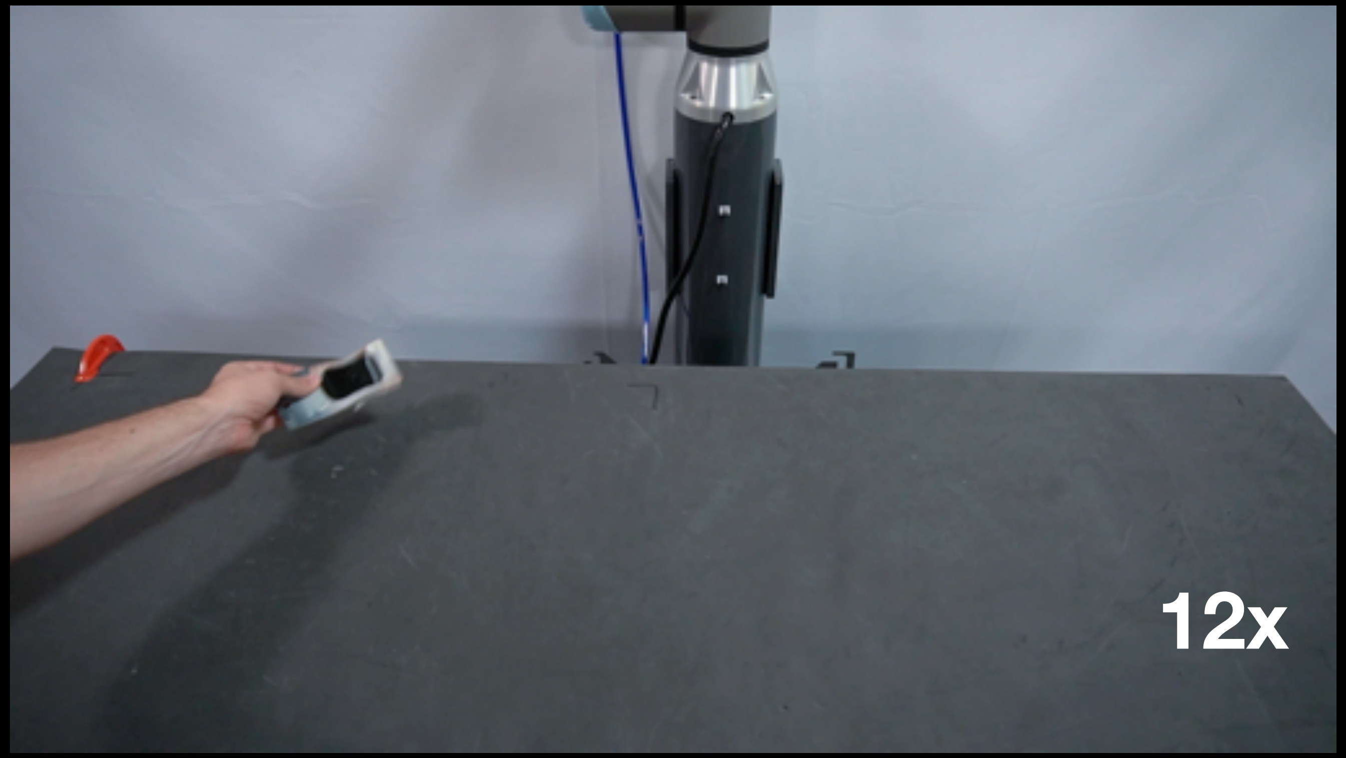


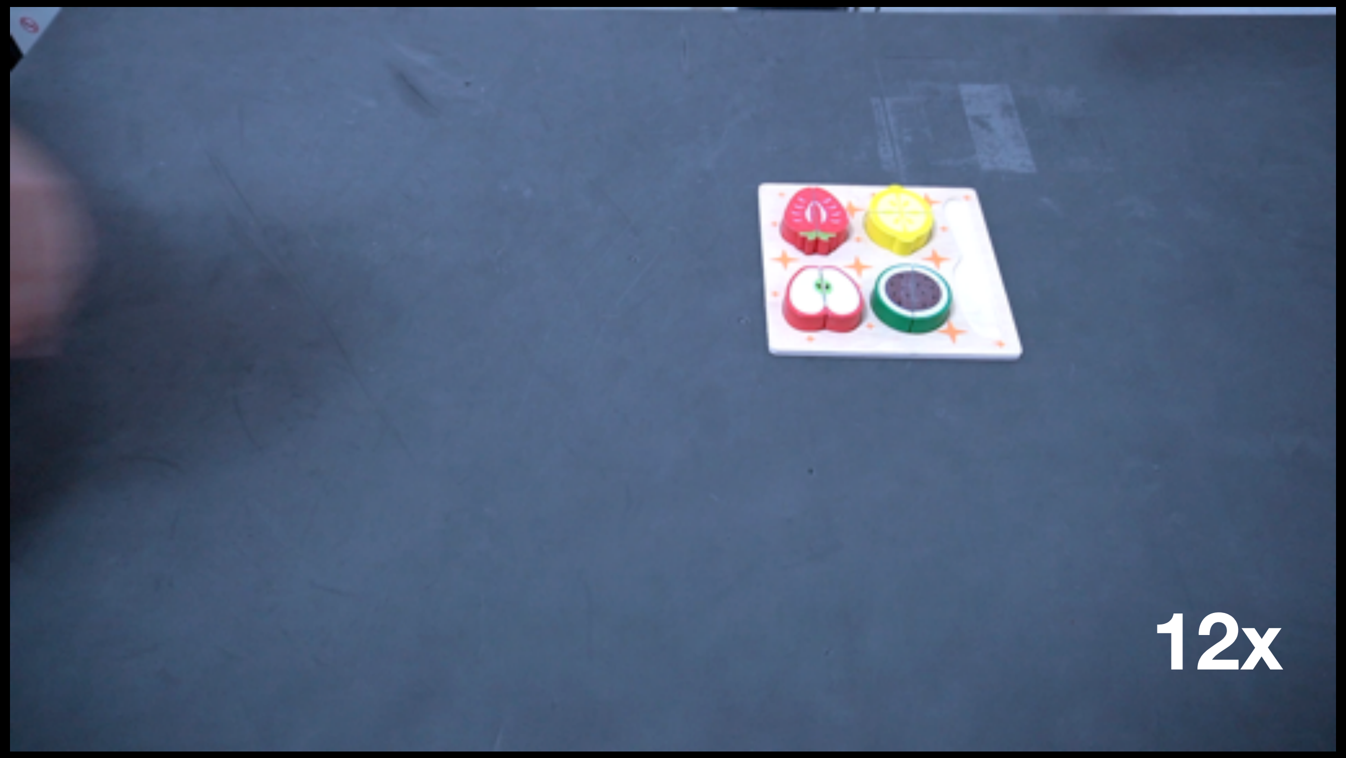



model trained and tested on each kit

# Varying Initial Conditions



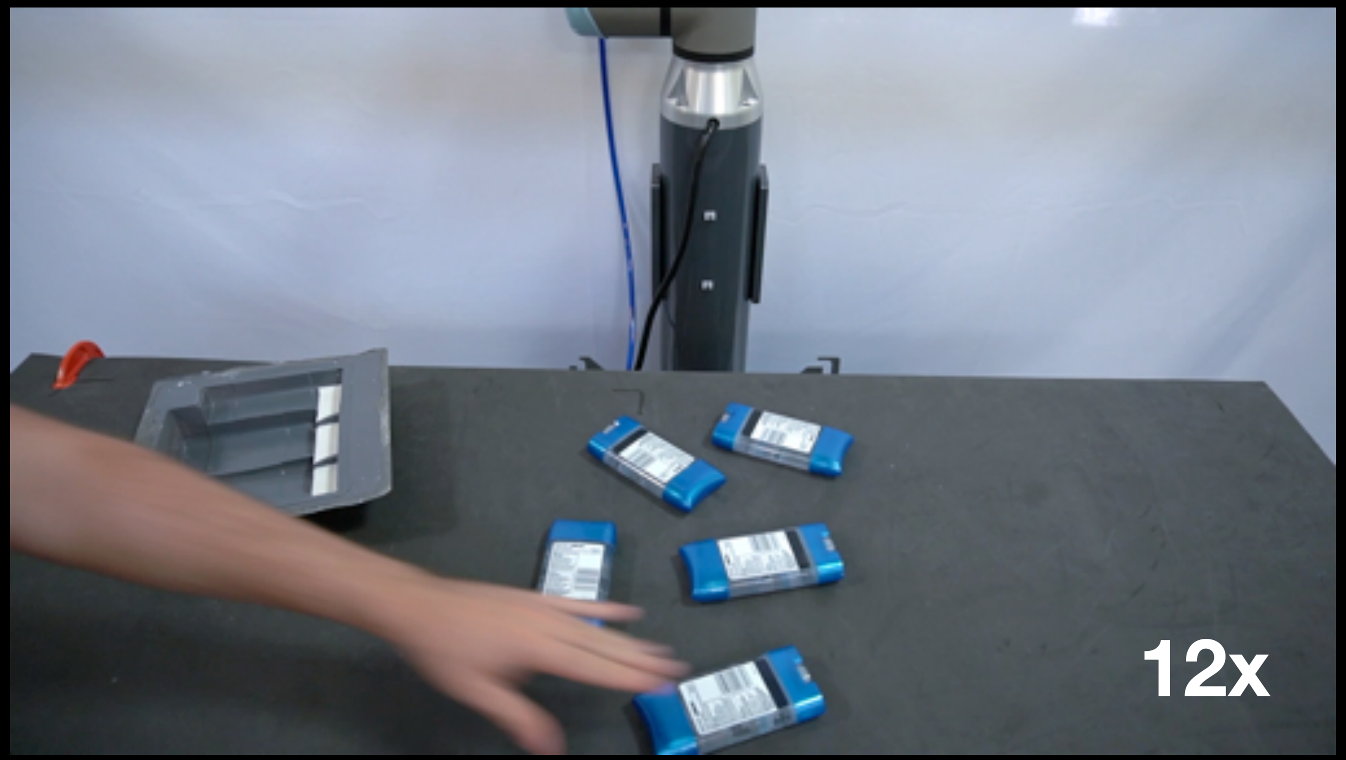


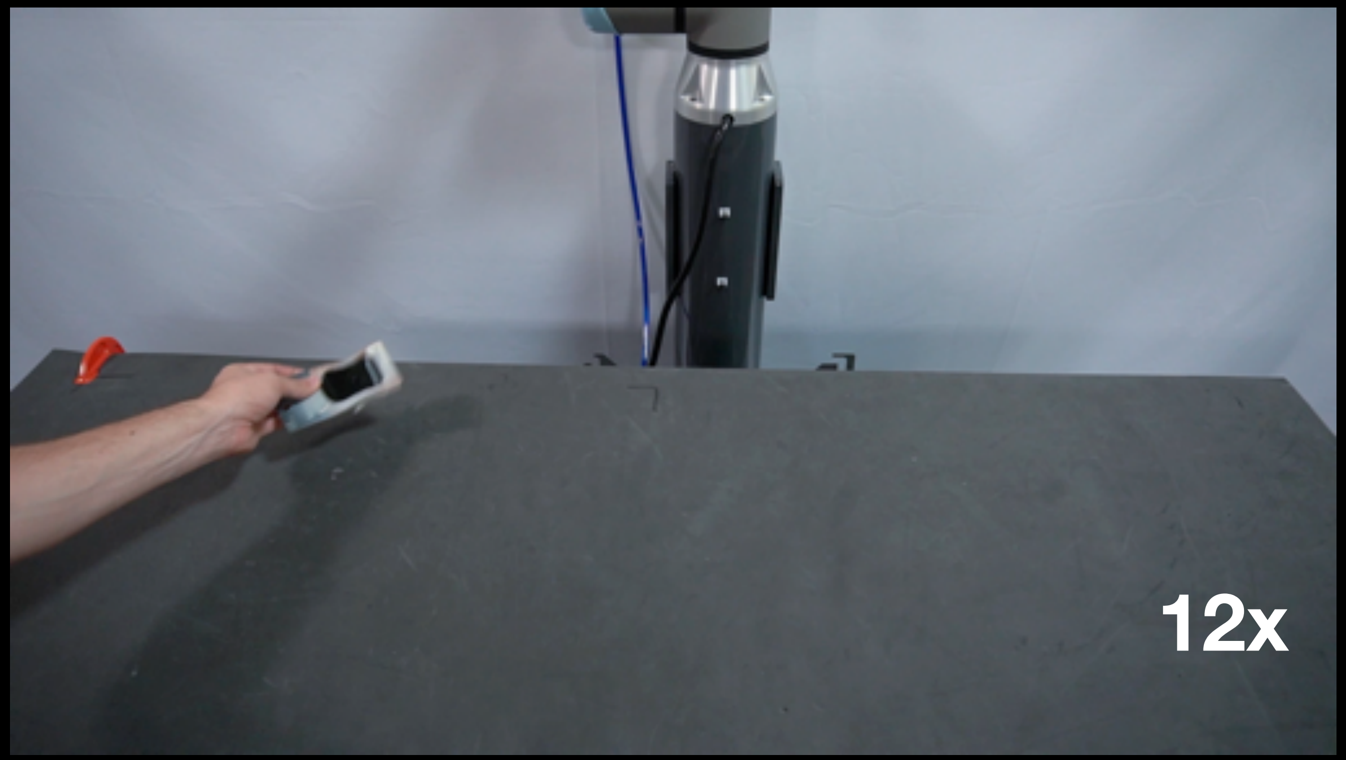


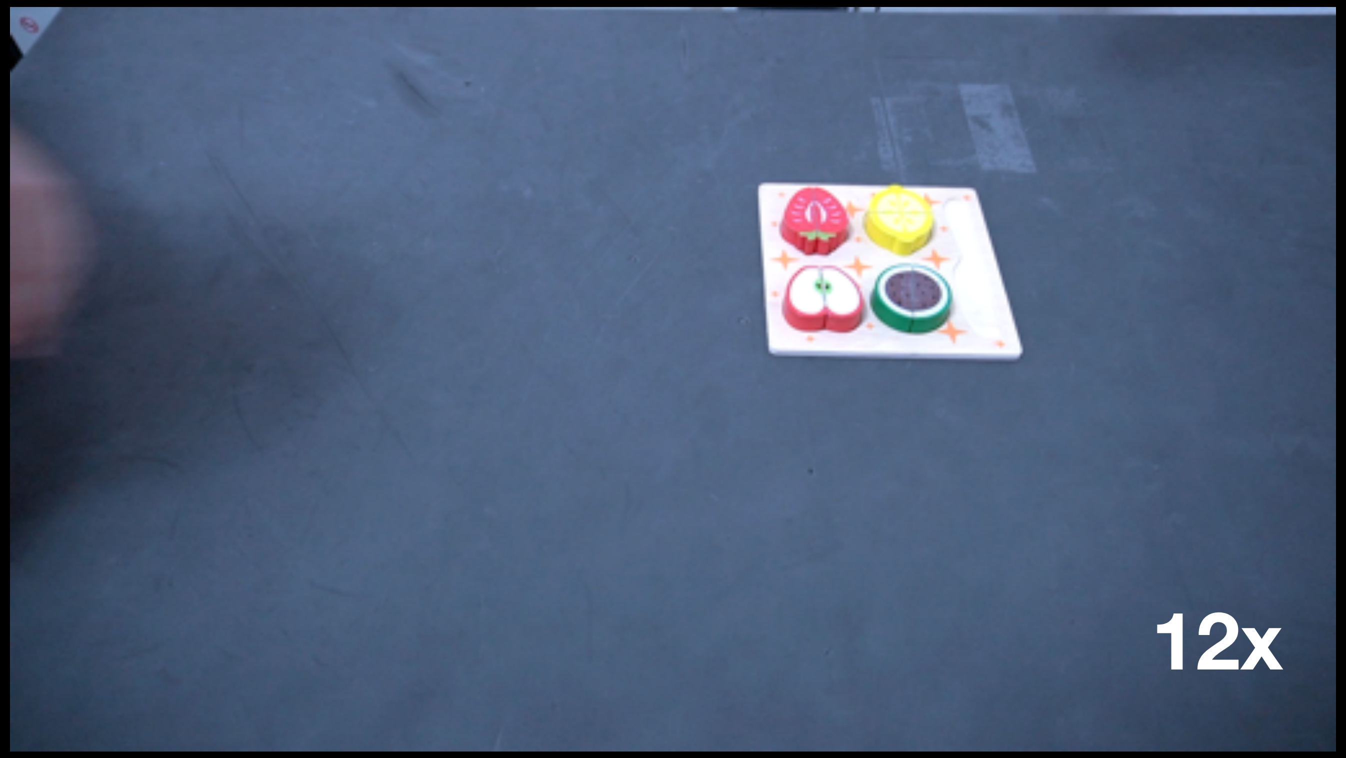
model trained and tested on each kit

# Varying Initial Conditions







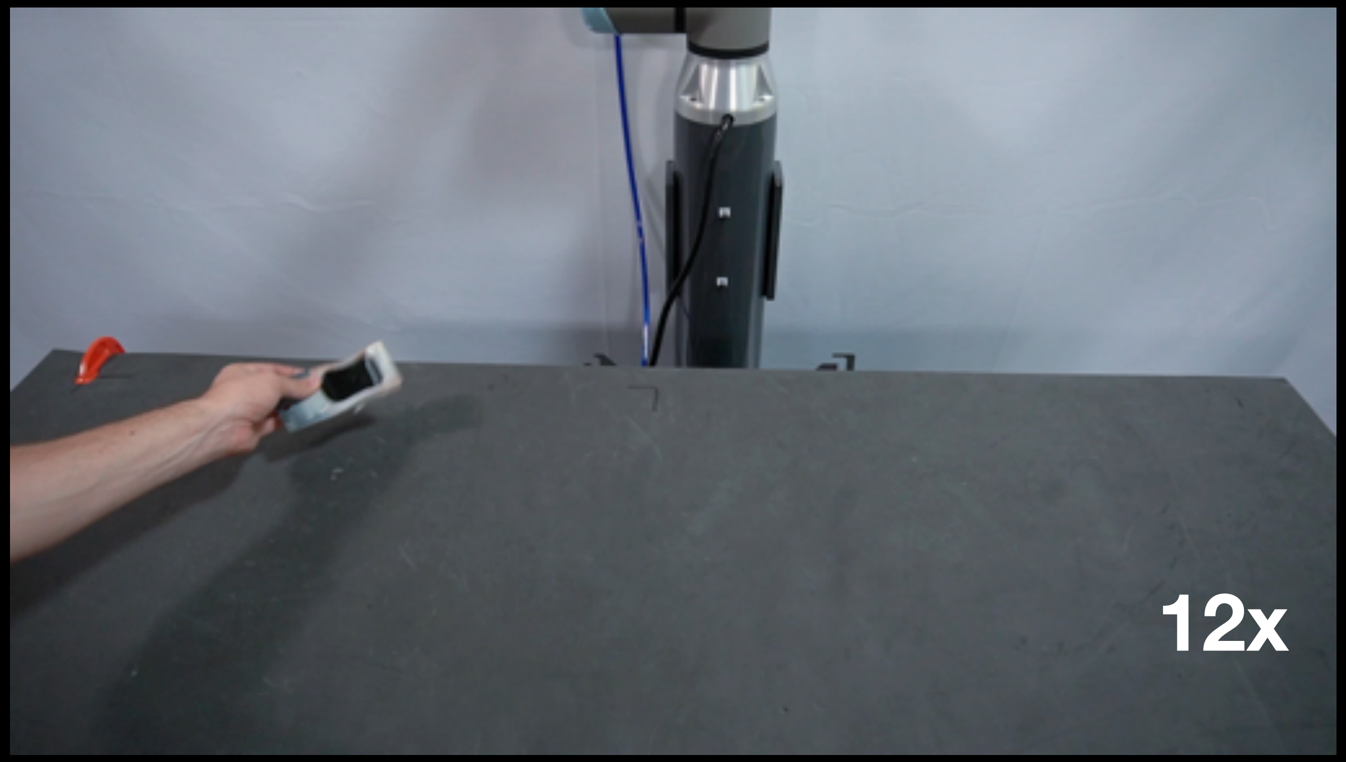



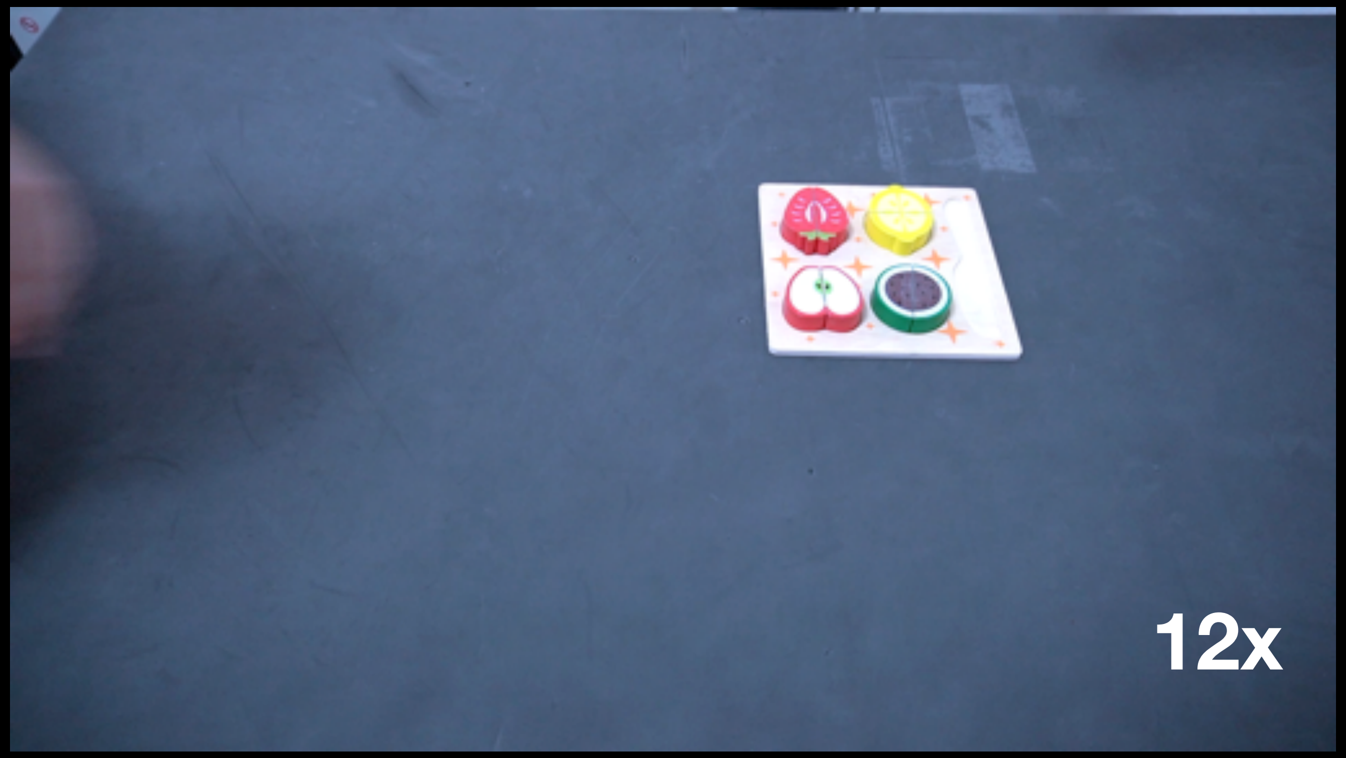


model trained and tested on each kit

# Varying Initial Conditions







model trained and tested on each kit

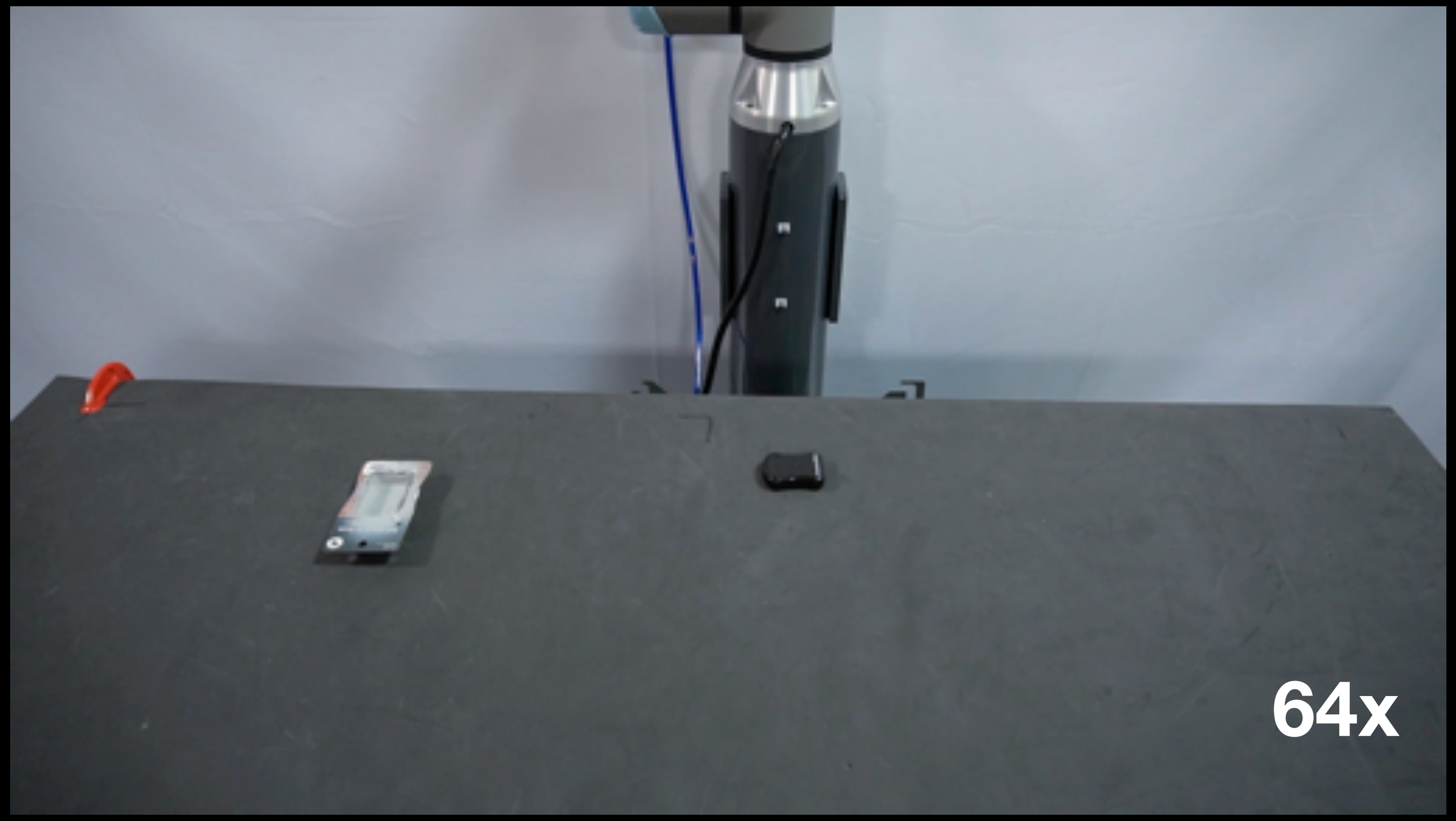
# Generalization to Novel Settings

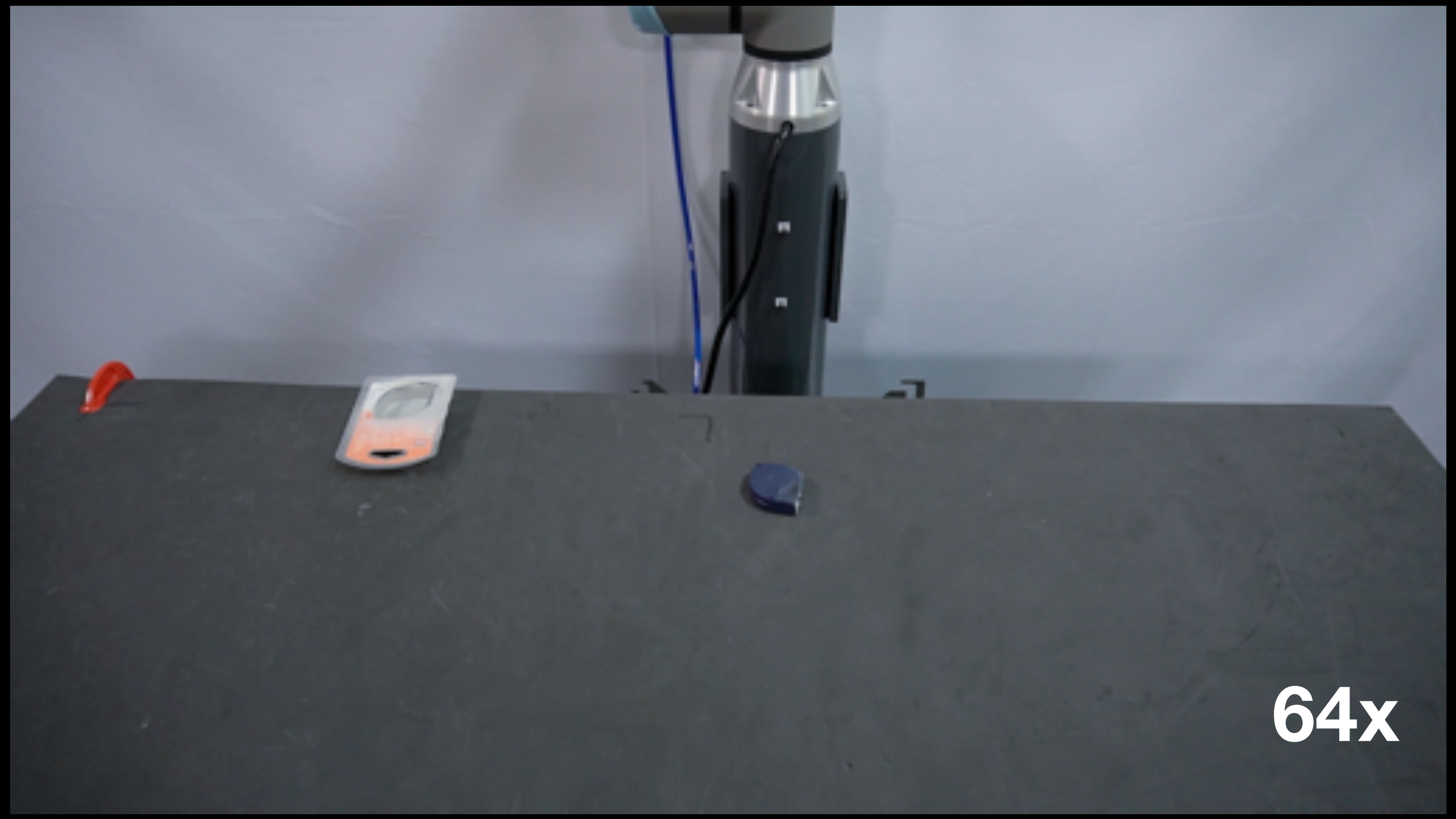
# Generalization to Novel Settings

model trained on 2 kits: floss and tape

# Generalization to Novel Settings

**Individual**

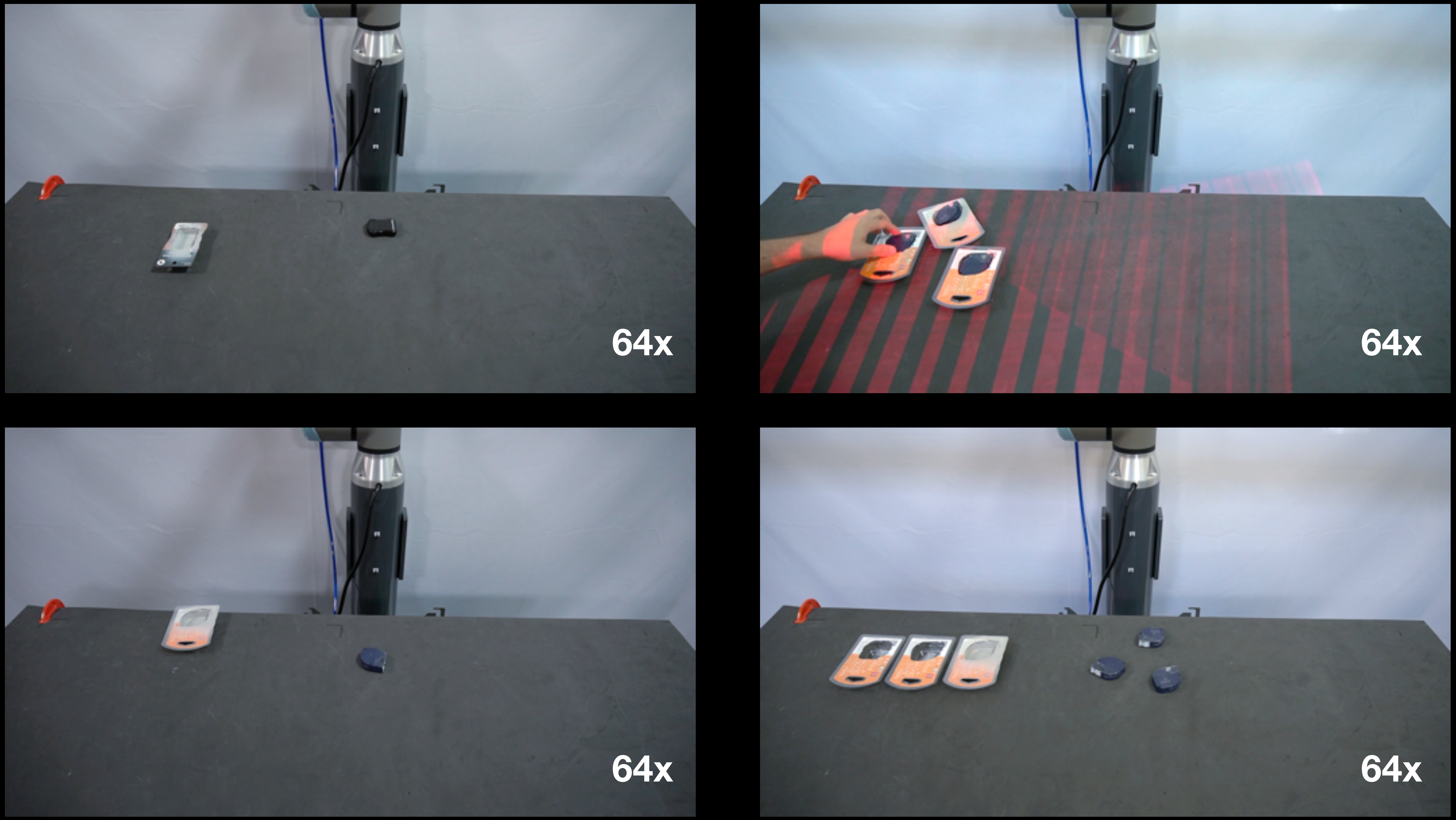



model trained on 2 kits: floss and tape

# Generalization to Novel Settings

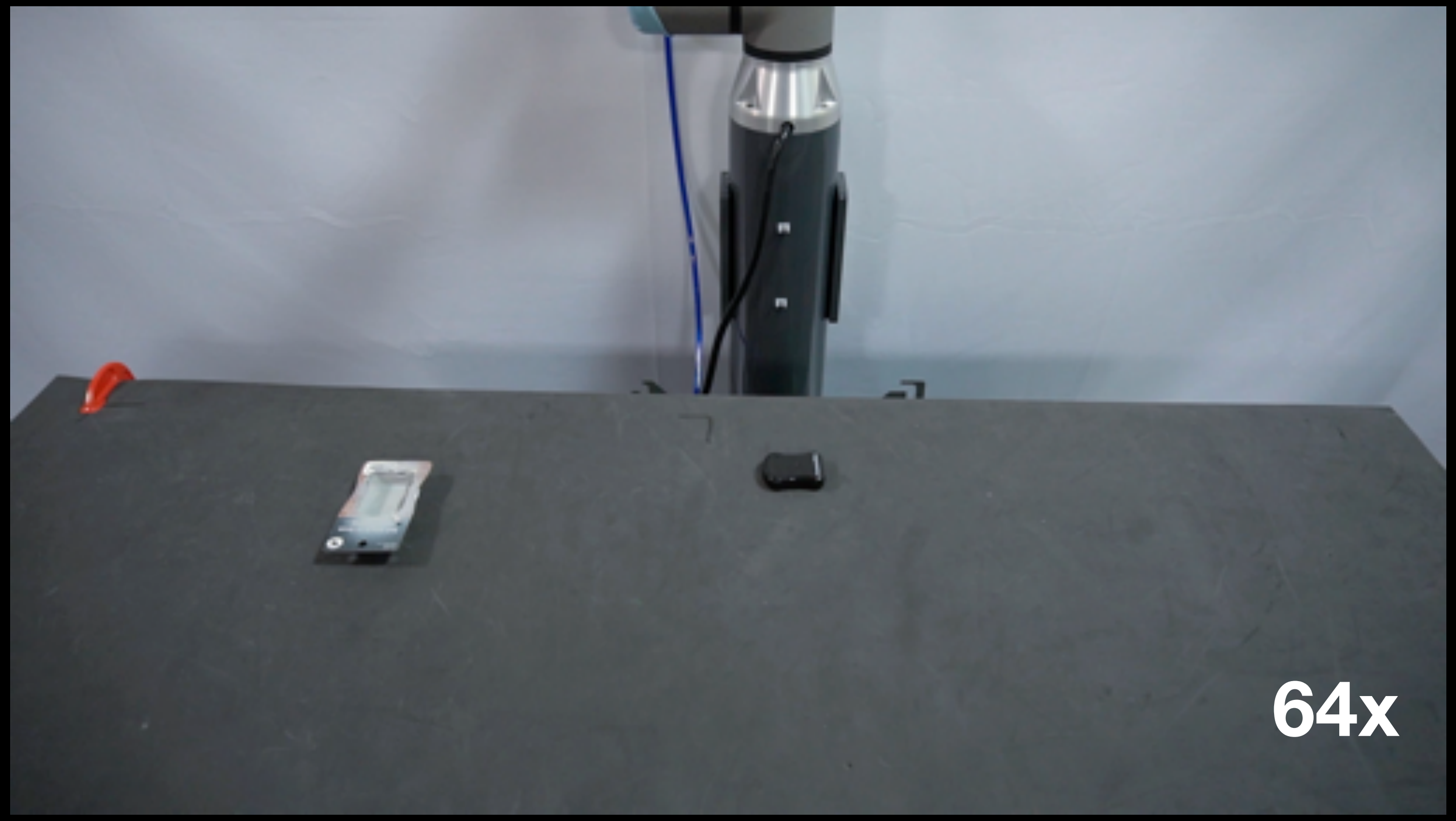
**Individual** **Multiple**

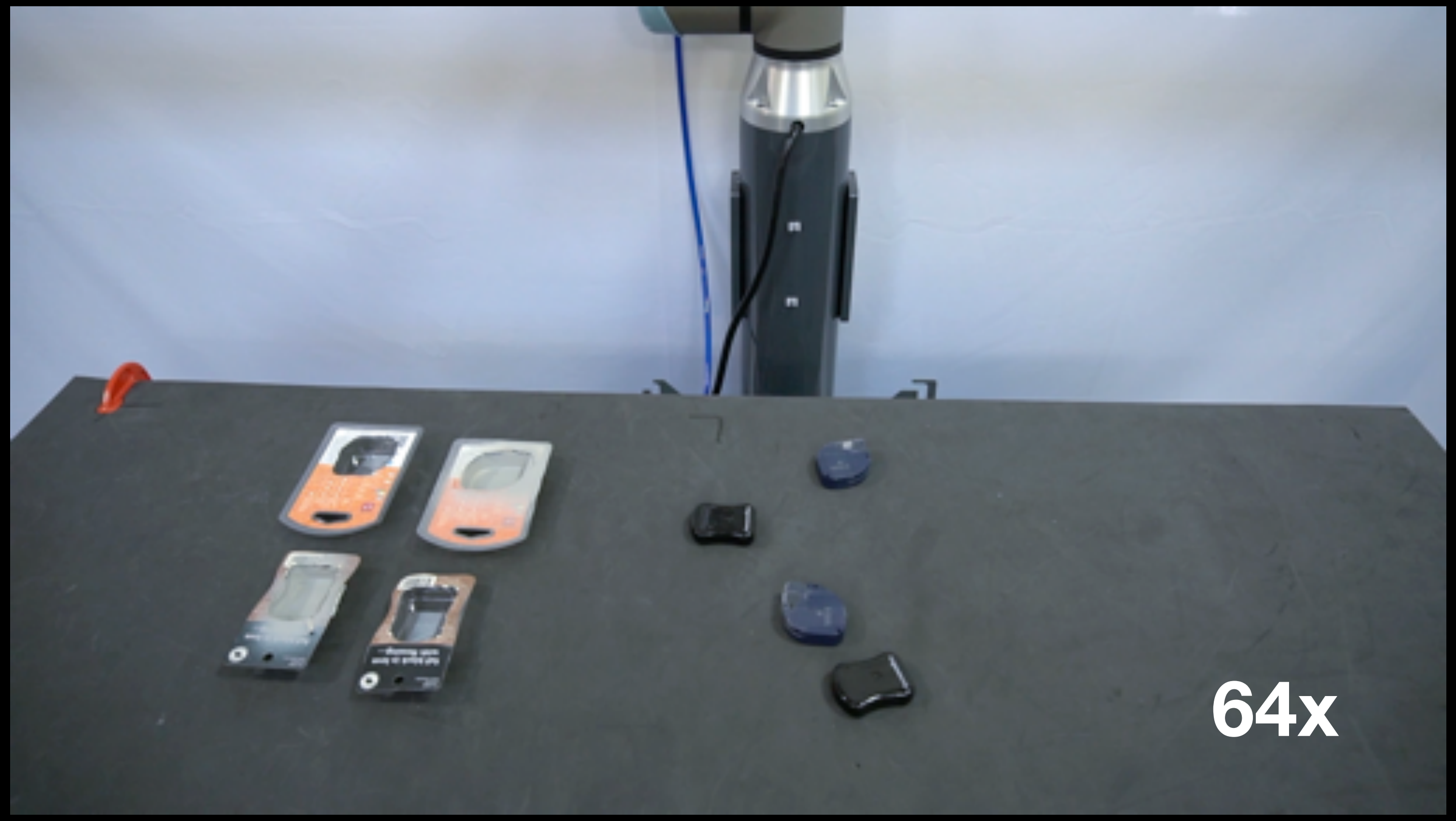


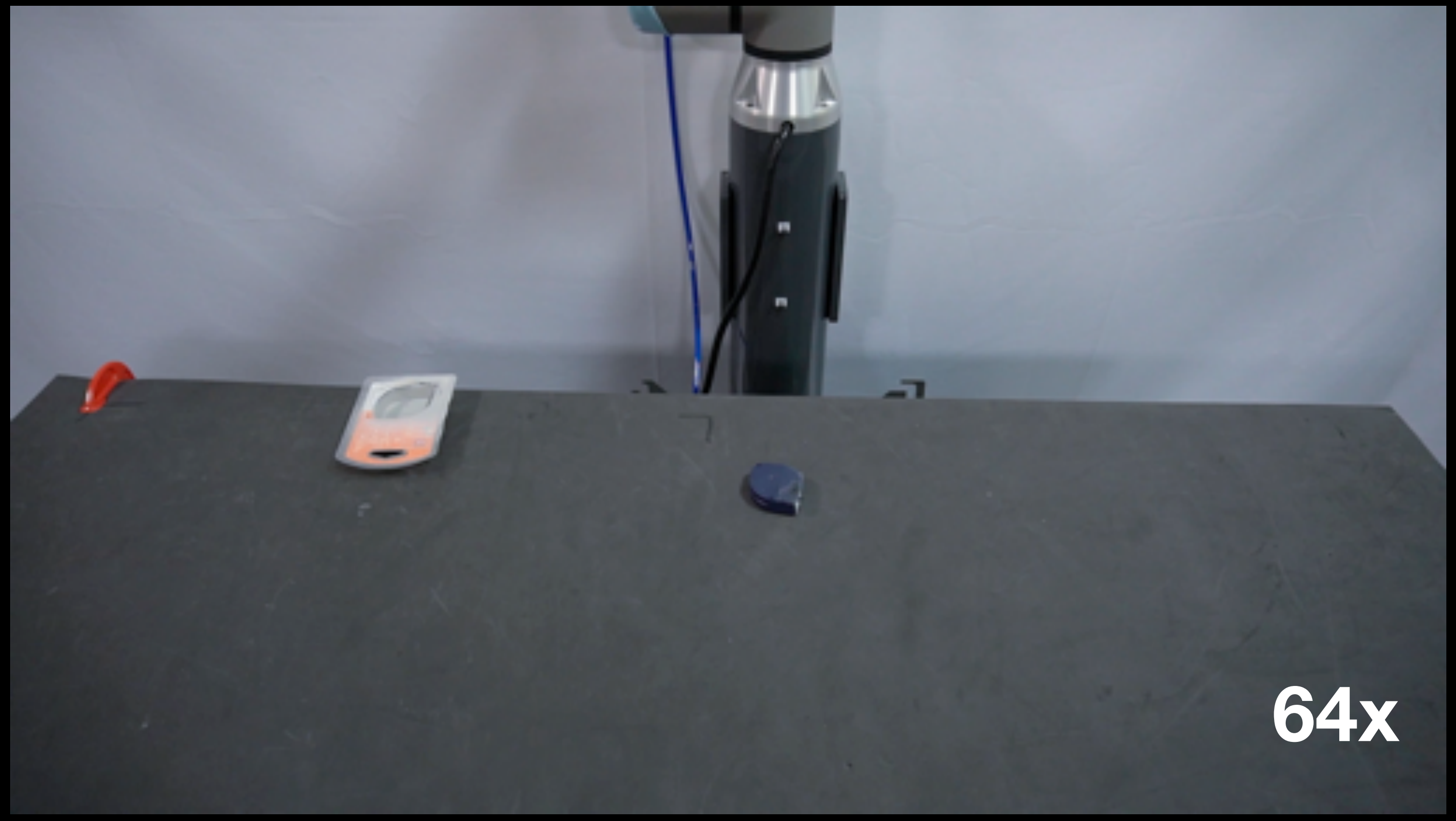

model trained on 2 kits: floss and tape

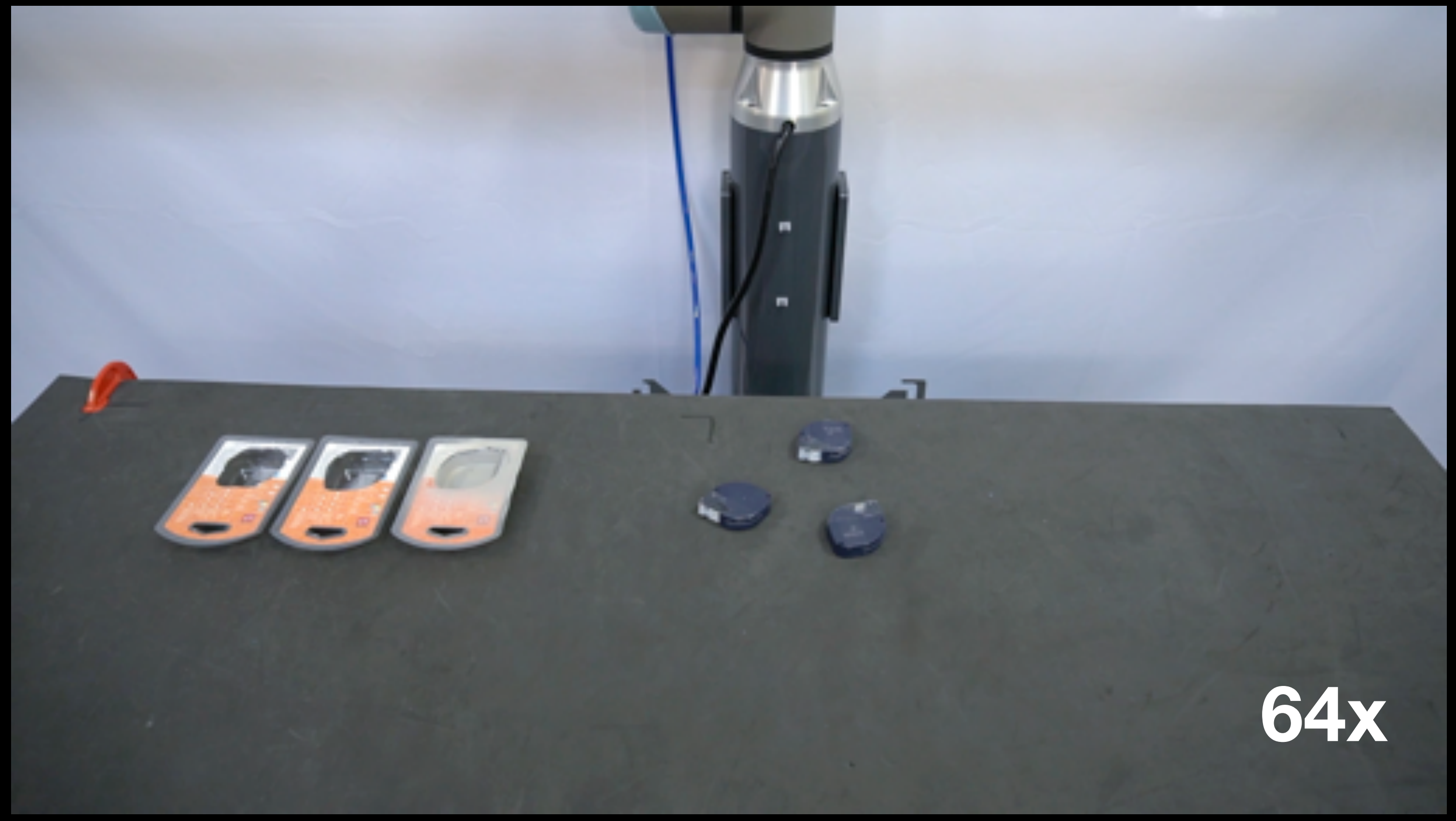
# Generalization to Novel Settings

**Individual** | **Multiple** | **Mixture**





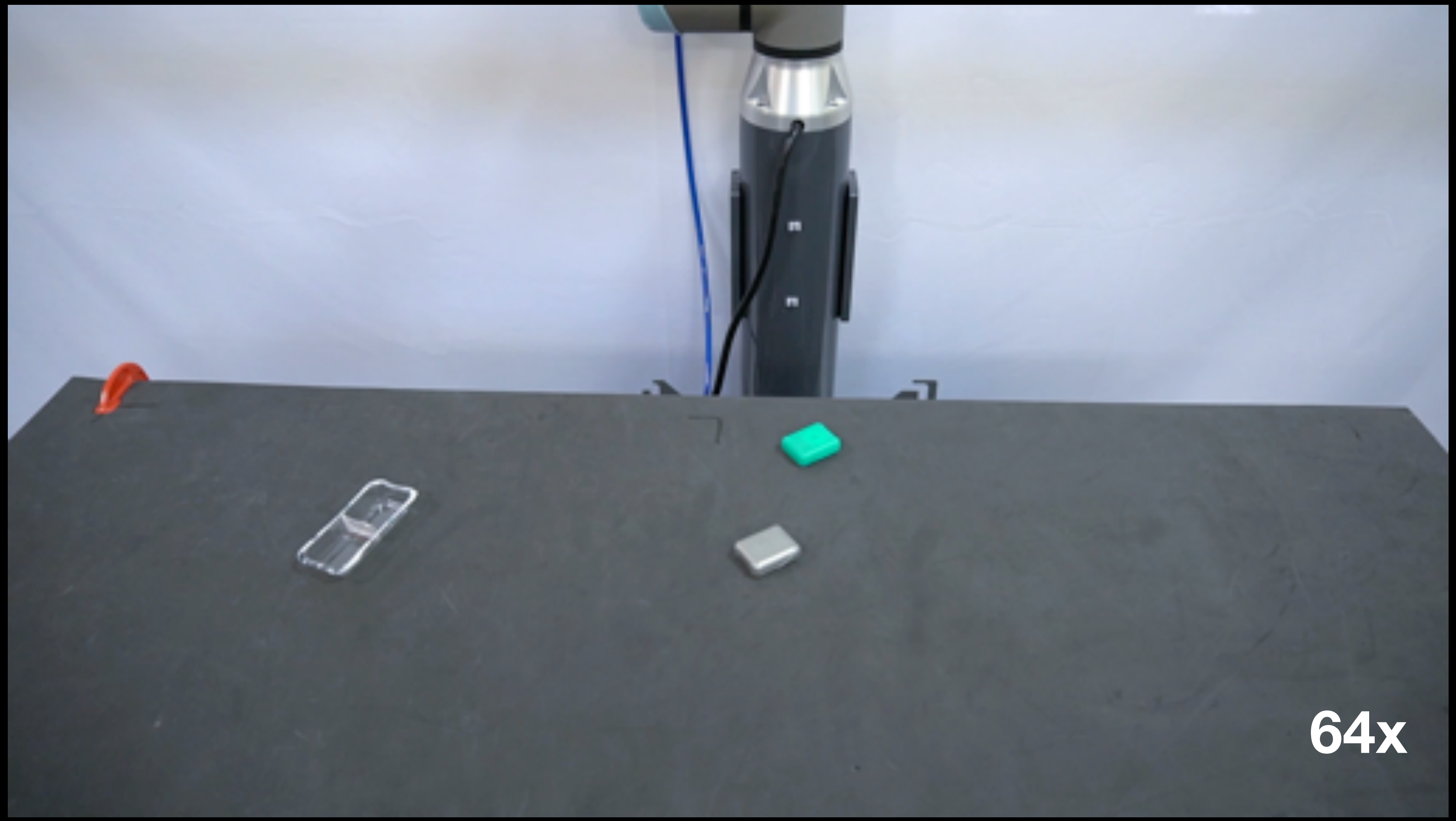


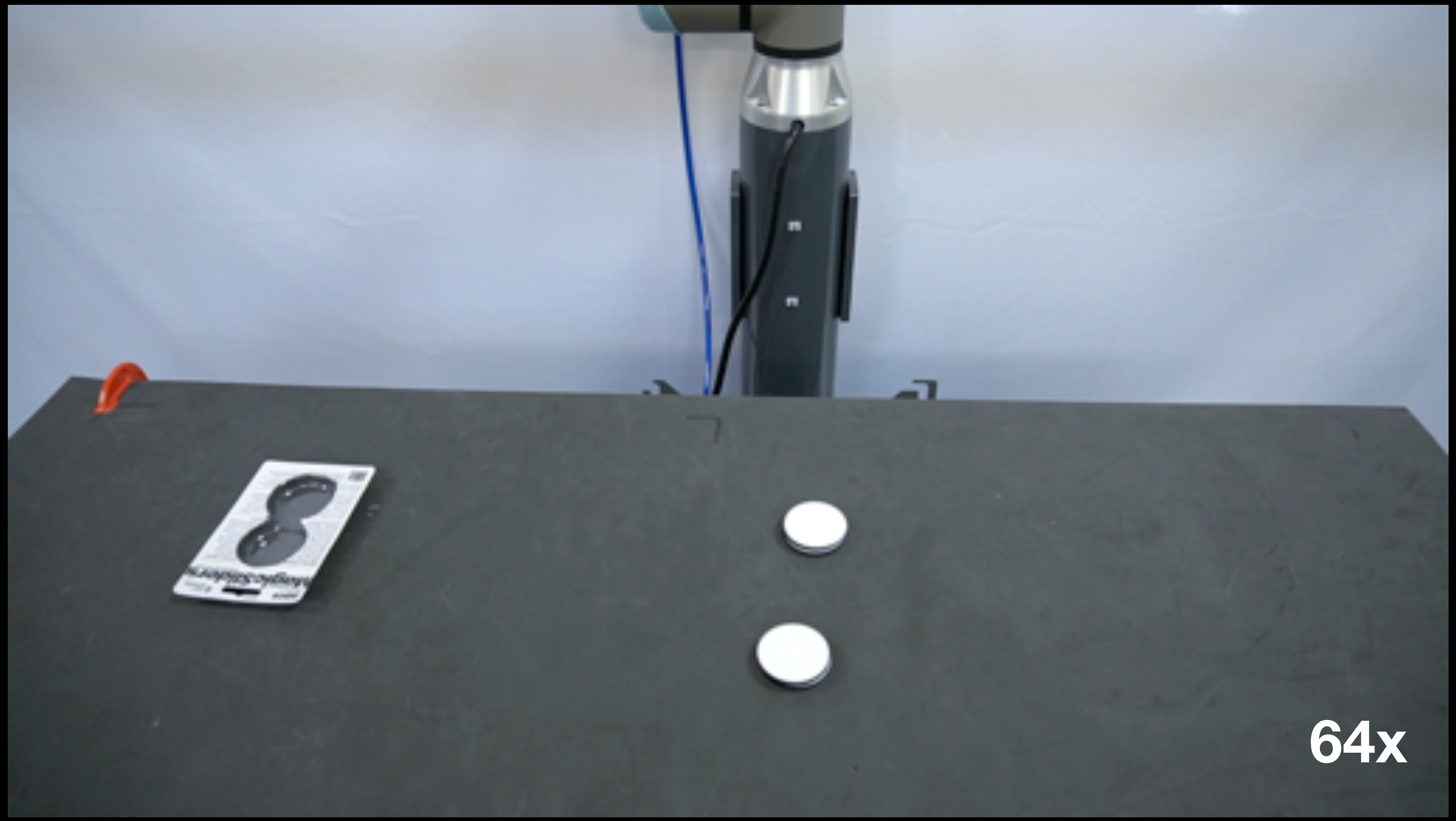


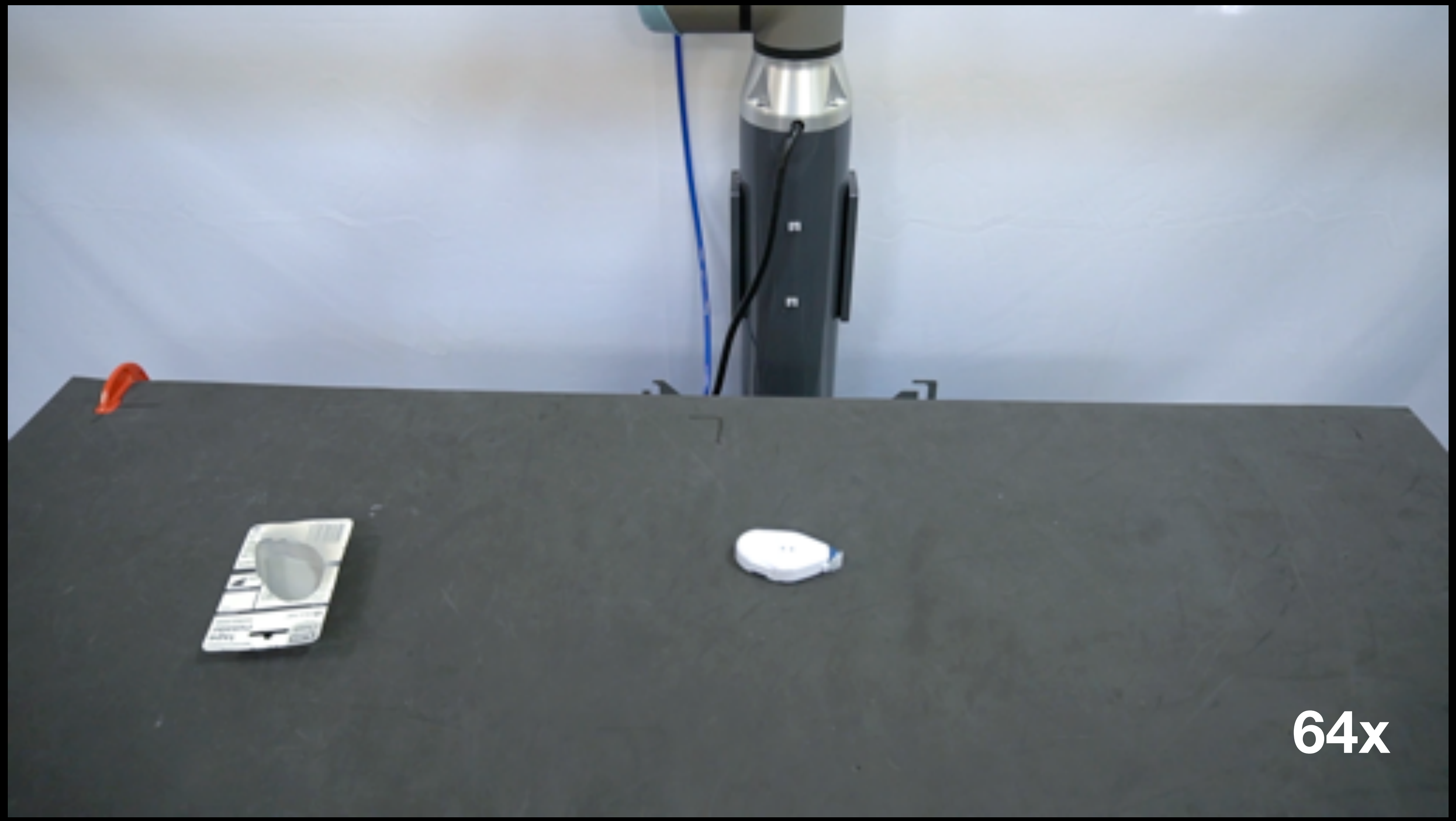

model trained on 2 kits: floss and tape

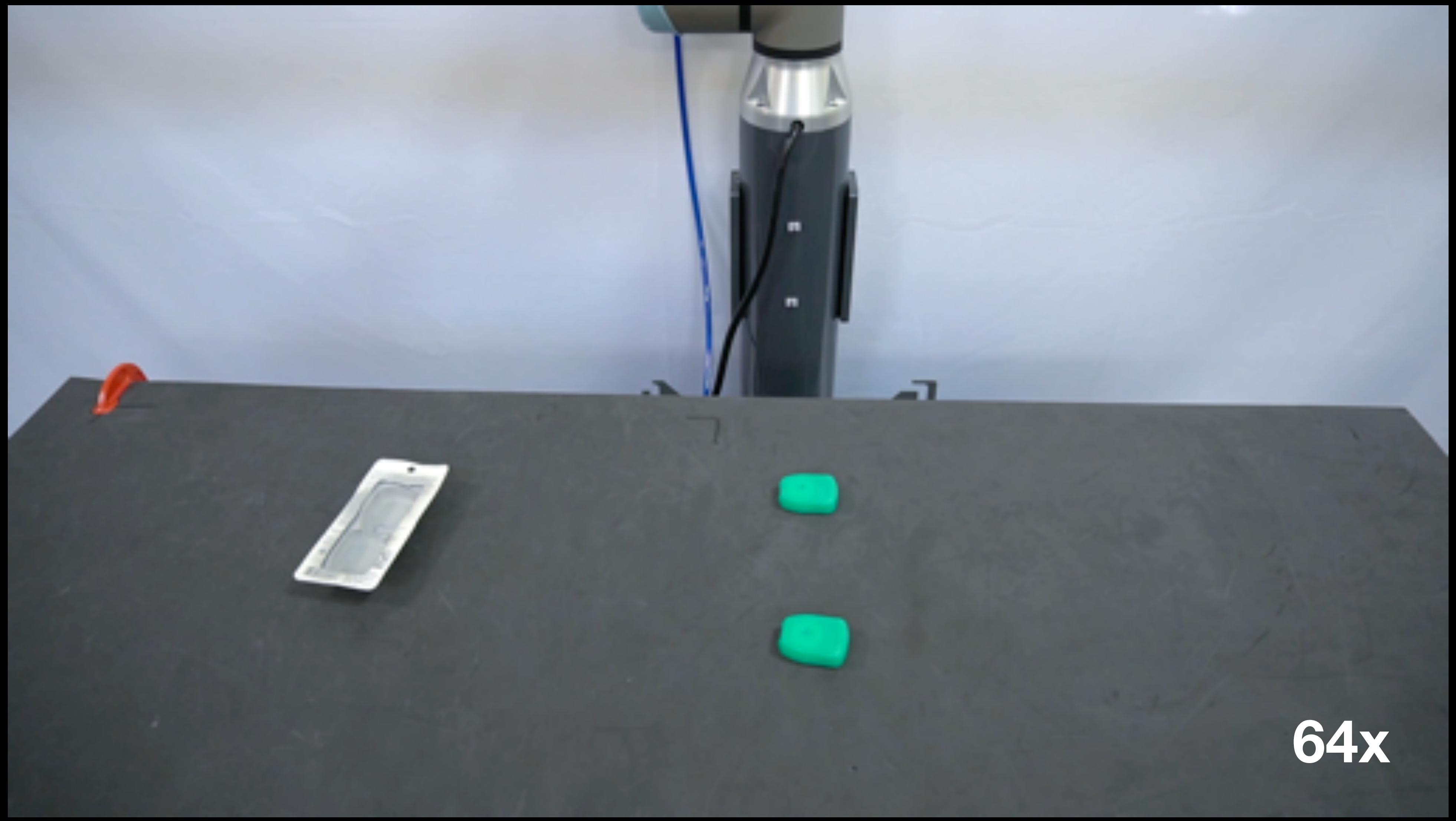
# Generalization to Novel Objects/Kits

# Generalization to Novel Objects/Kits



# Generalization to Novel Objects/Kits

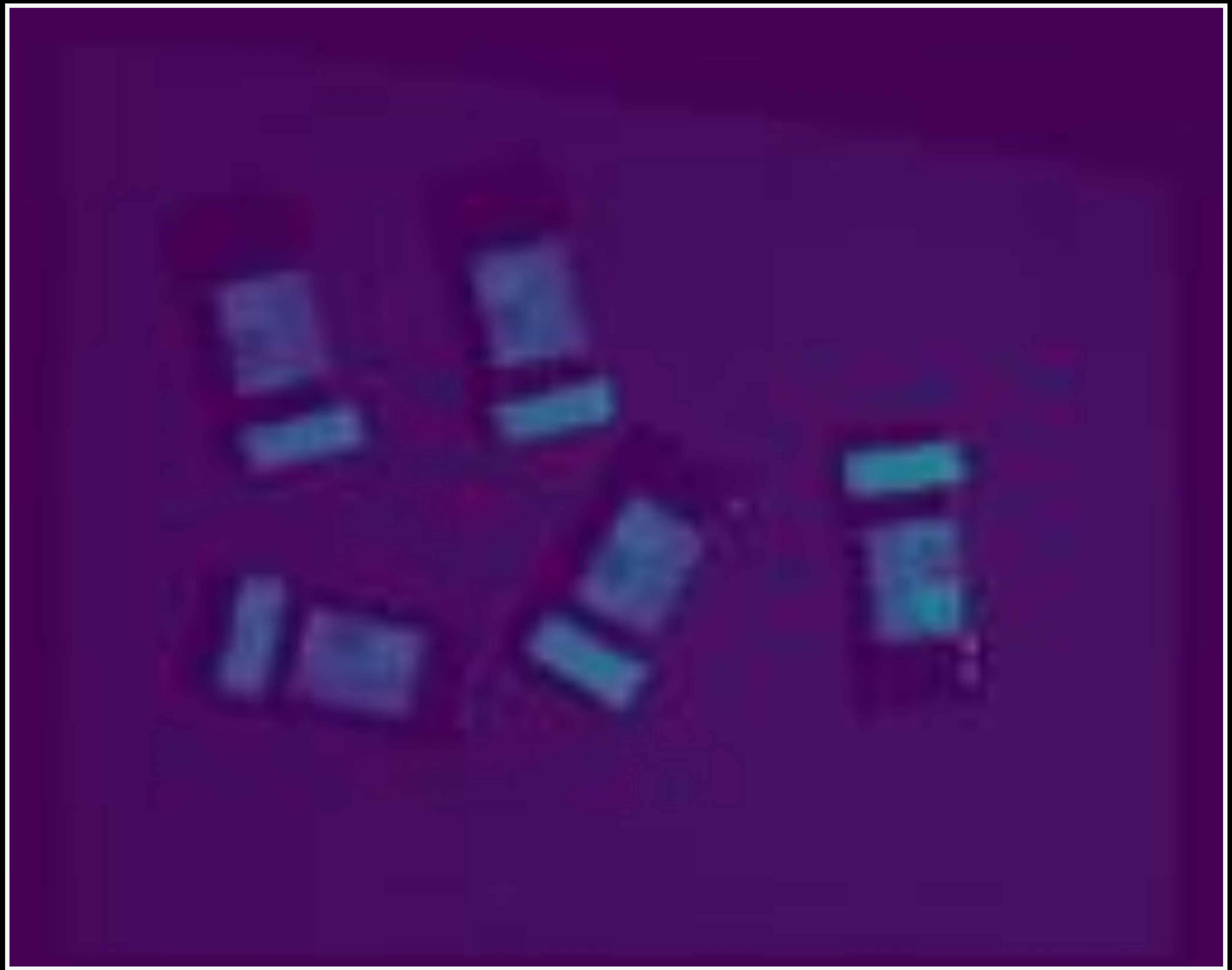

never before seen animals

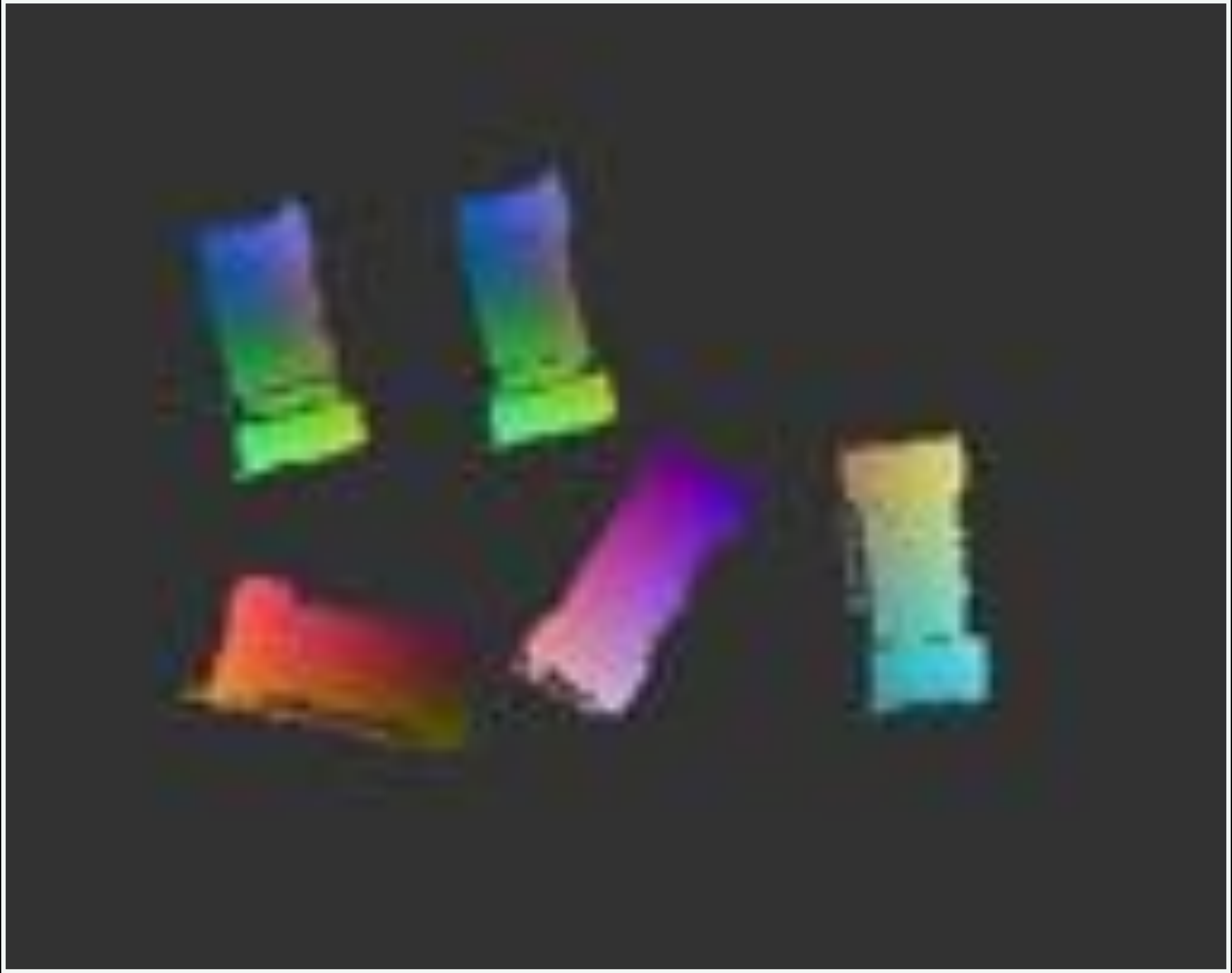
# Generalization to Novel Objects/Kits


never before seen animals

# What Has Form2Fit Learned?

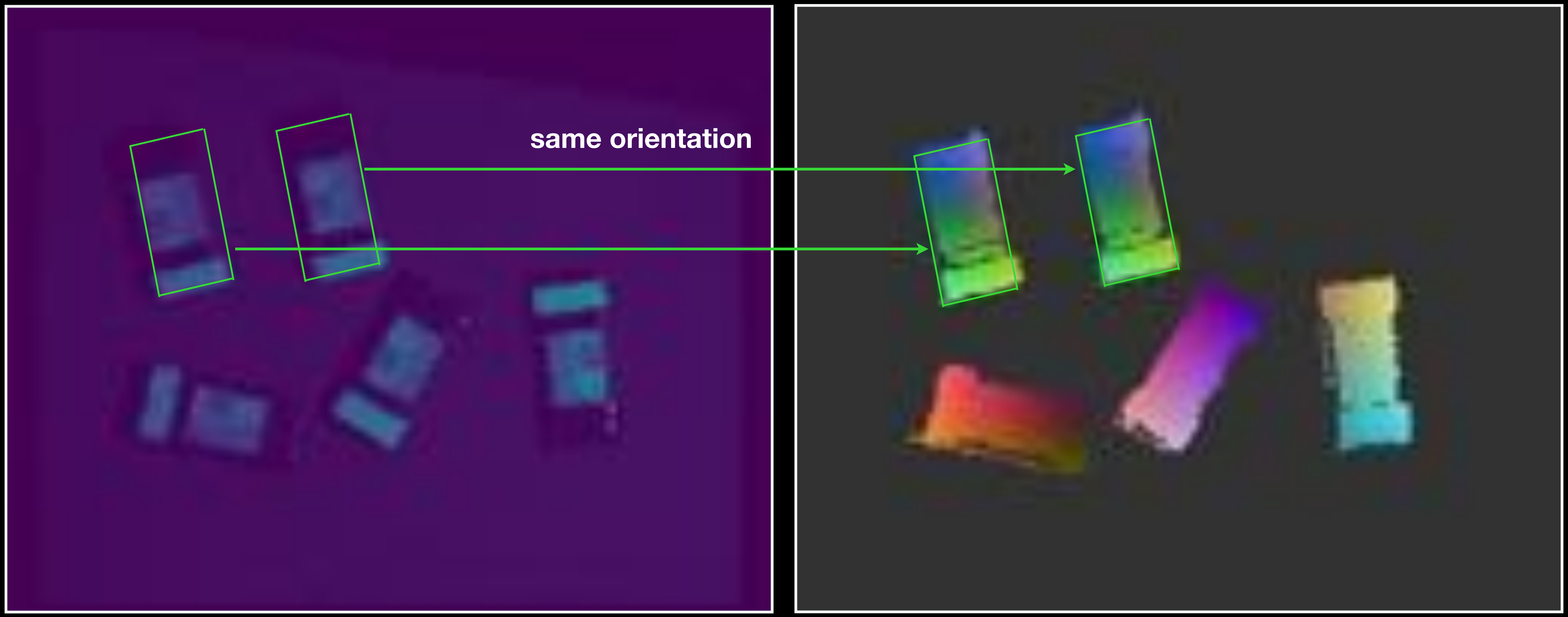
# Descriptor Visualization





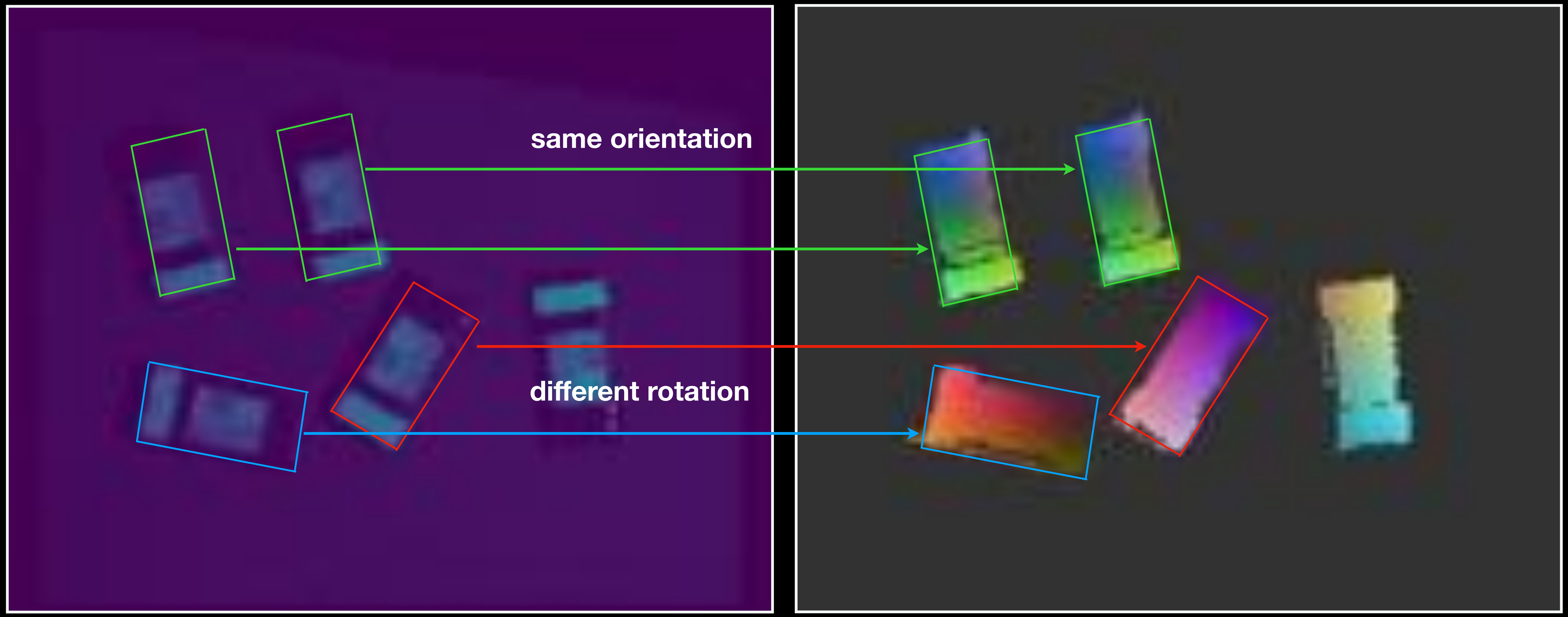
descriptors encode object orientation

# Descriptor Visualization



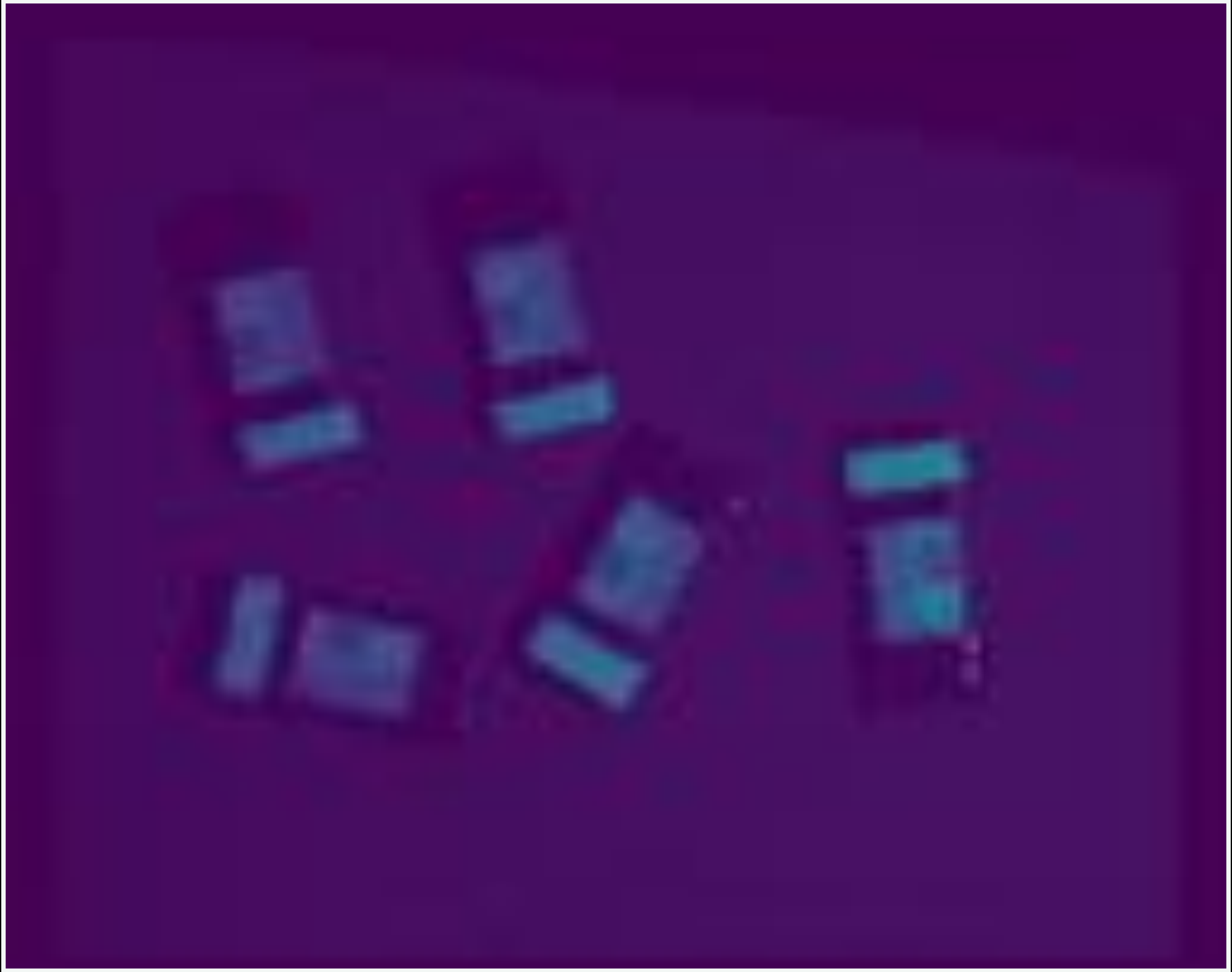

descriptors encode object orientation

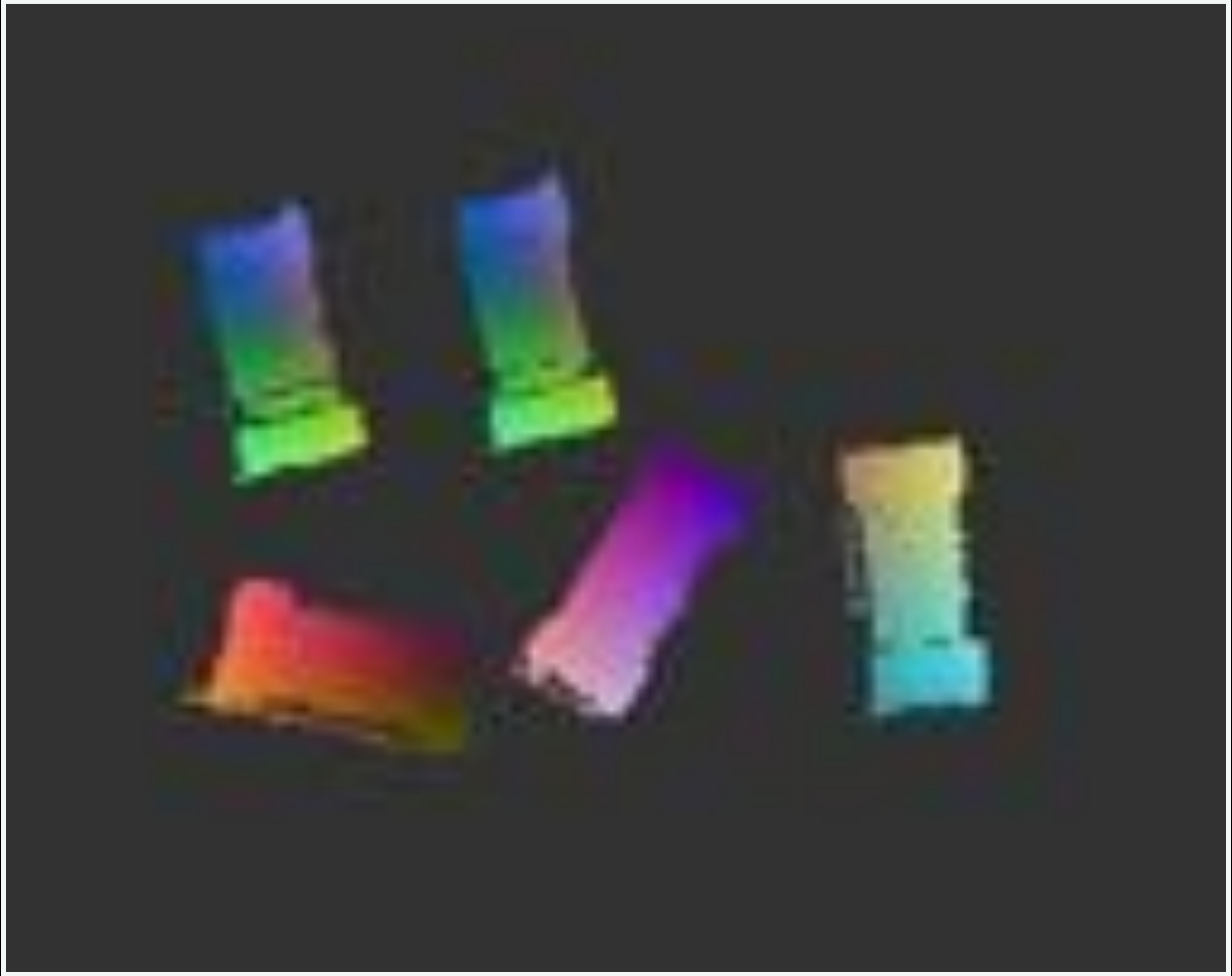
# Descriptor Visualization




descriptors encode object orientation

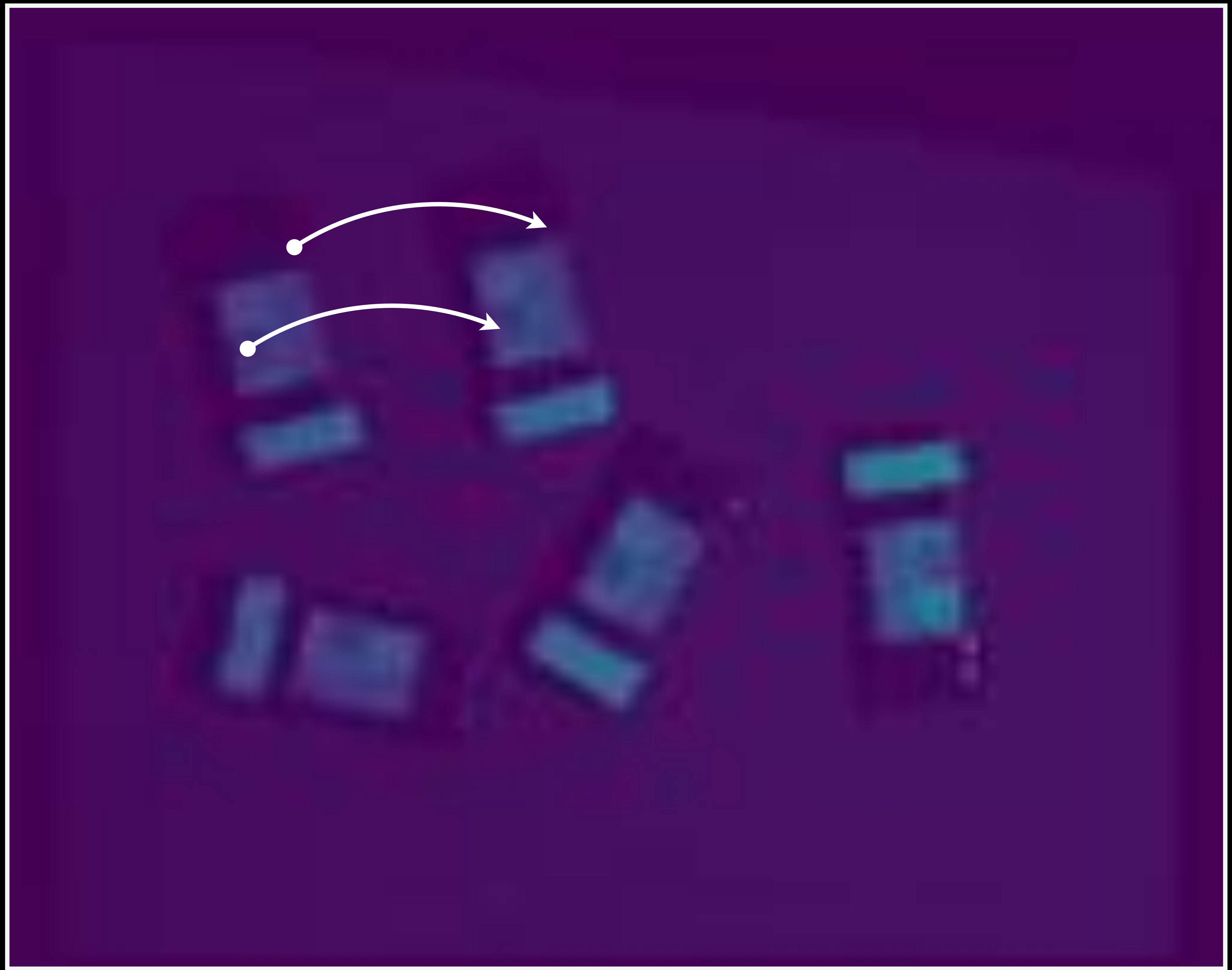
# Descriptor Visualization

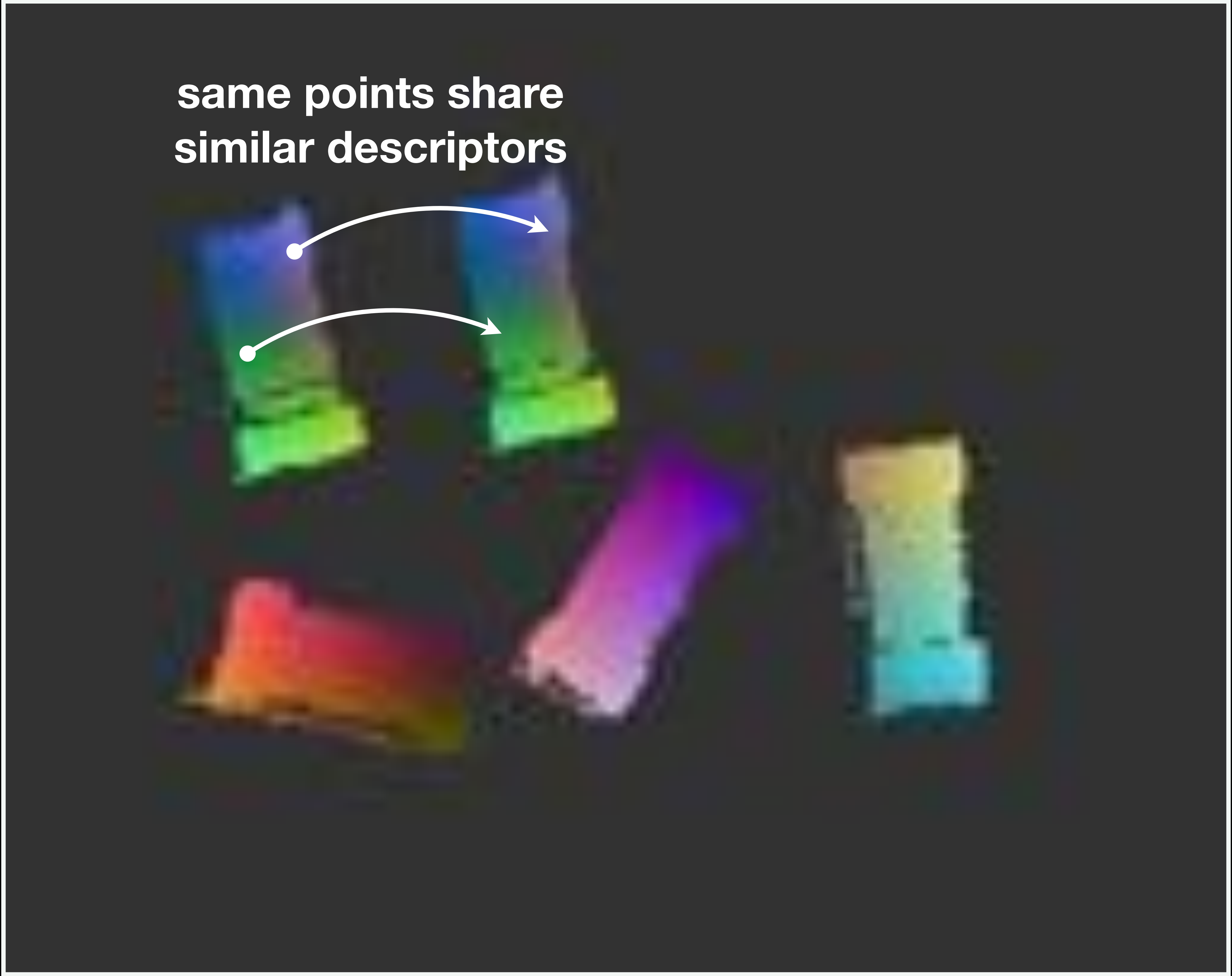




descriptors encode spatial correspondence

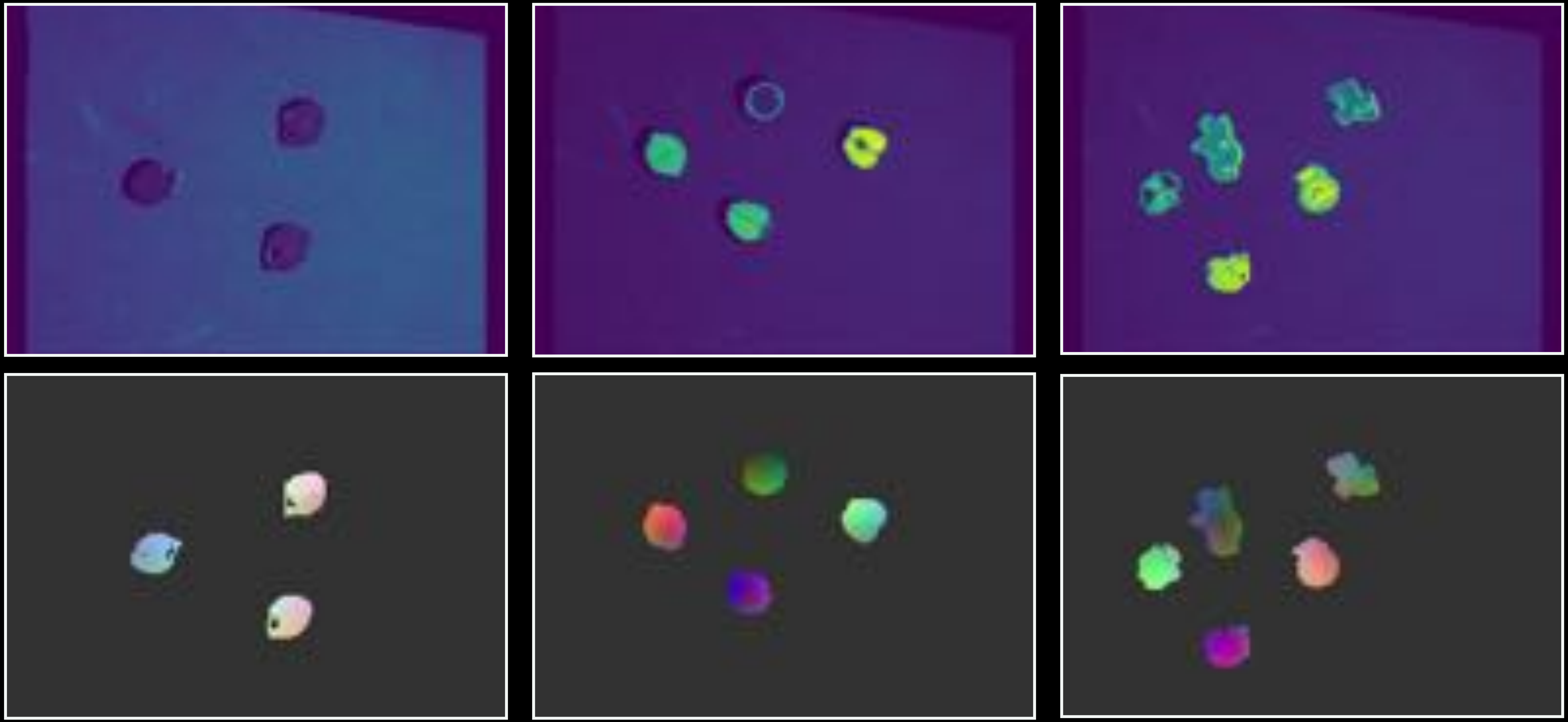
# Descriptor Visualization





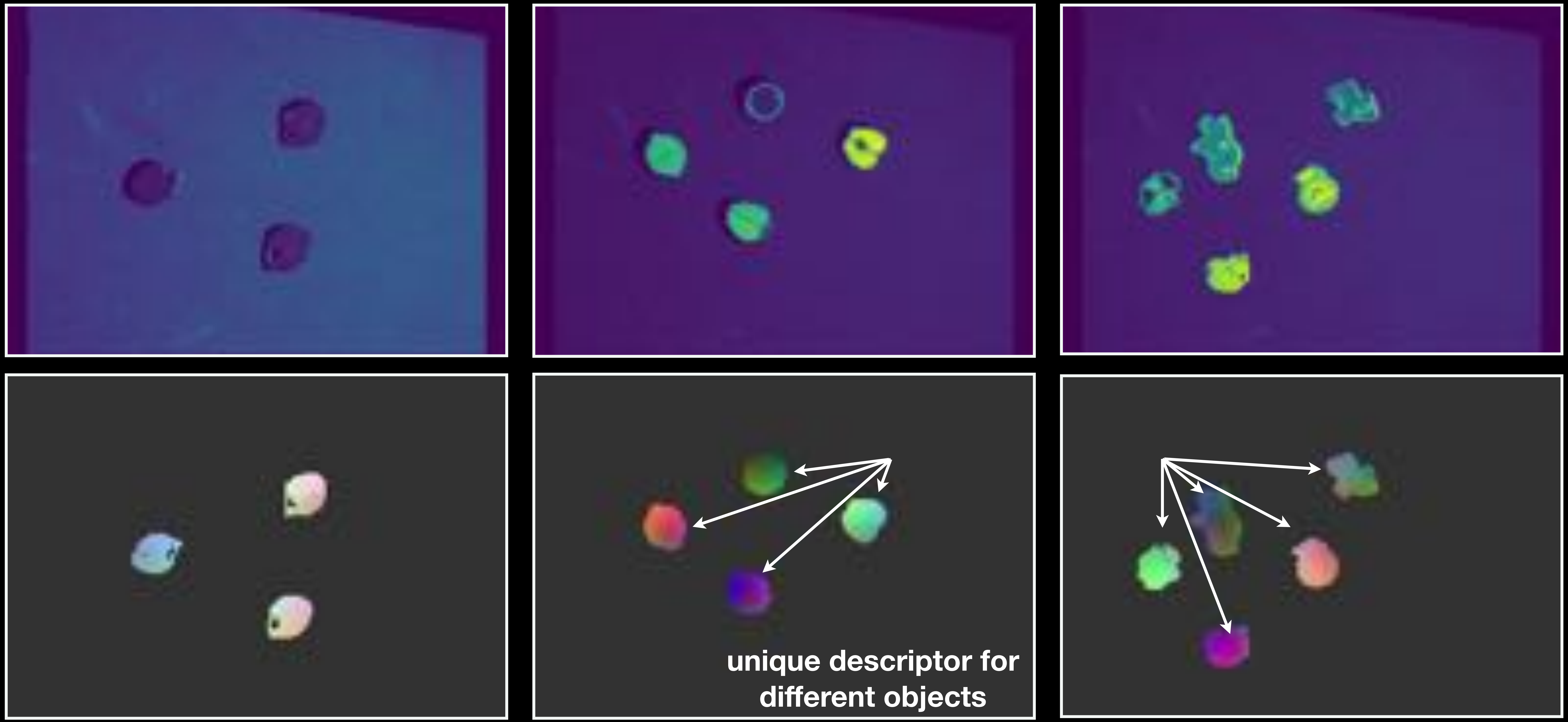

descriptors encode spatial correspondence

# Descriptor Visualization



descriptors encode object identity

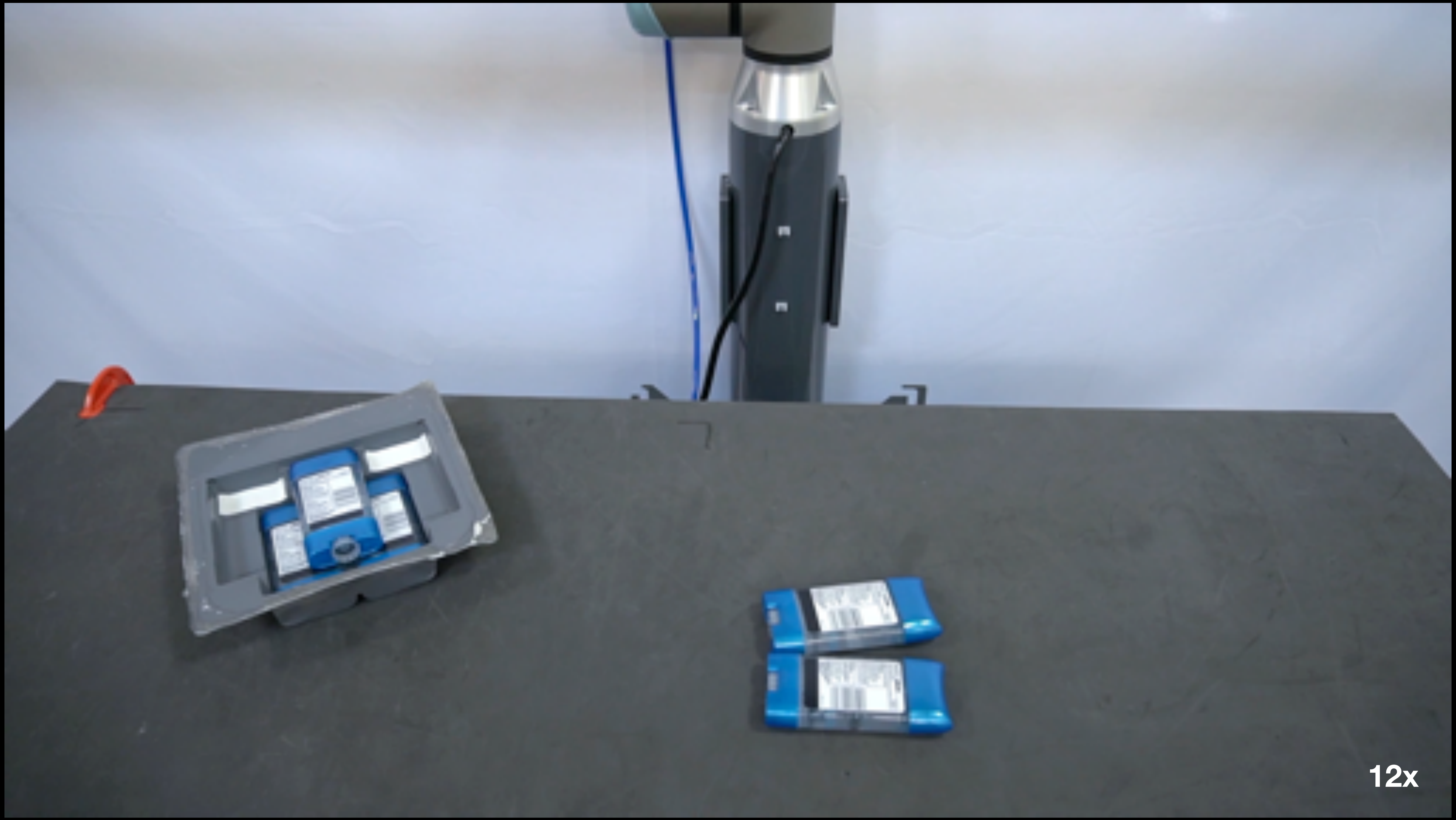
# Descriptor Visualization




descriptors encode object identity

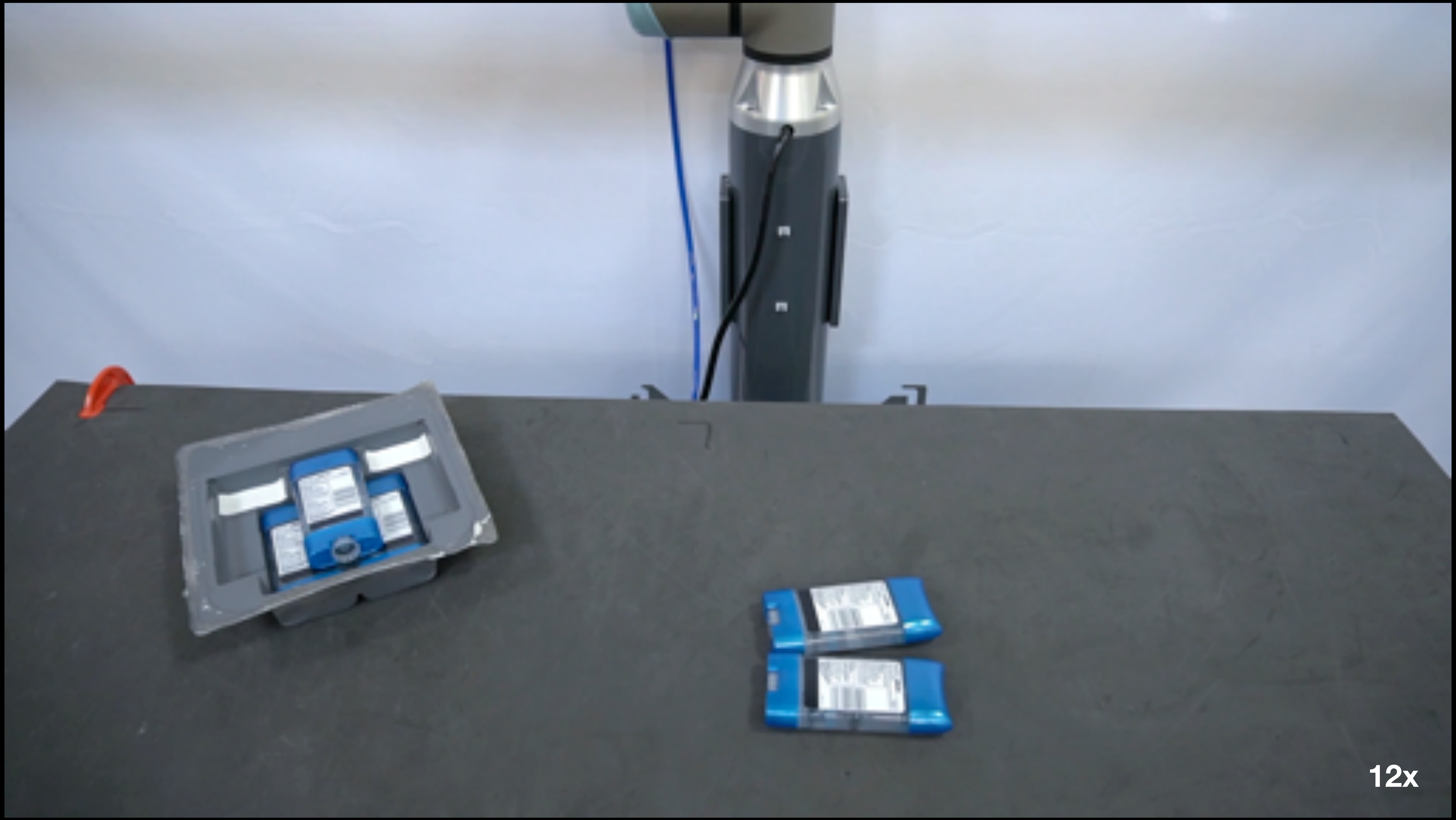
# Limitations & Future Work

# Typical Failure Case



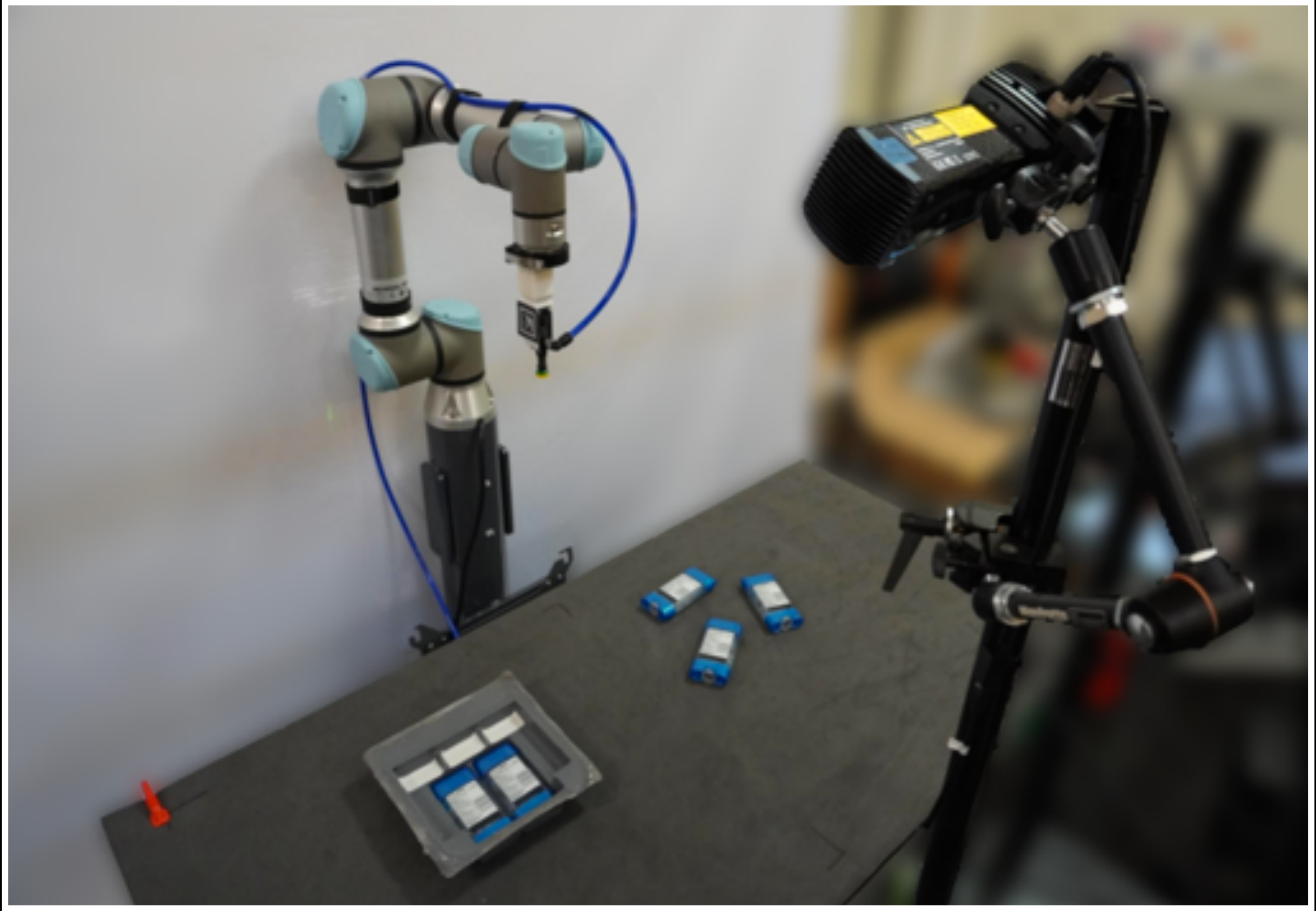

180º rotational flips

# Typical Failure Case

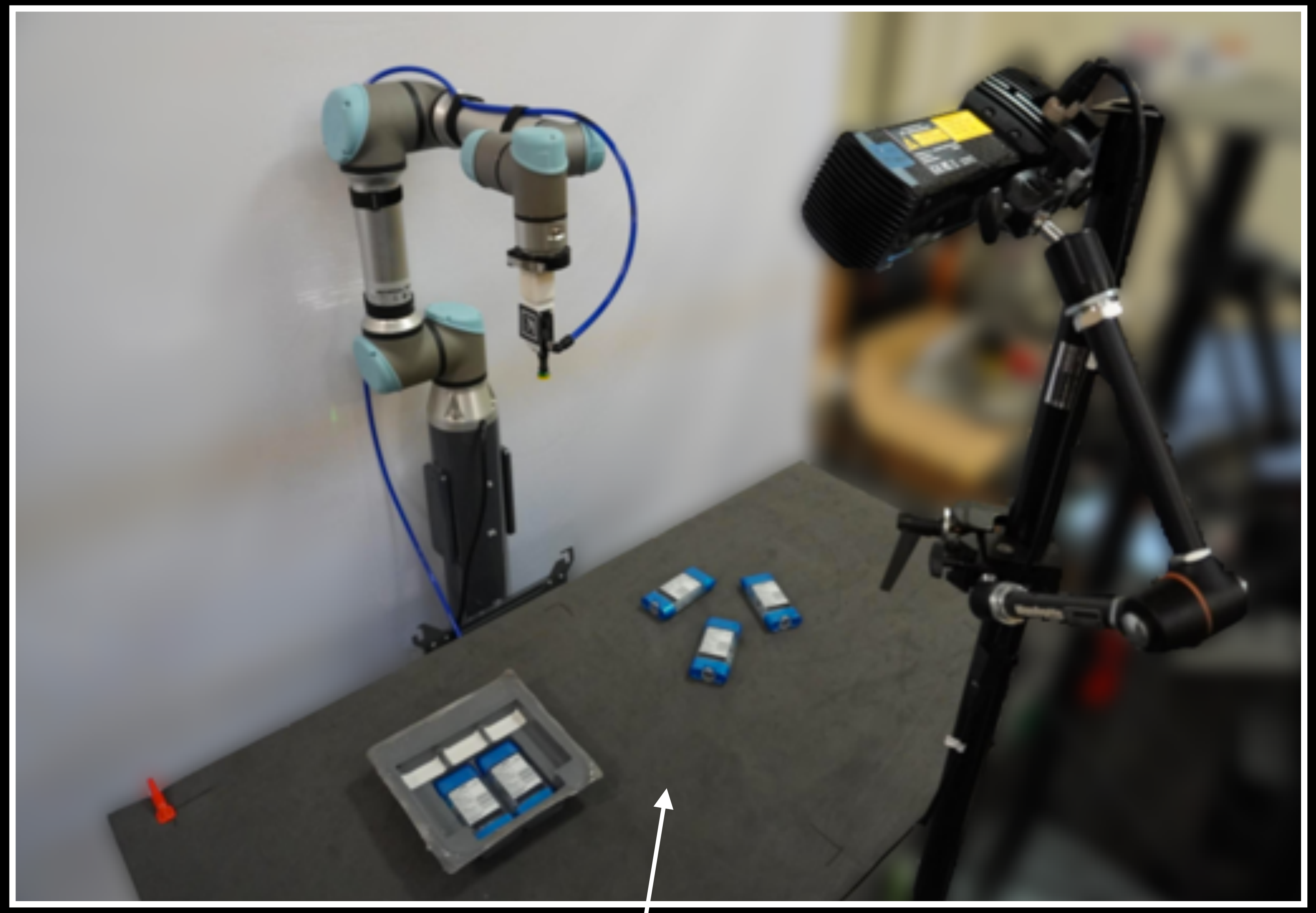



180º rotational flips

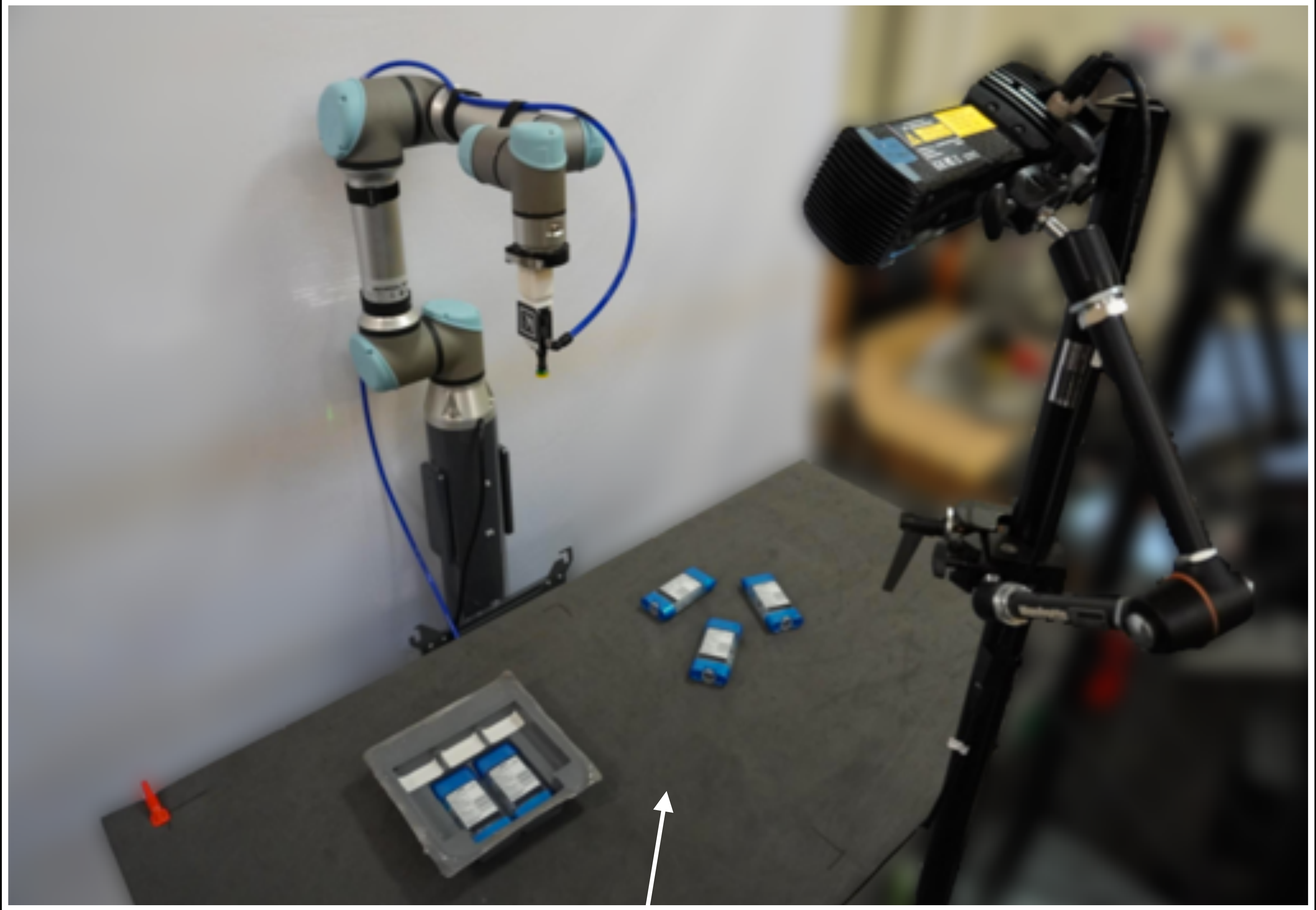
# Future Directions



# Future Directions

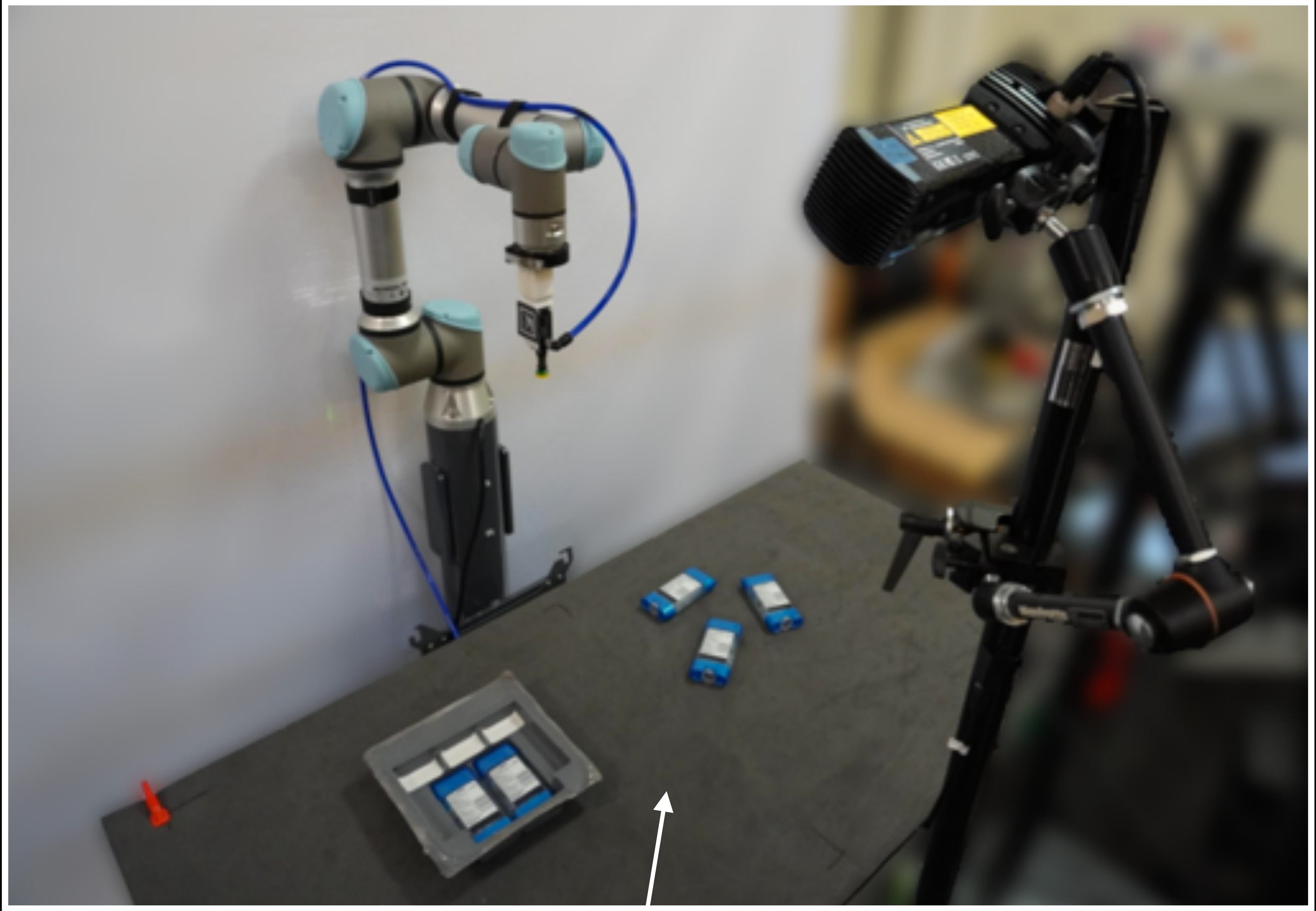


# Future Directions



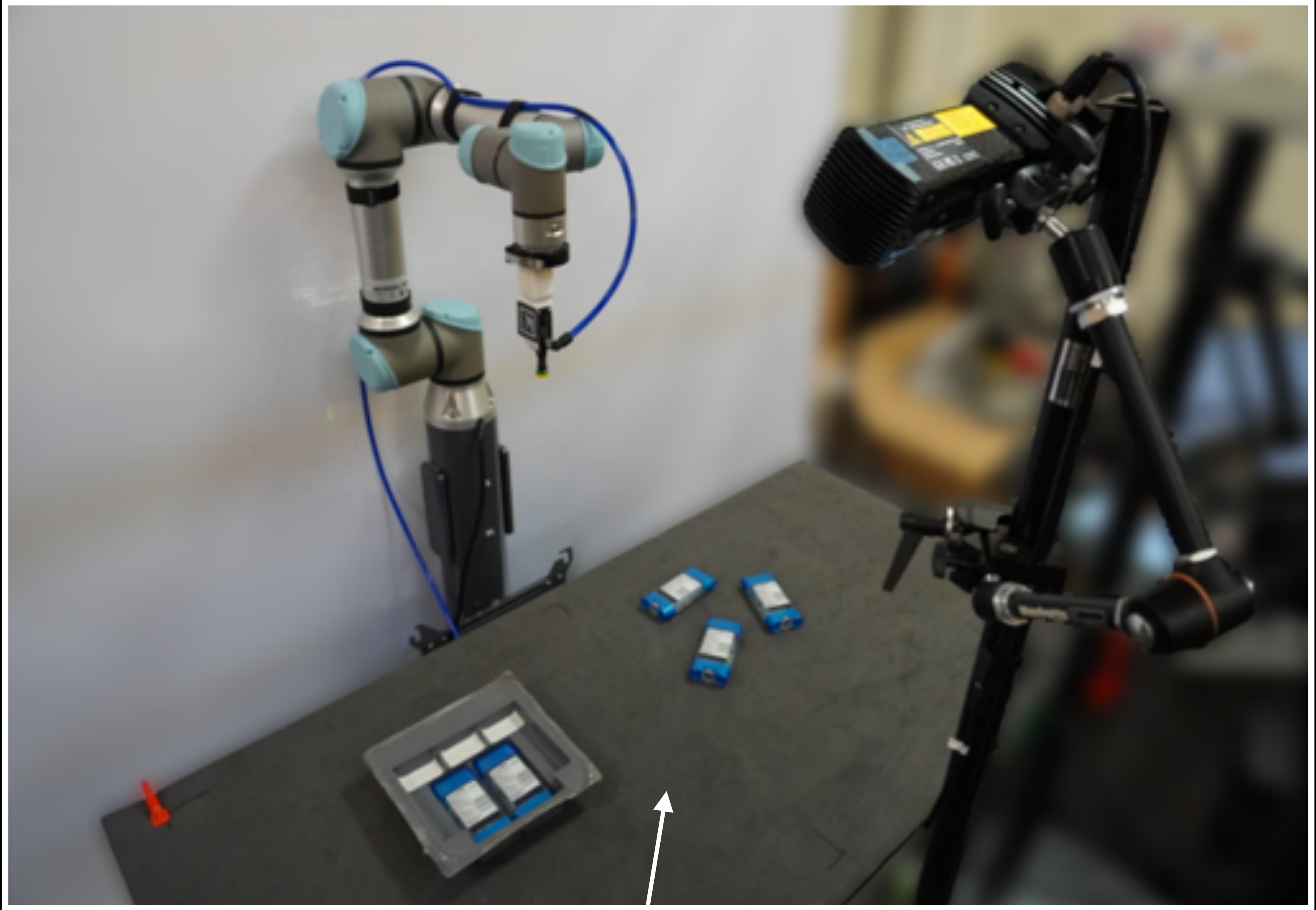
- restricted to planar manipulations

# Future Directions



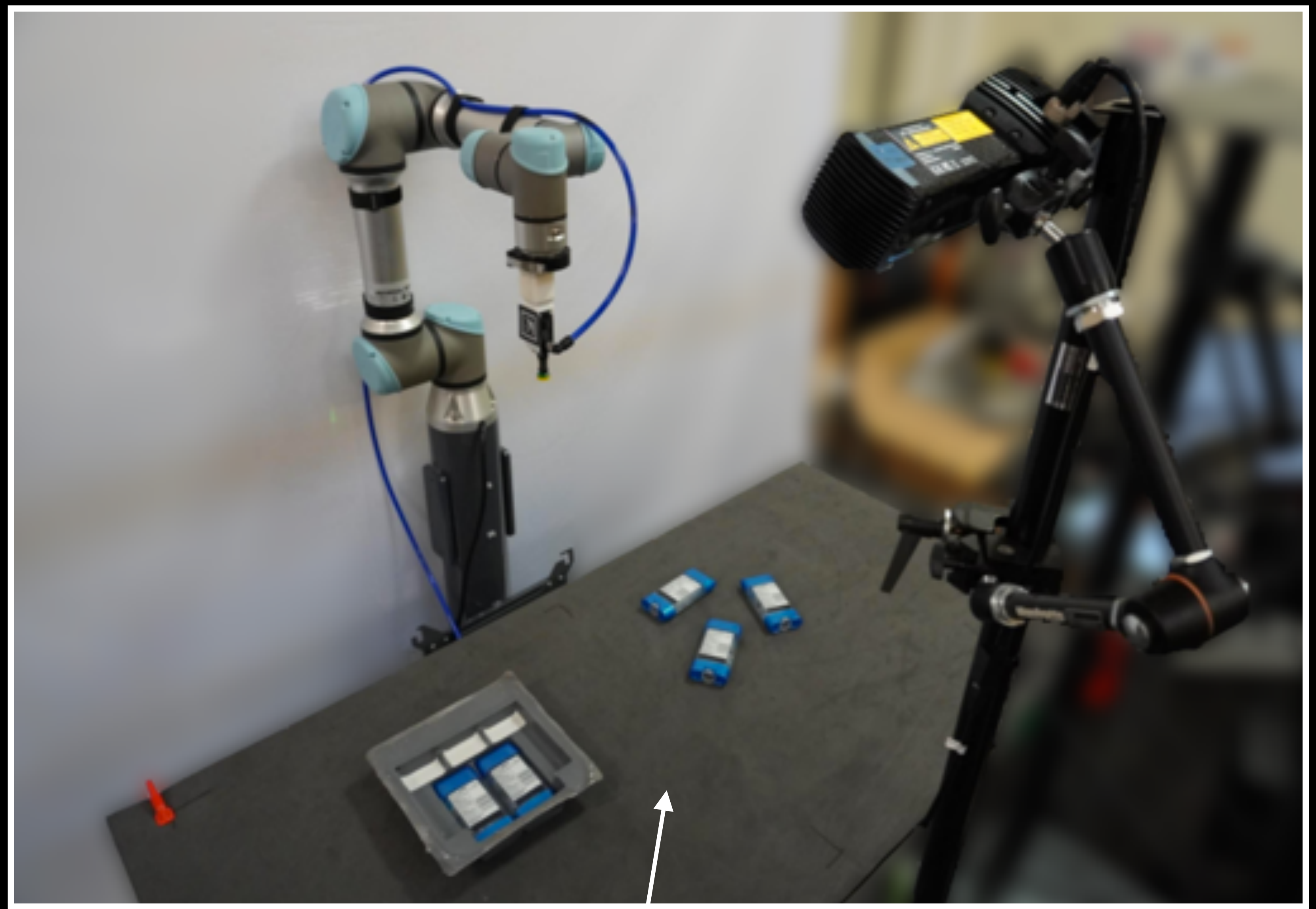
- restricted to planar manipulations

# Future Directions

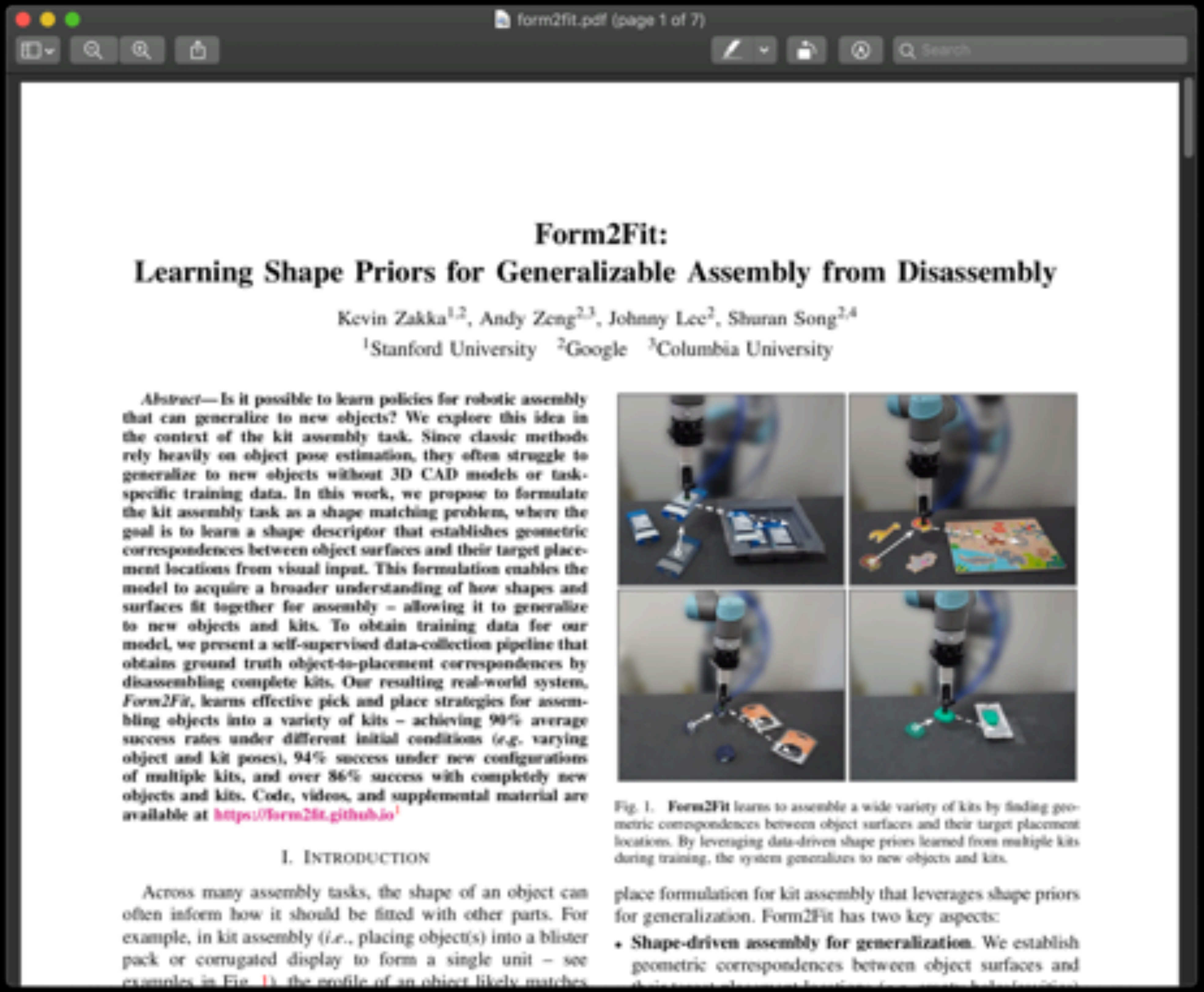



- restricted to planar manipulations
- can’t handle fully-transparent objects

# Future Directions



- restricted to planar manipulations
- can’t handle fully-transparent objects
- time-reversal currently restricted to quasi-static environment


form2fit.pdf (page 1 of 7)

Form2Fit:
Learning Shape Priors for Generalizable Assembly from Disassembly

Kevin Zakka[1,2], Andy Zeng[2,3], Johnny Lee[2], Shuran Song[2,4]
[1]Stanford University [2]Google [3]Columbia University

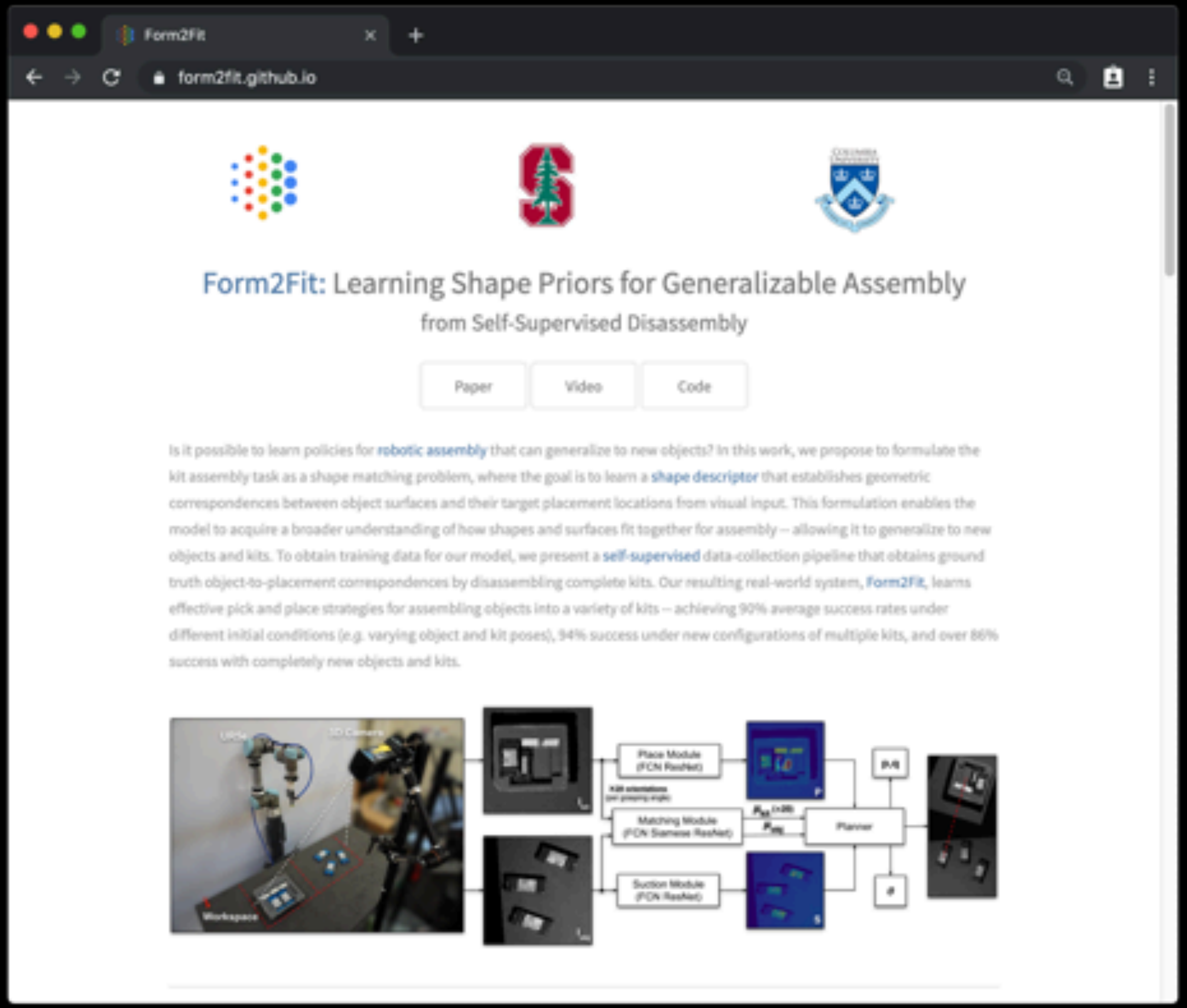
*Abstract*— Is it possible to learn policies for robotic assembly that can generalize to new objects? We explore this idea in the context of the kit assembly task. Since classic methods rely heavily on object pose estimation, they often struggle to generalize to new objects without 3D CAD models or task-specific training data. In this work, we propose to formulate the kit assembly task as a shape matching problem, where the goal is to learn a shape descriptor that establishes geometric correspondences between object surfaces and their target placement locations from visual input. This formulation enables the model to acquire a broader understanding of how shapes and surfaces fit together for assembly – allowing it to generalize to new objects and kits. To obtain training data for our model, we present a self-supervised data-collection pipeline that obtains ground truth object-to-placement correspondences by disassembling complete kits. Our resulting real-world system, *Form2Fit*, learns effective pick and place strategies for assembling objects into a variety of kits – achieving 90% average success rates under different initial conditions (*e.g.* varying object and kit poses), 94% success under new configurations of multiple kits, and over 86% success with completely new objects and kits. Code, videos, and supplemental material are available at https://form2fit.github.io[1]

Fig. 1. Form2Fit learns to assemble a wide variety of kits by finding geometric correspondences between object surfaces and their target placement locations. By leveraging data-driven shape priors learned from multiple kits during training, the system generalizes to new objects and kits.

I. Introduction

Across many assembly tasks, the shape of an object can often inform how it should be fitted with other parts. For example, in kit assembly (*i.e.*, placing object(s) into a blister pack or corrugated display to form a single unit – see examples in Fig. 1), the profile of an object likely matches

place formulation for kit assembly that leverages shape priors for generalization. Form2Fit has two key aspects:
- **Shape-driven assembly for generalization.** We establish geometric correspondences between object surfaces and


Form2Fit
form2fit.github.io

Form2Fit: Learning Shape Priors for Generalizable Assembly
from Self-Supervised Disassembly

Paper Video Code

Is it possible to learn policies for robotic assembly that can generalize to new objects? In this work, we propose to formulate the kit assembly task as a shape matching problem, where the goal is to learn a shape descriptor that establishes geometric correspondences between object surfaces and their target placement locations from visual input. This formulation enables the model to acquire a broader understanding of how shapes and surfaces fit together for assembly – allowing it to generalize to new objects and kits. To obtain training data for our model, we present a self-supervised data-collection pipeline that obtains ground truth object-to-placement correspondences by disassembling complete kits. Our resulting real-world system, Form2Fit, learns effective pick and place strategies for assembling objects into a variety of kits – achieving 90% average success rates under different initial conditions (e.g. varying object and kit poses), 94% success under new configurations of multiple kits, and over 86% success with completely new objects and kits.



For details, videos and code, visit:

https://form2fit.github.io