Form2Fit: Learning Shape Priors for Generalizable Assembly from Disassembly

Kevin Zakka, Andy Zeng, Johnny Lee, Shuran Song
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

Choi et. al.
Towards Flexible Assembly

state-of-the-art robo-Kitting solution

Choi et al.

3D Sensing Pose Estimation Grasping

CAD Model
Towards Flexible Assembly

state-of-the-art robo-kitting solution

• require prior knowledge and manual engineering
Towards Flexible Assembly

state-of-the-art robo-kitting solution

require prior knowledge and manual engineering
cannot quickly adapt to new objects and settings
Towards Flexible Assembly

state-of-the-art robo-kitting solution

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 $\rightarrow$ Shape Matching

- learns geometric shape descriptors
Key Ideas

Kit Assembly $\rightarrow$ 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 $\rightarrow$ 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
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

Kit Heightmap

Object Heightmap

Suction Network

*suction* network ingests object heightmap and outputs suction heatmap
Overview of Form2Fit

**Kit Heightmap**

**Object Heightmap**

**Suction Network**

The suction network ingests object heightmap and outputs suction heatmap.
Overview of Form2Fit

**Kit Heightmap**

**Object Heightmap**

**Suction Network**

*place* network ingests kit heightmap and outputs place heatmap
Overview of Form2Fit

Kit Heightmap -> Place Network

Object Heightmap -> Suction Network

*place* network ingests kit heightmap and outputs place heatmap
Overview of Form2Fit

place network ingests kit heightmap and outputs place heatmap
Overview of Form2Fit

Kit Heightmap → Place Network → Suction Network

Object Heightmap → Suction Network → corresponding pick and place candidates
Overview of Form2Fit

Kit Heightmap → Place Network → Suction Network

Object Heightmap → Suction Network

corresponding **pick** and **place** candidates
Overview of Form2Fit

Kit Heightmap → Place Network → Suction Network

Object Heightmap → Suction Network

corresponding pick and place candidates
Overview of Form2Fit

Kit Heightmap → Place Network → Suction Network

Object Heightmap → Suction Network

corresponding **pick** and **place** candidates
Overview of Form2Fit

Kit Heightmap -> Place Network

Object Heightmap -> Suction Network

corresponding pick and place candidates
Overview of Form2Fit

Kit Heightmap → Place Network → Suction Network → corresponding pick and place candidates

Object Heightmap → Suction Network
Overview of Form2Fit

Kit Heightmap → Place Network → Object Heightmap → Suction Network

Matching network ingests heightmaps and outputs descriptor maps.
Overview of Form2Fit

Kit Heightmap → Place Network → Matching Network → pixel-wise descriptors → Suction Network → Object Heightmap

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

Kit Heightmap → Place Network → Matching Network → pixel-wise descriptors

Object Heightmap → Suction Network

closer descriptor distances indicate better object-to-placement correspondences
Overview of Form2Fit

closer descriptor distances indicate better object-to-placement correspondences
Overview of Form2Fit

Kit Heightmap

Place Network

Matching Network

Object Heightmap

Suction Network

Descriptors are rotation-sensitive
Overview of Form2Fit

Kit Heightmap

× 20

Place Network

Matching Network

× 20

pixel-wise descriptors

Object Heightmap

Suction Network

descriptors are rotation-sensitive
Overview of Form2Fit

Kit Heightmap → Place Network → Matching Network

Object Heightmap → Suction Network

Pixel-wise descriptors

Descriptors are rotation-sensitive
Overview of Form2Fit

Kit Heightmap

Object Heightmap

Place Network

Matching Network

Suction Network

pixel-wise descriptors

descriptors are rotation-sensitive
Overview of Form2Fit

The planner integrates information to produce suction/place poses & end-effector rotation.
Overview of Form2Fit

Kit Heightmap → Place Network → Matching Network × 20 → Pixel-wise descriptors → Planner

Object Heightmap → Suction Network

Planner integrates information to produce suction/place **poses** & end-effector **rotation**
Overview of Form2Fit

Kit Heightmap

Place Network

× 20

Matching Network

descriptors × 20

Pixel-wise

Planner

Object Heightmap

Suction Network

p

q

 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

Kit Heightmap

Place Network

Matching Network

Pixel-wise descriptors × 20

Planner

Place Position

Pick Position

Object Heightmap

Suction Network

planner integrates information to produce suction/place poses & end-effector rotation
Overview of Form2Fit

Kit Heightmap → Place Network → Matching Network → Planner → Place Position

Object Heightmap → Suction Network → Planner → Pick Position

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 \((q, \theta)\)
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

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

same orientation
Descriptor Visualization

descriptors encode object orientation

same orientation

different rotation

descriptors encode object orientation
Descriptor Visualization

descriptors encode spatial correspondence
Descriptor Visualization

descriptors encode spatial correspondence

same points share similar descriptors
Descriptor Visualization

descriptors encode object identity
Descriptor Visualization

Descriptors encode object identity
Limitations & Future Work
Typical Failure Case

180° rotational flips
Typical Failure Case

180° 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
For details, videos and code, visit:

https://form2fit.github.io