Robotic Pick-and-Place of Novel Objects
The task

Pick → recognize → place **novel** objects

storage system  shipping box  competition footage
Team MIT-Princeton @ the Amazon Robotics Challenge

1st place in stowing task

Andy Zeng  Shuran Song  Kuan-Ting Yu  Elliott Donlon  Francois Hogan  Maria Bauza  Daolin Ma
Orion Taylor  Melody Liu  Eudald Romo  Nima Fazeli  Ferran Alet  Nikhil Dafle  Rachel Holladay
Isabella Morona  Prem Qu Nair  Druck Green  Ian Taylor  Weber Liu  Thomas Funkhouser Alberto Rodriguez
Team MIT-Princeton @ the Amazon Robotics Challenge

1st place in stowing task

"The Beast from the East"

setup

competition footage
The challenges

- Dense clutter in tabletop bin/box scenario
The challenges

- Dense clutter in tabletop bin/box scenario

Clutter in prior work with novel objects
The challenges

- **Dense clutter in tabletop bin/box scenario**

Rethink clutter:
- Objects not only tightly packed, but also tossed and stacked on top of each other
- Objects in corners and on bin edges

Clutter in prior work with novel objects
The challenges

- Dense clutter in tabletop bin/box scenario
The challenges

- Dense clutter in tabletop bin/box scenario
- Novel objects - no extra data collection or re-training time
The challenges

- **Dense** clutter in tabletop **bin/box** scenario
- **Novel** objects - no extra data collection or re-training time

~485k new products added to catalogue daily worldwide
The challenges

- Dense clutter in tabletop bin/box scenario
- Novel objects - no extra data collection or re-training time

~485k new products added to catalogue daily worldwide
The challenges

- Dense clutter in tabletop bin/box scenario
- Novel objects - no extra data collection or re-training time
The challenges

- Dense clutter in tabletop bin/box scenario
- Novel objects - no extra data collection or re-training time
  - Grasping: use model-free methods instead of model-based
The challenges

- **Dense** clutter in tabletop bin/box scenario
- **Novel** objects - no extra data collection or re-training time
  - Grasping: use model-free methods instead of model-based

![Model-based grasping](image1.png)

![Model-free grasping](image2.png)

- Pose estimation  →  Grasp planning  →  Visual data  →  Grasp planning
The challenges

- **Dense** clutter in tabletop bin/box scenario
- **Novel** objects - no extra data collection or re-training time
  - Grasping: use model-free methods instead of model-based
  - Recognition: use product images available at test time
The challenges

- **Dense** clutter in tabletop **bin/box** scenario
- **Novel** objects - no extra data collection or re-training time
  - Grasping: use model-free methods instead of model-based
  - Recognition: use product images available at test time
The challenges

- **Dense** clutter in tabletop **bin/box** scenario
- **Novel** objects - no extra data collection or re-training time
  - Grasping: use model-free methods instead of model-based
  - Recognition: use product images available at test time
  - Can include anomalies (e.g. objects with missing depth)
The challenges

- **Dense** clutter in tabletop **bin/box** scenario
- **Novel** objects - no extra data collection or re-training time
  - Grasping: use model-free methods instead of model-based
  - Recognition: use product images available at test time
  - Can include anomalies (e.g. objects with missing depth)
The challenges

- **Dense** clutter in tabletop **bin/box** scenario
- **Novel** objects - no extra data collection or re-training time
  - Grasping: use model-free methods instead of model-based
  - Recognition: use product images available at test time
  - Can include anomalies (e.g. objects with missing depth)
The challenges

- **Dense** clutter in tabletop *bin/box* scenario
- **Novel** objects - no extra data collection or re-training time
  - Grasping: use model-free methods instead of model-based
  - Recognition: use product images available at test time
  - Can include anomalies (e.g. objects with missing depth)
- **Speed and efficiency requirements**
  - Human performance (industrial standard): 600 picks per hour
The challenges

- **Dense** clutter in tabletop bin/box scenario
- **Novel** objects - no extra data collection or re-training time
  - Grasping: use model-free methods instead of model-based
  - Recognition: use product images available at test time
  - Can include anomalies (e.g. objects with missing depth)
- **Speed and efficiency requirements**
  - Human performance (industrial standard): 600 picks per hour
  - \( \leq 10 \text{ seconds} \)
Our approach

- “Pick first, ask questions later”
  - Grasping (model-free)
  - Recognition (using product images, no re-training)
    - **Very challenging**: objects in isolation helps

“The Beast from the East”
setup
competition footage
Grasping overview

Input: multi-view RGB-D images
Grasping overview

Input: multi-view RGB-D images
Grasping overview

Input: multi-view RGB-D images

Output: grasp proposals and confidence scores for 4 primitive grasping behaviors:

suction down
Grasping overview

Input: multi-view RGB-D images

Output: grasp proposals and confidence scores for 4 primitive grasping behaviors:

- suction down
- suction side
Grasping overview

Input: multi-view RGB-D images

Output: grasp proposals and confidence scores for 4 primitive grasping behaviors:

- suction down
- suction side
- grasp down
Grasping overview

Input: multi-view RGB-D images

Output: dense grasp proposals and affordance scores for 4 primitive grasping behaviors:
suction down  suction side  grasp down  flush grasp
Dense pixel-wise affordances with FCNs

Input RGB-D images → fully convolutional ResNet-50

RGB-D heightmaps → predicts horizontal grasp affordances

✓ suction down
X suction side

✓ grasp down
X flush grasp
Generalization from hardware capabilities

- High-powered deployable suction
- Actuated spatula
Recognition: cross domain image matching (training)

product images

observed images

\[ \ell^2 \text{ distance ratio loss} \]

match?

softmax loss for K-Net only
Recognition: cross domain image matching (testing)
Recognition: cross domain image matching (testing)
Recognition: cross domain image matching (testing)
Recognition: cross domain image matching (testing)

known

novel

Pre-trained ImageNet features

feature embedding

match!

input
Paper and code are available:

arc.cs.princeton.edu