DeepRob

Seminar 2
3D Perception: Point Cloud Processing
University of Michigan and University of Minnesota

This Week: 3D Perception

- Seminar 1: RGB-D Architectures

  1. PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes, Xiang et al., 2018
  2. A Unified Framework for Multi-View Multi-Class Object Pose Estimation, Li et al., 2018
  3. PVN3D: A Deep Point-Wise 3D Keypoints Voting Network for 6DoF Pose Estimation, He et al., 2020
  4. Learning RGB-D Feature Embeddings for Unseen Object Instance Segmentation, Li et al., 2021

- Seminar 2: Point Cloud Processing

  1. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, Qi et al., 2017
  2. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space, Qi et al., 2017
  3. PointFusion: Deep Sensor Fusion for 3D Bounding Box Estimation, Xu et al., 2018
  4. DenseFusion: 6D Object Pose Estimation by Iterative Dense Fusion, Wang et al., 2019

Today: Point Cloud Processing

- Seminar 1: RGB-D Architectures
  1. PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes, Xiang et al., 2018
  2. A Unified Framework for Multi-View Multi-Class Object Pose Estimation, Li et al., 2018
  3. PVN3D: A Deep Point-Wise 3D Keypoints Voting Network for 6DoF Pose Estimation, He et al., 2020
  4. Learning RGB-D Feature Embeddings for Unseen Object Instance Segmentation, Li et al., 2021

- Seminar 2: Point Cloud Processing
  1. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, Qi et al., 2017
  2. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space, Qi et al., 2017
  3. PointFusion: Deep Sensor Fusion for 3D Bounding Box Estimation, Xu et al., 2018
  4. DenseFusion: 6D Object Pose Estimation by Iterative Dense Fusion, Wang et al., 2019

DenseFusion

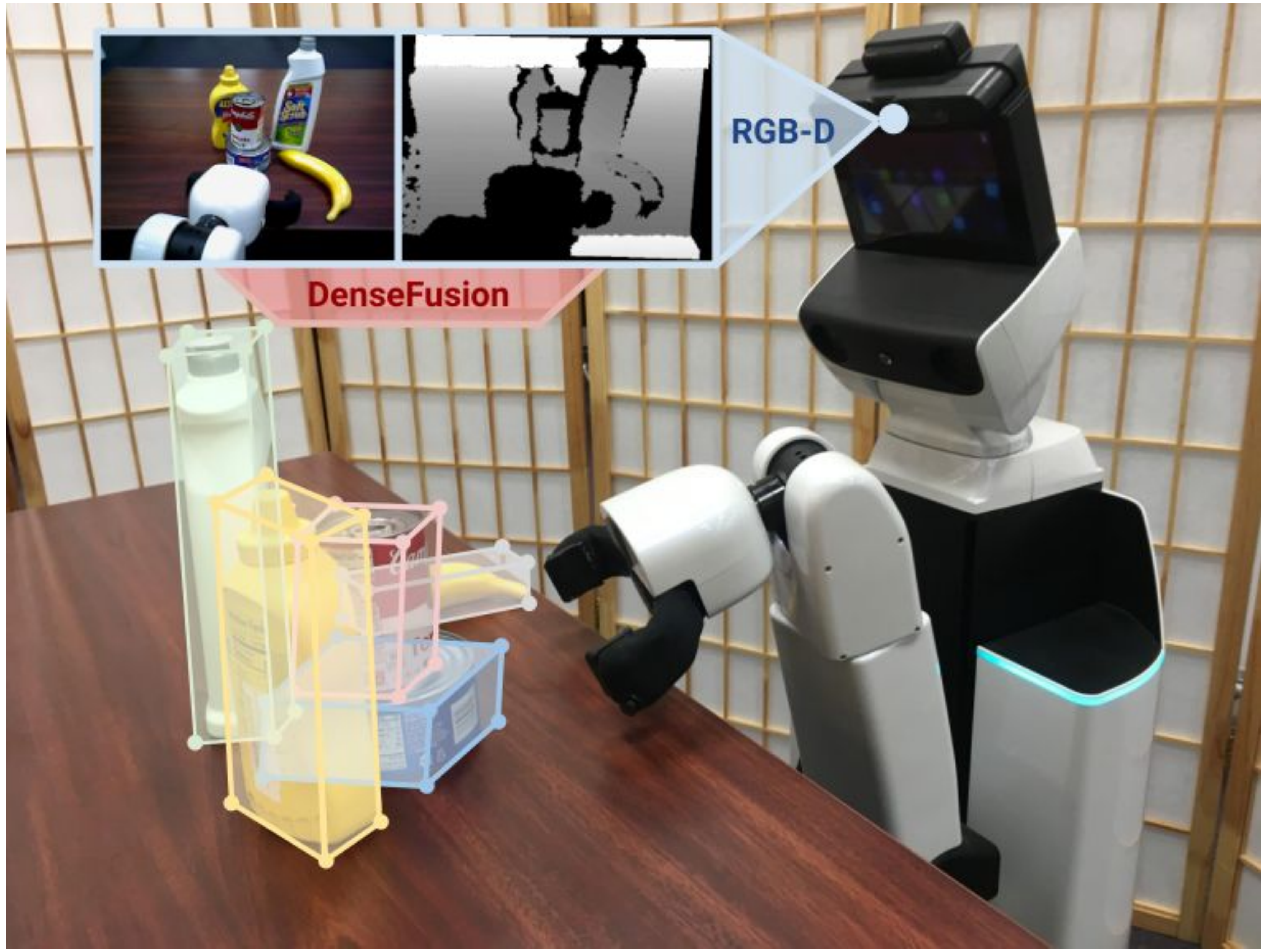
6D Object Pose Estimation by Iterative Dense Fusion

By: Chen Wang, Danfei Xu, Luke Zhu, Roberto Martín-Martín
Cewu Lu, Li Fei-Fei, Silvio Savarese

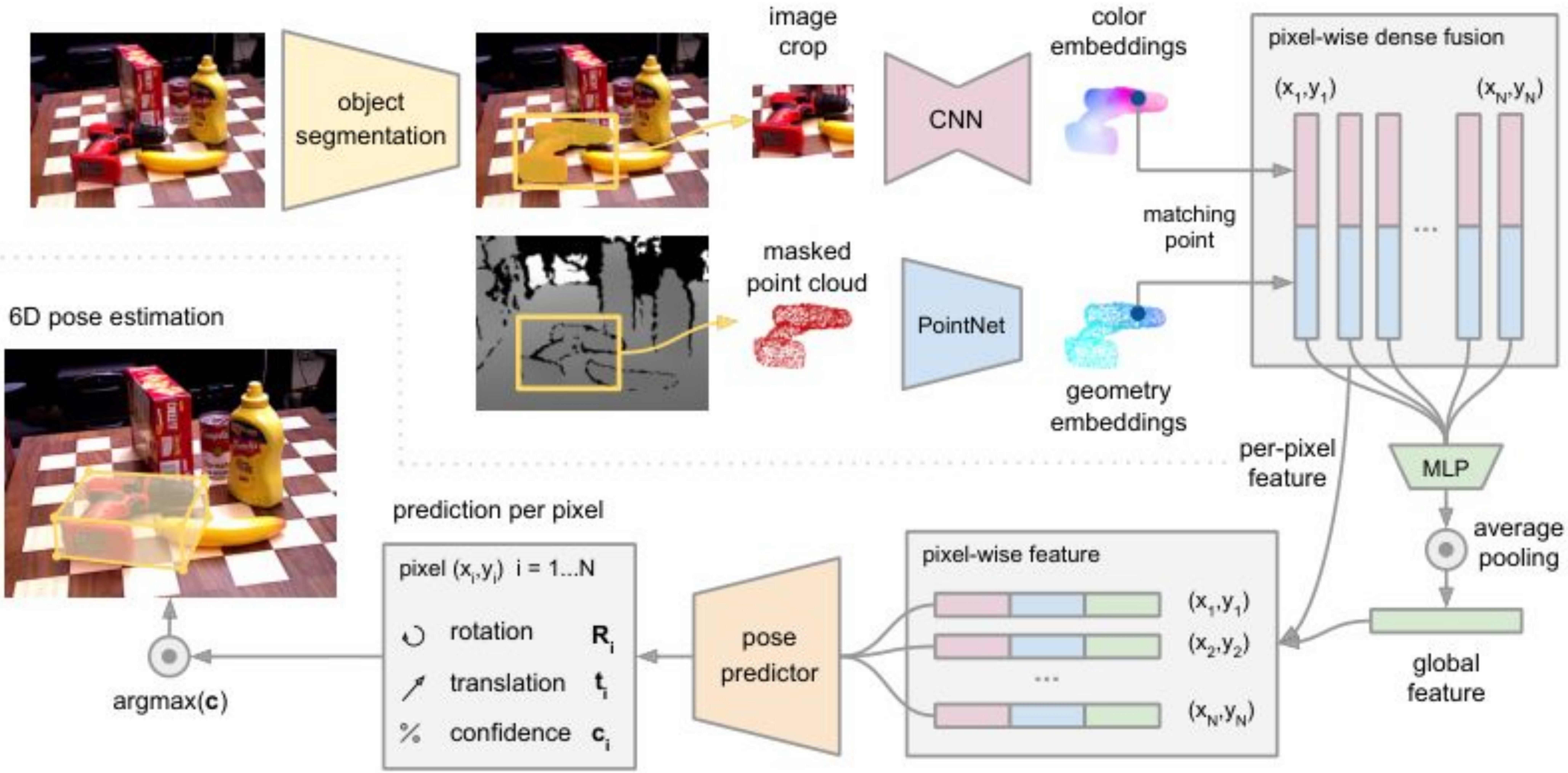
Presented by: Yogi Sahu

Objective

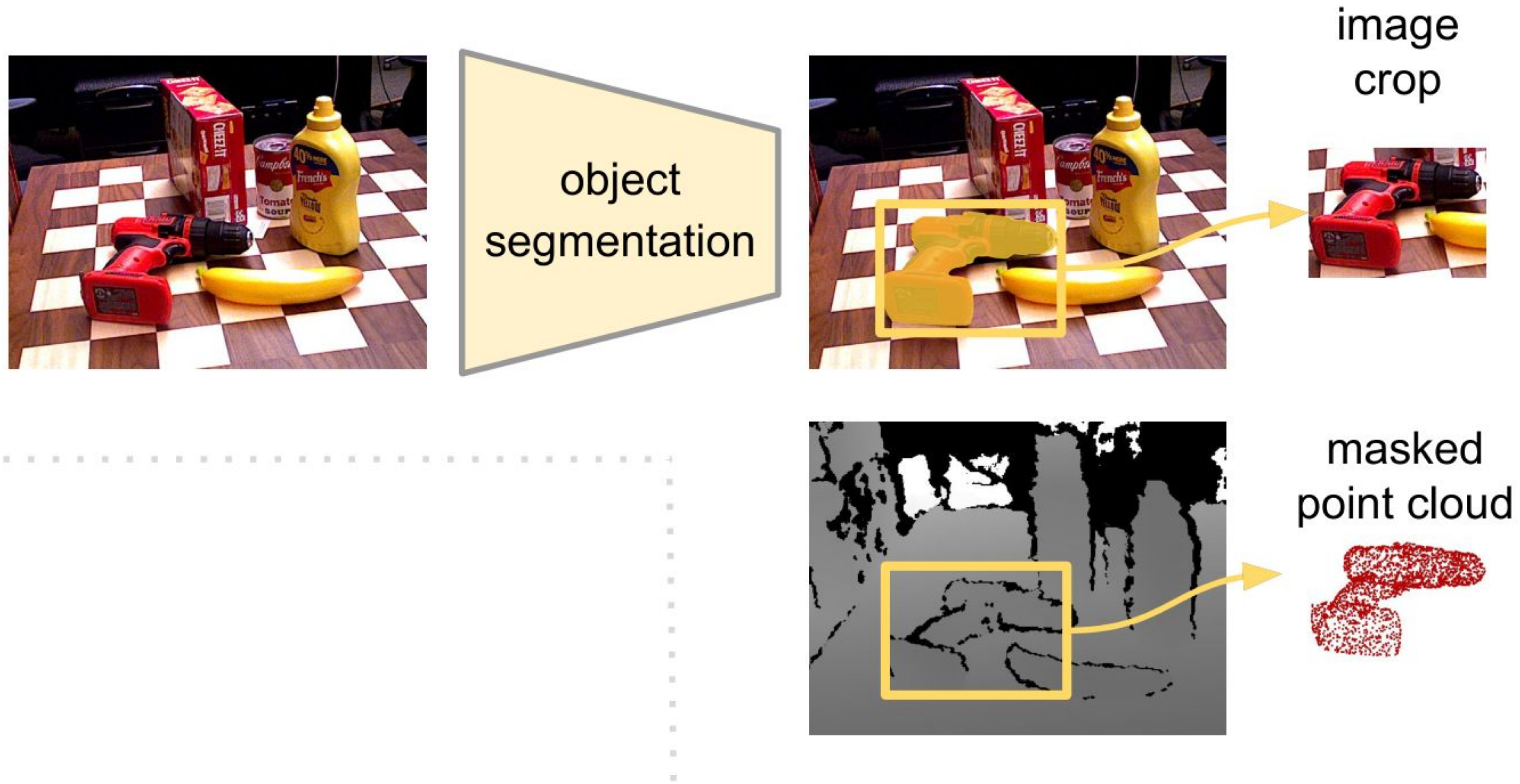
Perform fast and accurate 6D pose estimation for real-time applications such as robot grasping and manipulation



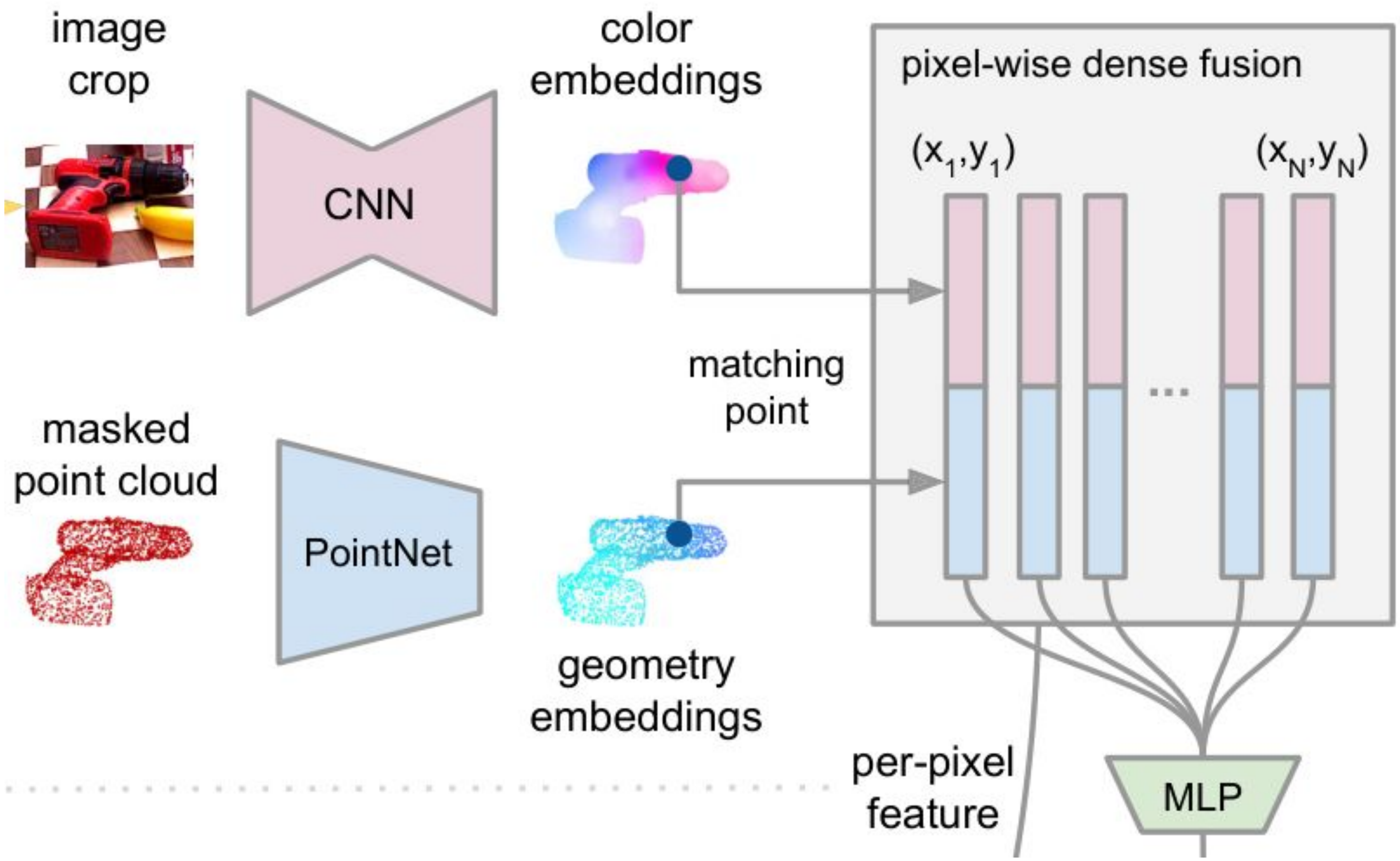
Model Overview



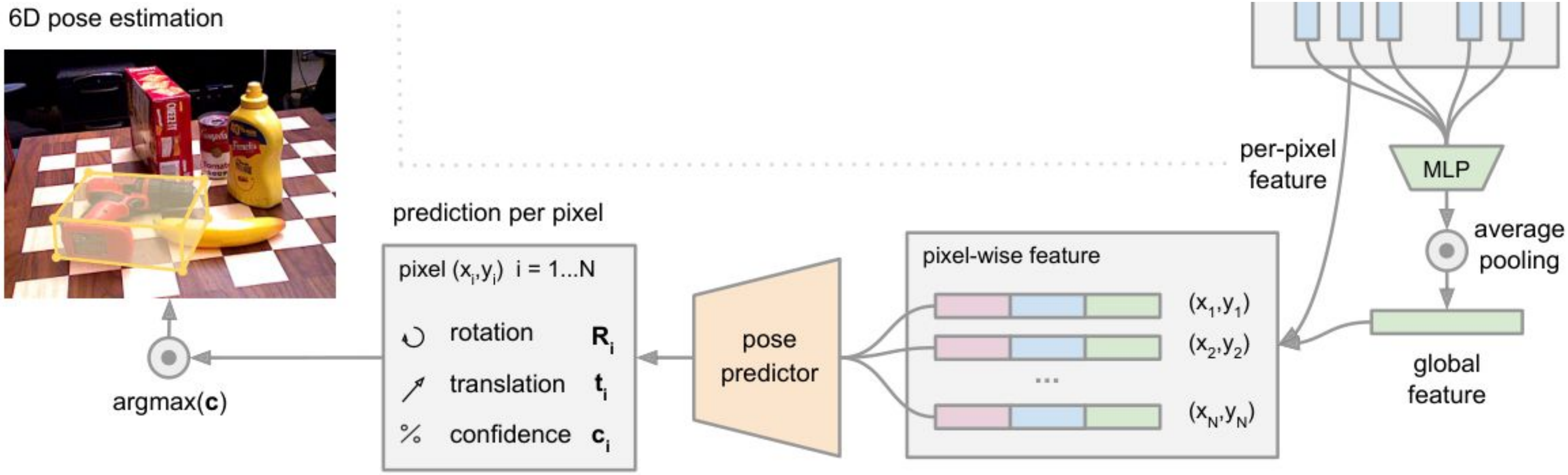
Object Segmentation



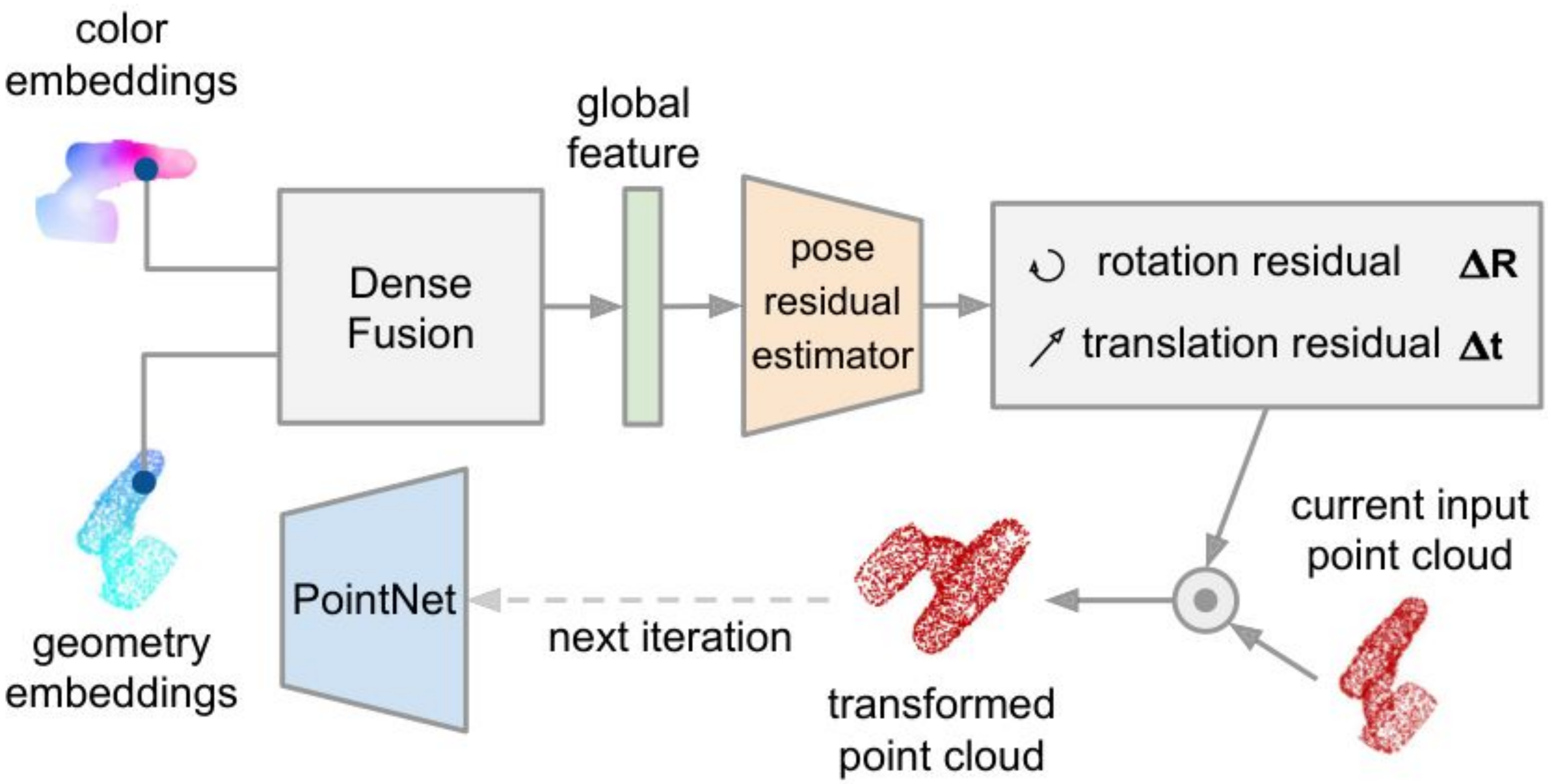
Fuse RGB and Point Cloud Data



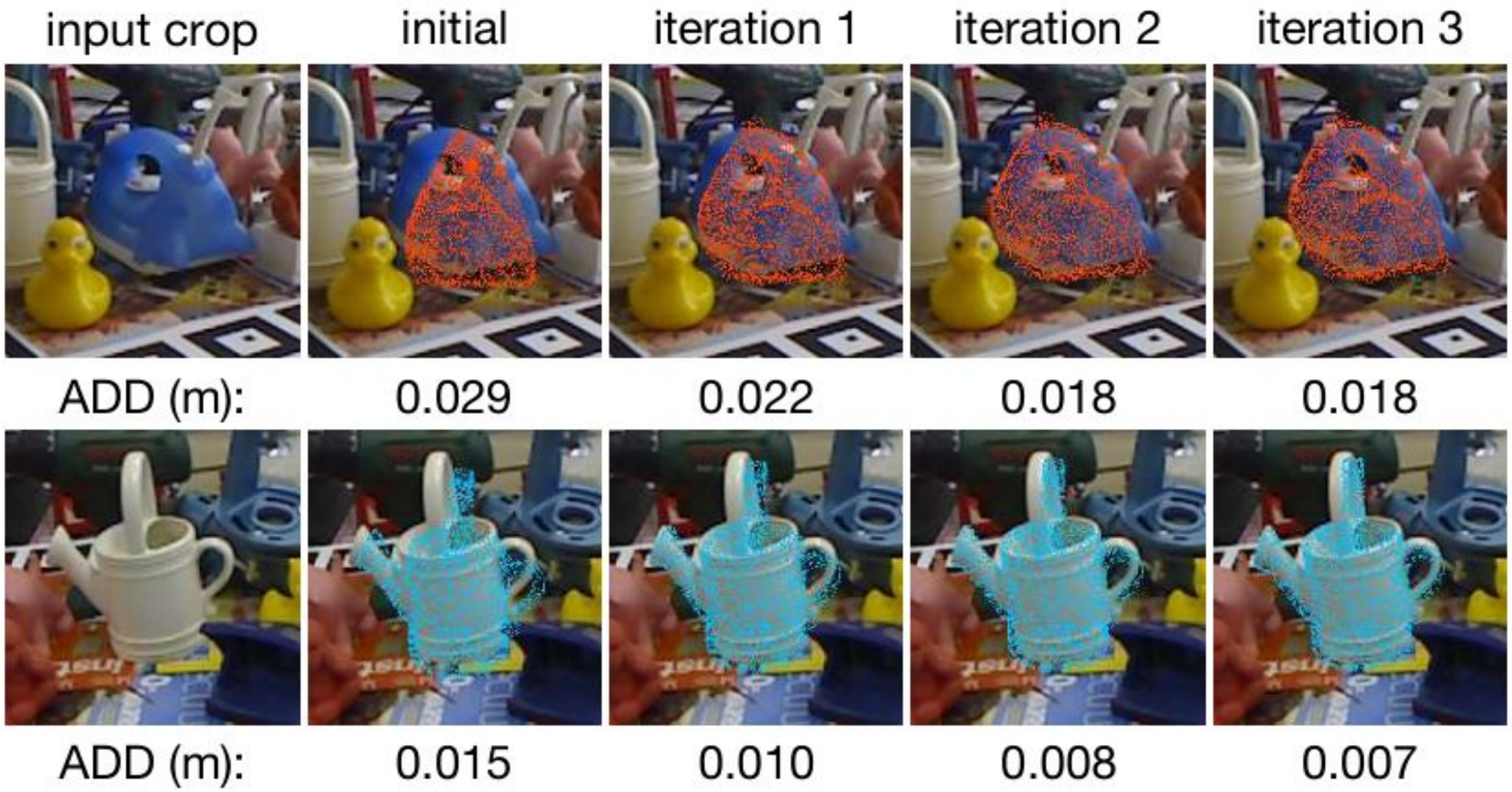
Pose Prediction



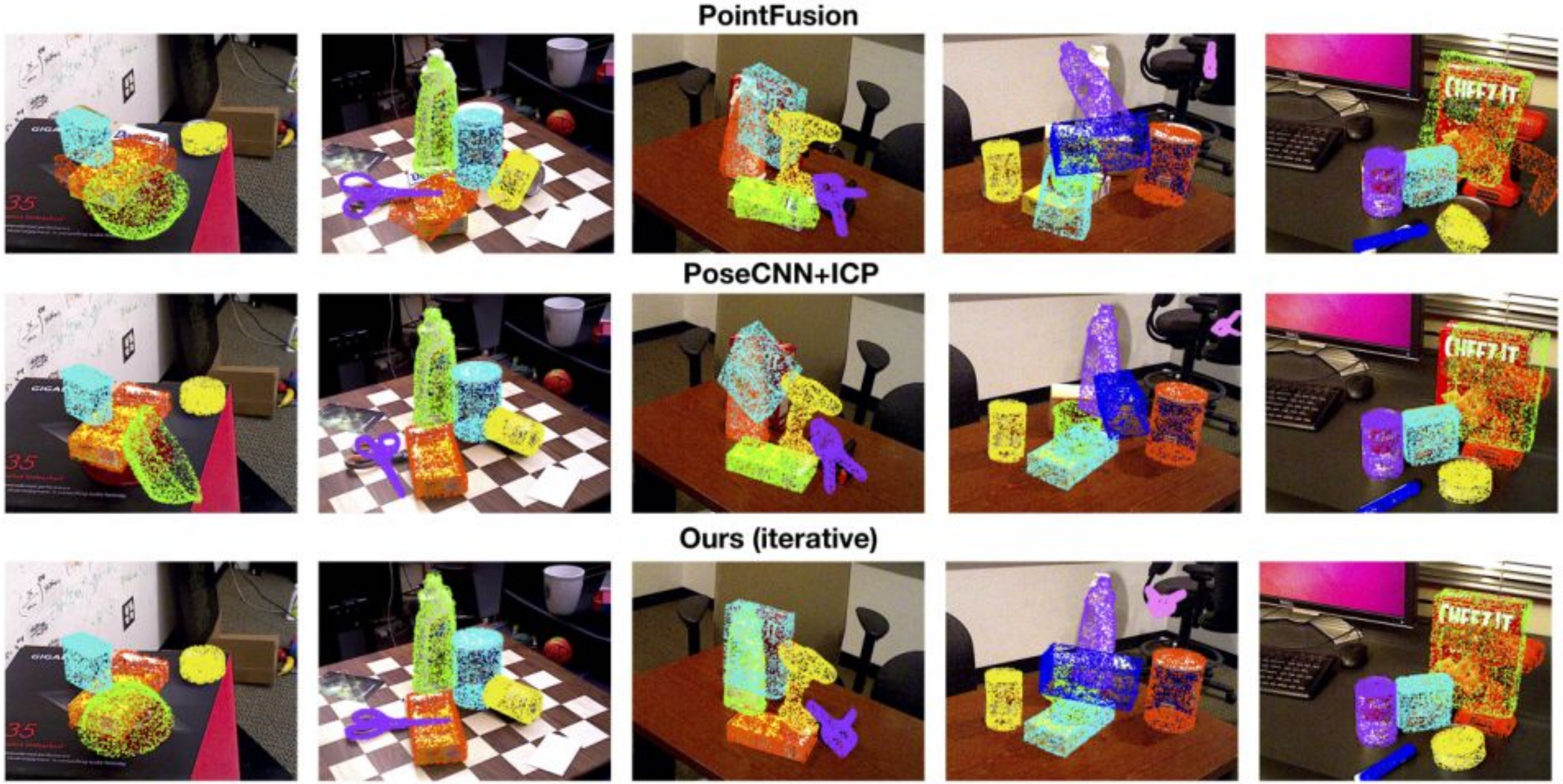
Iterative Pose Refinement



Iterative Pose Refinement Example



Qualitative Results on the YCB-Video Dataset



Quantitative Results on the YCB-Video Dataset

	PointFusion [42]		PoseCNN+ICP [41]		Ours (single)		Ours (per-pixel)		Ours (iterative)	
	AUC	<2cm	AUC	<2cm	AUC	<2cm	AUC	<2cm	AUC	<2cm
002_master_chef_can	90.9	99.8	95.8	100.0	93.9	100.0	95.2	100.0	**96.4**	100.0
003_cracker_box	80.5	62.6	92.7	91.6	90.8	98.4	92.5	99.3	**95.5**	**99.5**
004_sugar_box	90.4	95.4	**98.2**	100.0	94.4	99.2	95.1	100.0	97.5	100.0
005_tomato_soup_can	91.9	96.9	94.5	96.9	92.9	96.7	93.7	96.9	**94.6**	96.9
006_mustard_bottle	88.5	84.0	**98.6**	100.0	91.2	97.8	95.9	100.0	97.2	100.0
007_tuna_fish_can	93.8	99.8	**97.1**	100.0	94.9	100.0	94.9	100.0	96.6	100.0
008_pudding_box	87.5	96.7	**97.9**	100.0	88.3	97.2	94.7	100.0	96.5	100.0
009_gelatin_box	95.0	100.0	**98.8**	100.0	95.4	100.0	95.8	100.0	98.1	100.0
010_potted_meat_can	86.4	88.5	**92.7**	**93.6**	87.3	91.4	90.1	93.1	91.3	93.1
011_banana	84.7	70.5	**97.1**	99.7	84.6	62.0	91.5	93.9	96.6	**100.0**
019_pitcher_base	85.5	79.8	**97.8**	100.0	86.9	80.9	94.6	100.0	97.1	100.0
021_bleach_cleanser	81.0	65.0	**96.9**	99.4	91.6	98.2	94.3	99.8	95.8	**100.0**
024_bowl	75.7	24.1	81.0	54.9	83.4	55.4	86.6	69.5	**88.2**	**98.8**
025_mug	94.2	99.8	95.0	99.8	90.3	94.7	95.5	**100.0**	**97.1**	**100.0**
035_power_drill	71.5	22.8	**98.2**	**99.6**	83.1	64.2	92.4	97.1	96.0	98.7
036_wood_block	68.1	18.2	87.6	80.2	81.7	76.0	85.5	93.4	**89.7**	**94.6**
037_scissors	76.7	35.9	91.7	95.6	83.6	75.1	96.4	**100.0**	**95.2**	**100.0**
040_large_marker	87.9	80.4	97.2	99.7	91.2	88.6	94.7	99.2	**97.5**	**100.0**
051_large_clamp	65.9	50.0	**75.2**	74.9	70.5	77.1	71.6	78.5	72.9	**79.2**
052_extra_large_clamp	60.4	20.1	64.4	48.8	66.4	50.2	69.0	69.5	**69.8**	**76.3**
061_foam_brick	91.8	100.0	**97.2**	100.0	92.1	100.0	92.4	100.0	92.5	100.0
MEAN	83.9	74.1	93.0	93.2	88.2	87.9	91.2	95.3	**93.1**	**96.8**

Quantitative Results on the LineMOD Dataset

		ape	ben.	cam	can	cat	drill.	duck	**egg.**	**glue**	hole.	iron	lamp	phone	MEAN
RGB	BB8 w ref. [25]	40	92	56	64	63	74	44	58	41	67	84	77	54	63
	DeepIM [17, 41]	77	98	94	97	82	95	78	97	99	53	98	98	88	89
RGB-D	Imp. [31]+ICP	21	64	63	76	72	42	32	99	96	50	63	92	71	65
	SSD6D [14]+ICP	65	80	78	86	70	73	66	100	100	49	78	73	79	79
	PointFusion [42]	70	81	61	61	79	47	63	100	99	72	83	62	79	74
	Ours (per-pixel)	80	84	77	87	89	78	76	100	99	79	92	92	88	86
	Ours (iterative)	**92**	**93**	**94**	**93**	**97**	**87**	**92**	100	100	**92**	**97**	**95**	**93**	**94**

Runtime

Table 2. Runtime breakdown (second per frame on YCB-Video Dataset). Our method is approximately 200x faster than PoseCNN+ICP. Seg means Segmentation, and PE means Pose Estimation.

PoseCNN+ICP [41]				Ours			
Seg	PE	ICP	ALL	Seg	PE	Refine	ALL
0.03	0.17	10.4	10.6	0.03	0.02	0.01	0.06

Conclusion

- Dense fusion has a clear advantage over the global fusion-by-concatenation method used in PointFusion because dense fusion baselines outperform PointFusion by a large margin
- Iterative refinement significantly improves the performance for texture-less symmetric objects (ex: bowl, banana, and extra_large_lamp in the YCB-Video dataset)
- The dense fusion method provides robustness towards occlusions
- This method is two orders of magnitude faster than PoseCNN+ICP

Thank you

Next Time: Rigid Body Objects

- Seminar 3: Object Pose, Geometry, SDF, Implicit Surfaces
  1. SUM: Sequential scene understanding and manipulation, Sui et al., 2017
  2. DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation, Park et al., 2019
  3. Implicit surface representations as layers in neural networks, Michalkiewicz et al., 2019
  4. iSDF: Real-Time Neural Signed Distance Fields for Robot Perception, Oriz et al., 2022

- Seminar 4: Dense Descriptors, Category-level Representations
  1. Dense Object Nets: Learning Dense Visual Object Descriptors By and For Robotic Manipulation, Florence et al., 2018
  2. Normalized Object Coordinate Space for Category-Level 6D Object Pose and Size Estimation, Wang et al., 2019
  3. kPAM: KeyPoint Affordances for Category-Level Robotic Manipulation, Manuelli et al., 2019
  4. Single-Stage Keypoint-Based Category-Level Object Pose Estimation from an RGB Image, Lin et al., 2022

DeepRob

Seminar 2
3D Perception: Point Cloud Processing
University of Michigan and University of Minnesota