



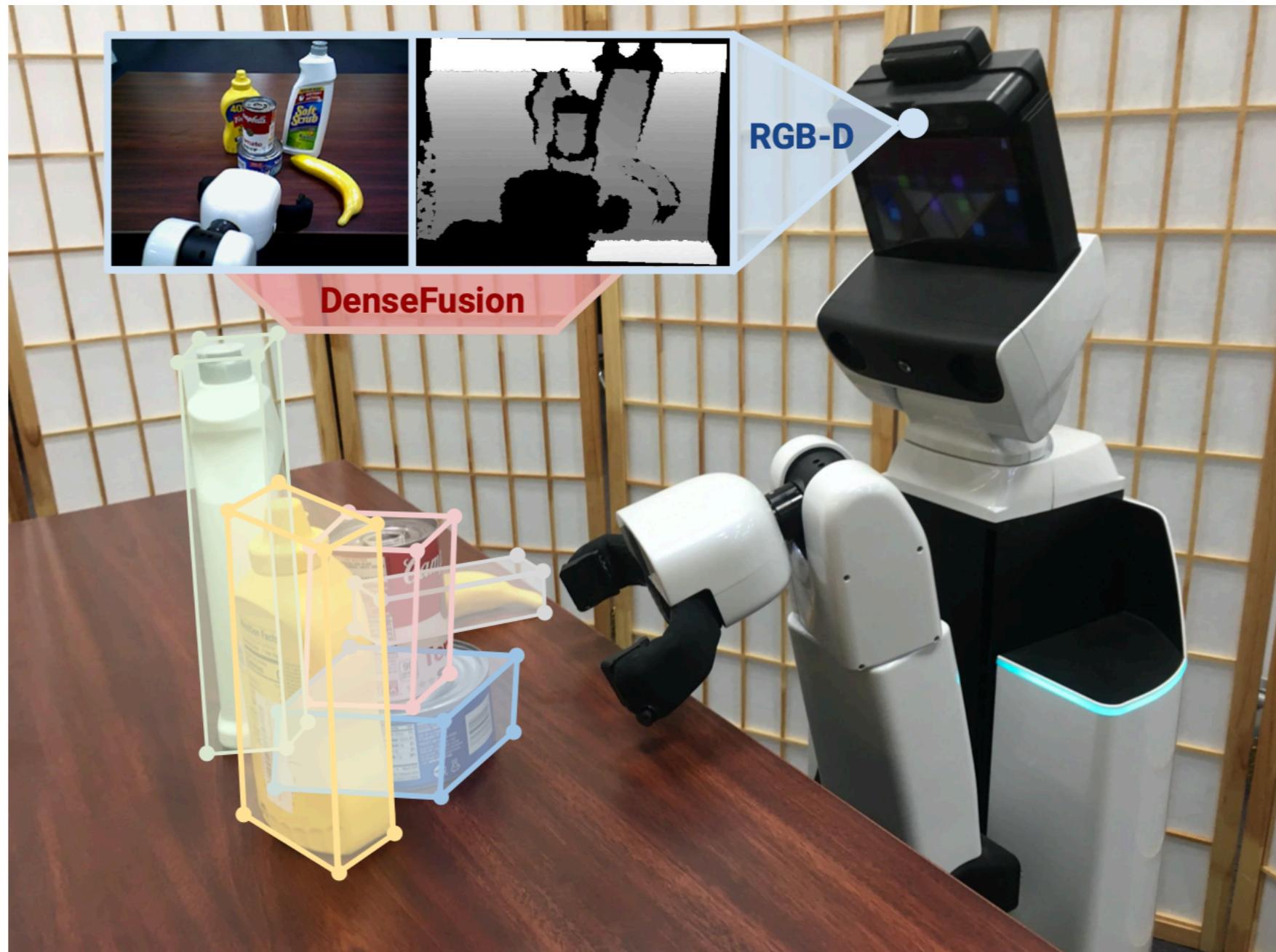
DENSEFUSION: 6D OBJECT POSE ESTIMATION BY ITERATIVE DENSE FUSION

Wang, C., Xu, D., Zhu, Y., Martín-Martín, R., Lu, C., Fei-Fei, L., & Savarese, S. (2019). Densefusion: 6d object pose estimation by iterative dense fusion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3343-3352).

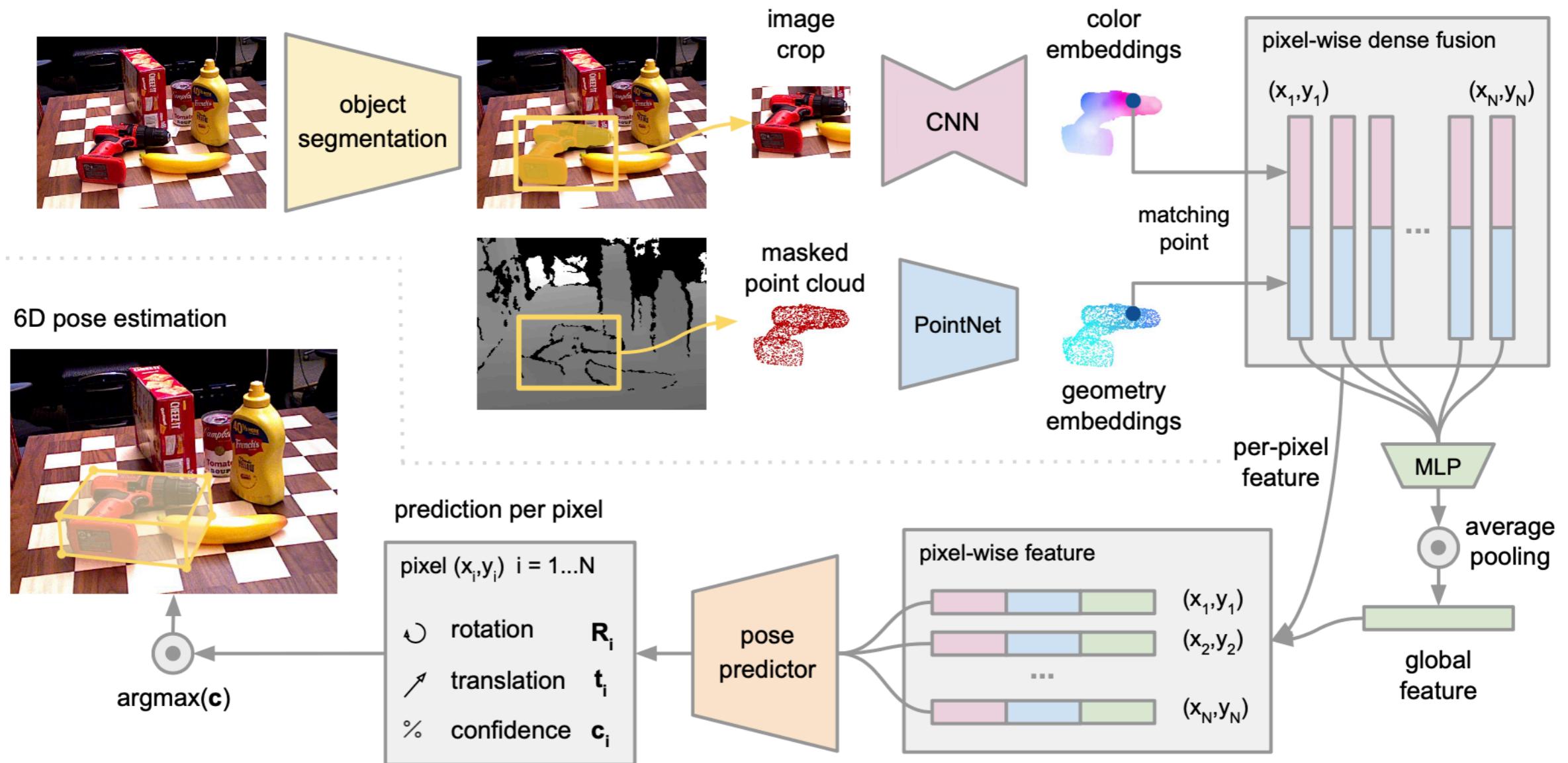
Name (Pinyin): Zhang Yizheng
Email: zhangyzh1@shanghaitech.edu.cn

OUTLINE OF DENSEFUSION

- A heterogeneous architecture that processes the two data sources individually and uses a novel dense fusion network to extract pixel-wise dense feature embedding, from which the pose is estimated.
- Integrate an end-to-end iterative pose refinement procedure that further improves the pose estimation while achieving near real-time inference.

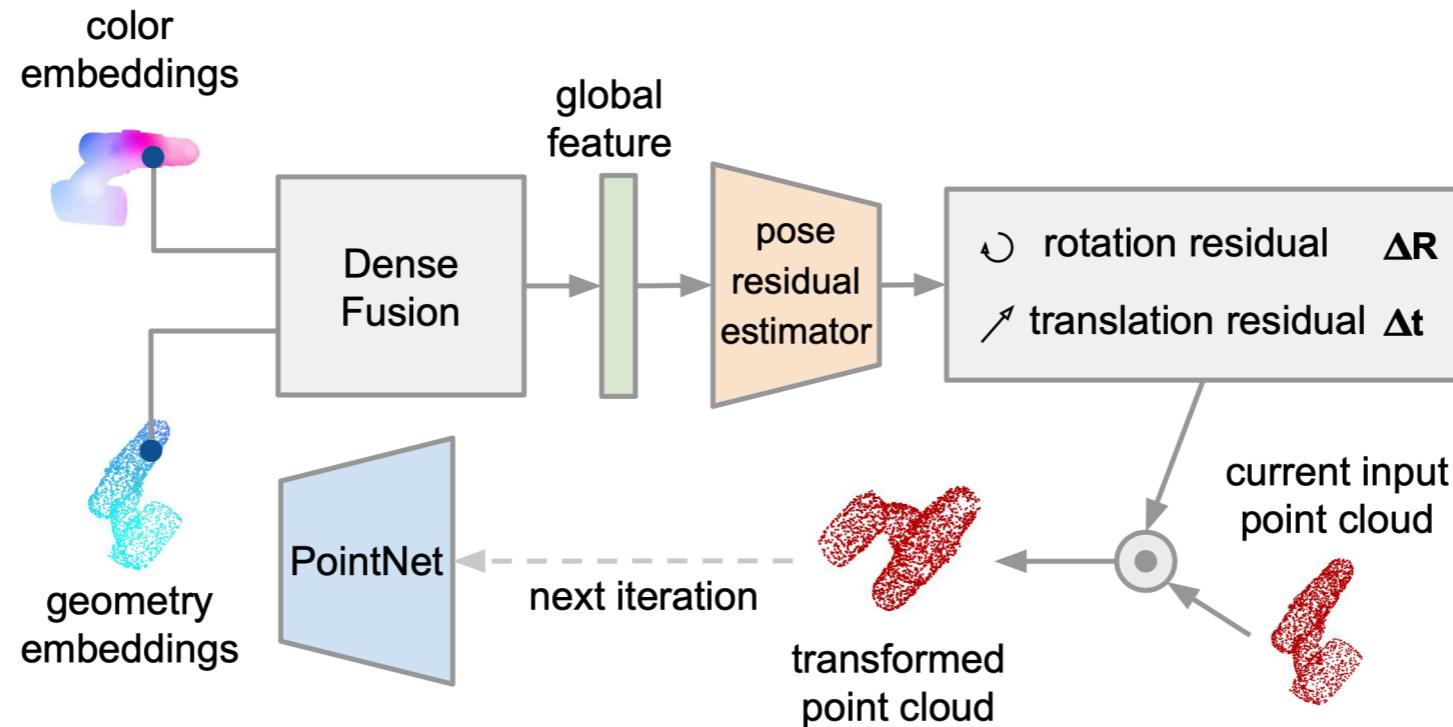


ARCHITECTURE OVERVIEW



- Takes color image as input and performs semantic segmentation for each known object category.
- Need the masked depth pixels (converted to 3D point cloud) as well as an image patch cropped by the bounding box of the mask.

ITERATIVE POSE REFINEMENT



- Key idea is to consider the previously predicted pose as an estimate of canonical frame of the target object and transform the input point cloud into this estimated canonical frame.
- Reuse the image feature embedding from the main network and perform dense fusion with the geometric features computed for the new transformed point cloud.

EXPERIMENT RESULT OF DENSEFUSION

- Time efficiency: 16fps, 5 objects in each frame
- Accuracy: the average closest point distance(ADD-S) is smaller than 2cm
- The robot succeeds on 73% of the grasps using our proposed approach to estimate the pose of the objects

		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

CORRELATIONS OF THE PROJECT



- Pose estimation of the tube holder.
- Pose estimation of the centrifuge tube and the screw lid.