MITSUBISHI ELECTRIC
Changes for the Better

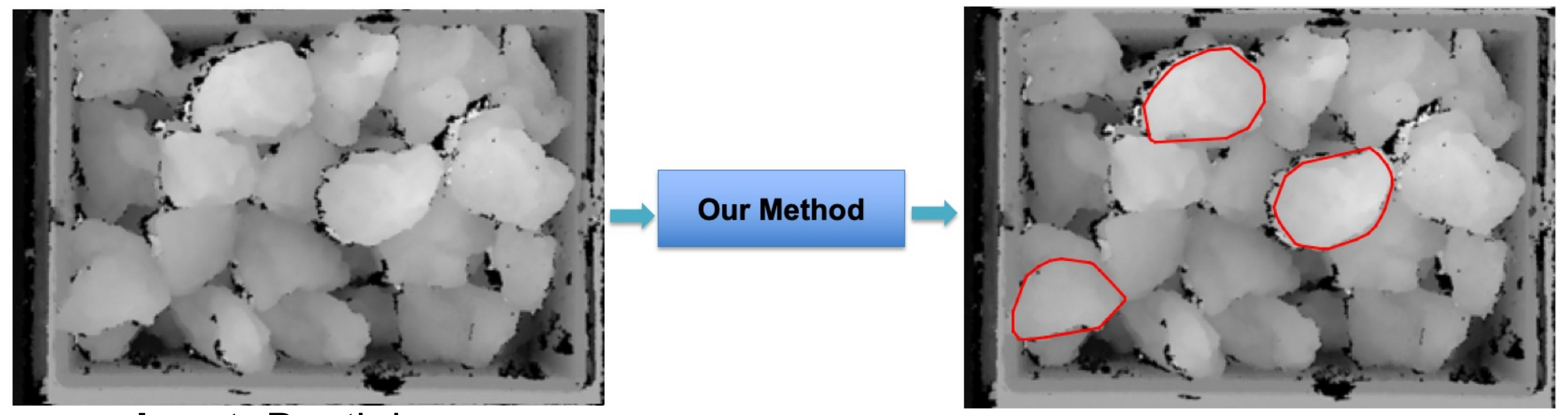
MITSUBISHI ELECTRIC RESEARCH LABORATORIES, INC

Discriminative 3D Shape Modeling for Few-Shot Instance Segmentation

Anoop Cherian   Siddarth Jain   Tim K. Marks   Alan Sullivan
Mitsubishi Electric Research Labs (MERL), Cambridge, MA
{cherian, sjain, tmarks, sullivan} @merl.com

1. Problem: Instance Segmentation

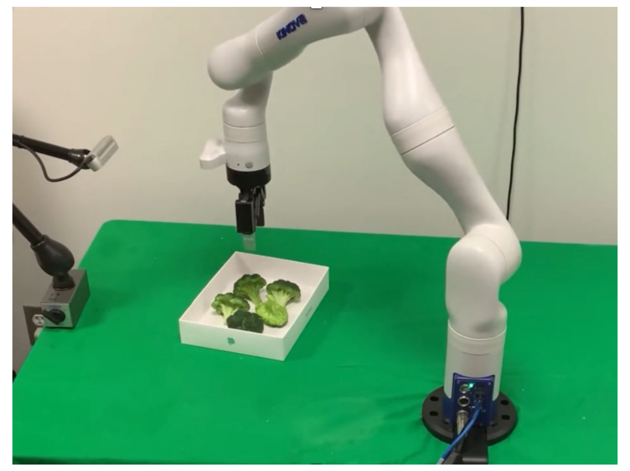
Given a depth image consisting of multiple non-rigid instances of an approximately convex object, our task is to predict the segmentation masks for the instances. We assume to have access to only a few annotated training examples.



Input: Depth image (with chicken nugget instances)

Output: Predicted instance segmentations for some instances

Why not predict all instance segmentations? We consider a **robotic bin-picking setting** where the robot picks one instance at a time. Thus, segmenting some pickable instances is sufficient, and once those are picked up, we could iteratively apply the method to the remaining instances.



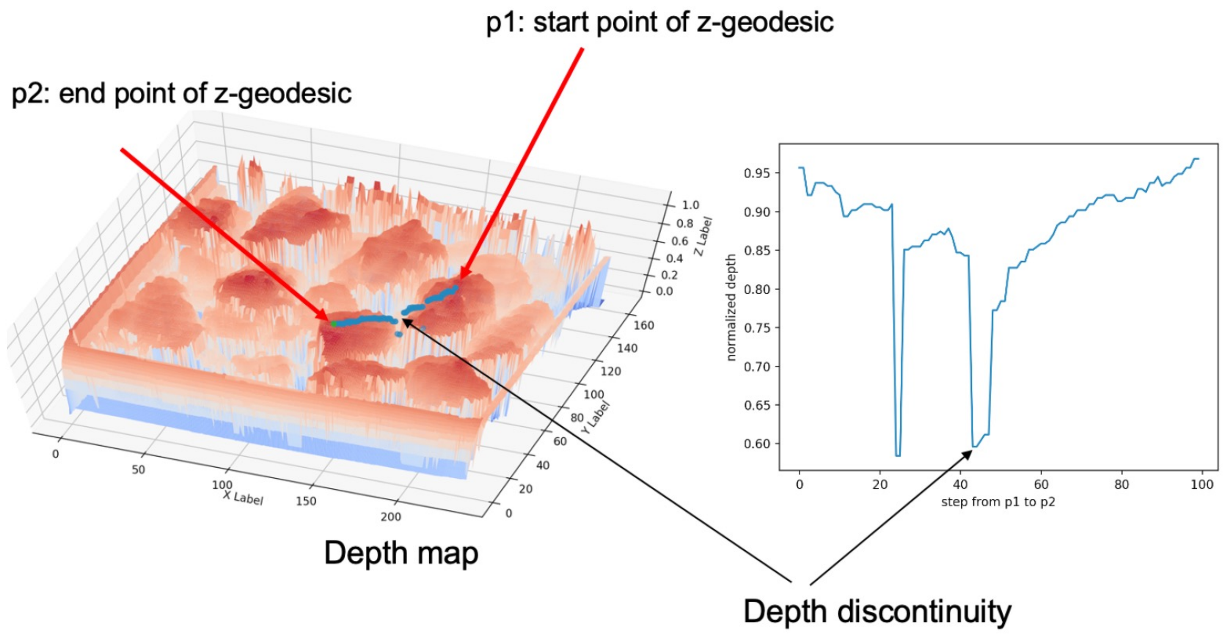
2. Contributions

- We present a simple approach – ***discriminative 3D shape modeling using surface geodesics*** – for instance segmentation in depth images
- Our method needs only **a few annotated examples** to train our model, is **very fast** to train and predict (10 min to train and 0.1s to predict using a CPU)
- Our method shows **large-margin improvements** in instance segmentation performance on our challenging Food-Items dataset.

3. Prior Works

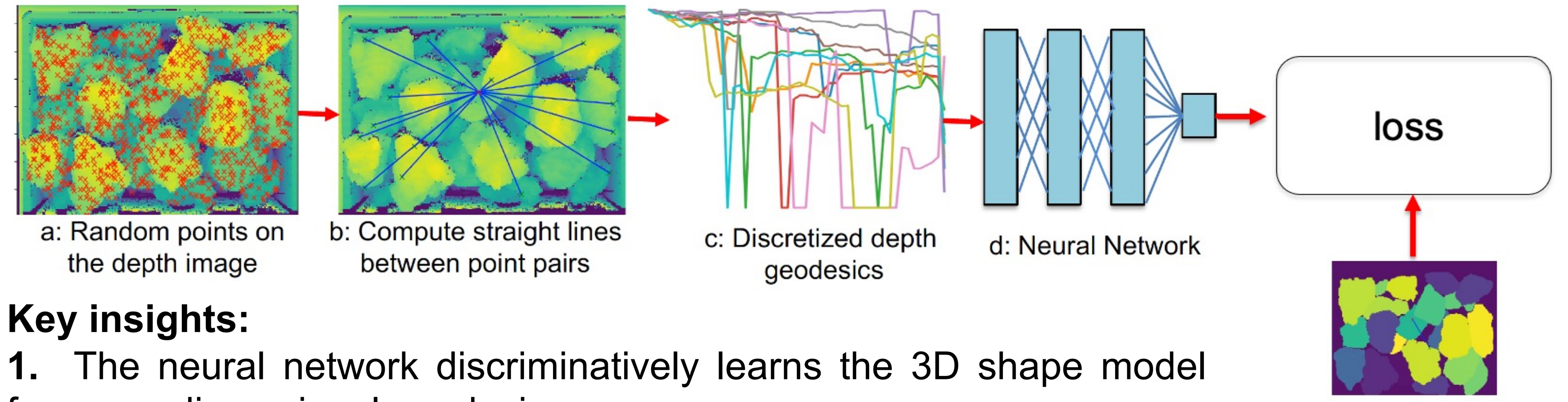
- **Supervised Methods:** Mask-RCNN and related methods
  - ➢ Needs many annotated training examples, could be expensive or challenging to gather in changing real-world conditions.
- **Unsupervised Methods**: InSeGAN, Slot Attention, IODINE, etc.
  - ➢ Needs a large-sized (e.g., thousands of images) of unannotated data
  - ➢ May be time consuming
- Our method needs only few-shot annotated examples to train our model.

4. Discriminative 3D Shapes Using Depth Geodesics



Key idea: Take any two points on the depth surface and draw a surface geodesic/curve between the points. Can we learn the subtilities in the geodesic as it traverses on the instance's surface to predict if its end-points belong to the same instance or multiple instances? We assume the 3D shape is approximately convex.

4a. Training Pipeline

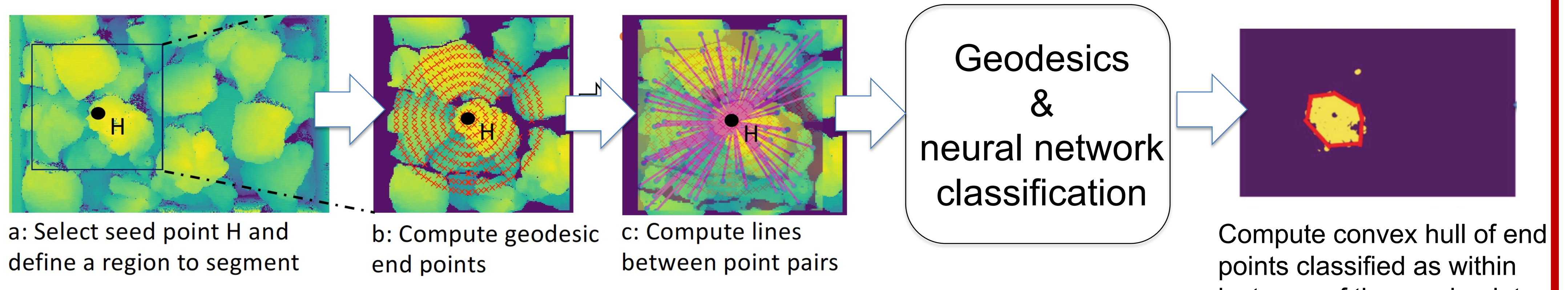


Key insights:

1. The neural network discriminatively learns the 3D shape model from one-dimensional geodesics.

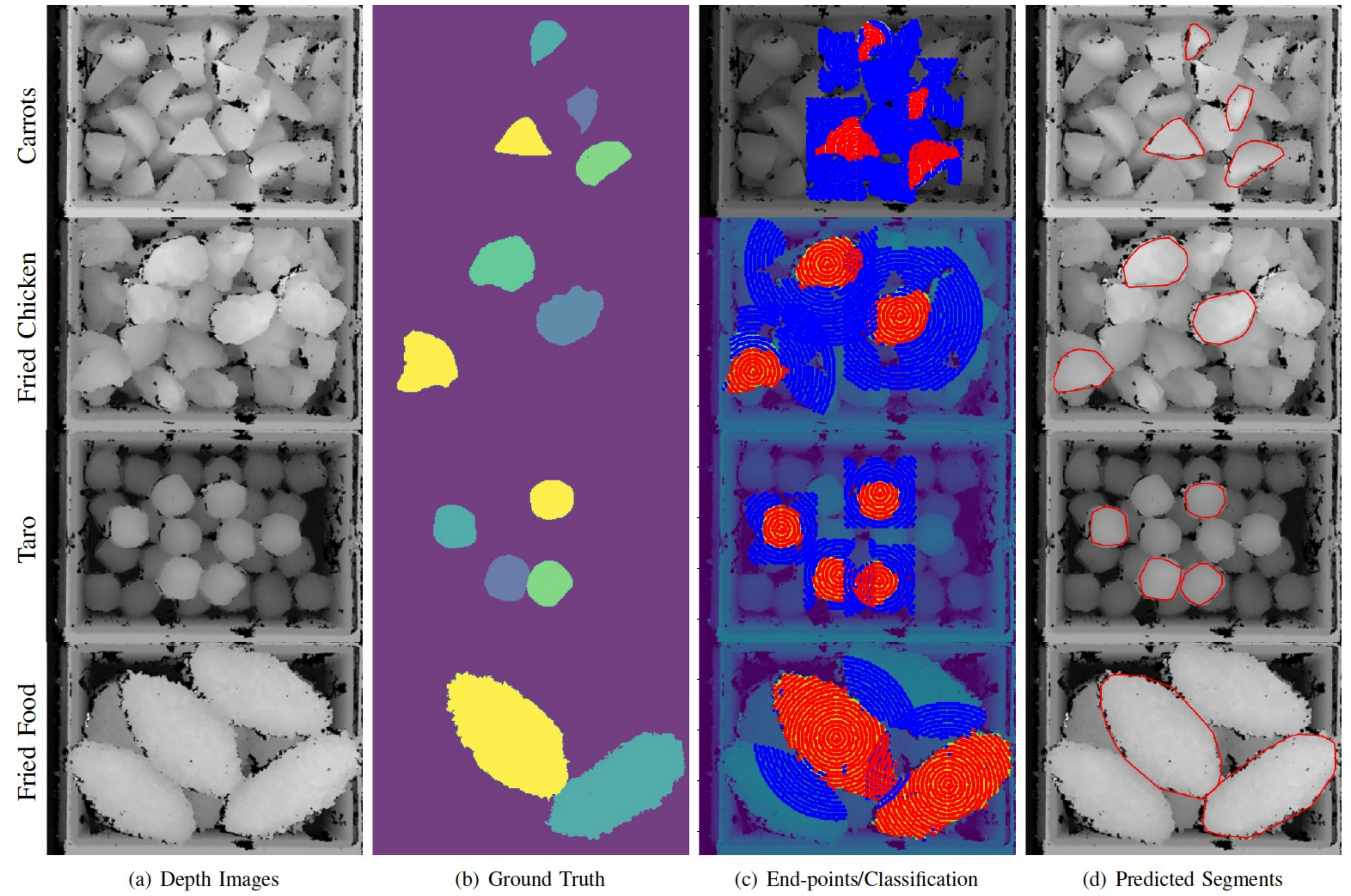
2. We may produce a very large training set of surface geodesics with very few instance annotations by considering all pairwise randomly selected points.

4b. Inference Pipeline



We use the trained neural network for classifying the geodesics.

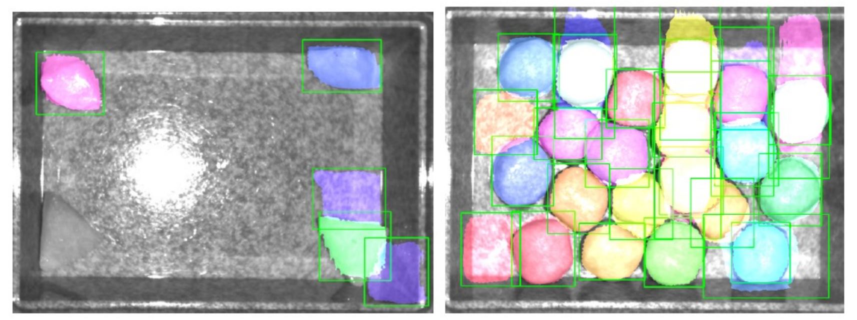
6. Experiments: Food-Items Dataset



- Four object categories
- Captured using Ensenso depth camera
- 20 annotated images per category
- 2–30 instances per image
- Used 3 images for training per category
- Used a simple neural network with 5 MLP layers.
- Used mIoU for evaluation using ground truth annotations.

Method	Carrot	Taro	Chicken	Fried Food
Ours	**0.807**	**0.844**	**0.851**	**0.944**
KMeans	0.510	0.461	0.480	0.435
GMM (full)	0.518	0.409	0.439	0.442
GMM (diag)	0.459	0.446	0.466	0.578
Spectral [39]	0.487	0.434	0.477	0.572
Watershed [14]	0.687	0.339	0.585	0.862
LCCP [29]	0.486	0.437	0.440	0.501
SLIC [25]	0.420	0.357	0.370	0.429
C2NO [26]	0.261	0.232	0.280	0.444
Mask-RCNN	0.659	0.712	0.591	0.262

mIoU comparisons against other methods.



Results using Mask-RCNN (see false positives and over-segmentations?)

Method	KMeans	Spectral	LCCP	MRCNN	Ours
time (s)	0.082	0.324	0.059	1.59 (0.135)	0.170

Time taken in seconds using a CPU (GPU)