THU-PM-066

Markerless Camera-to-Robot Pose Estimation via Self-supervised Sim-to-Real Transfer

Jingpei Lu, Florian Richter, and Michael C. Yip

University of California, San Diego

UC San Diego JACOBS SCHOOL OF ENGINEERING Electrical and Computer Engineering

ARClab UCSD Advanced Robotics and Controls

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

THU-PM-066

Markerless Camera-to-Robot Pose Estimation via Self-supervised Sim-to-Real Transfer

Jingpei Lu, Florian Richter, and Michael C. Yip

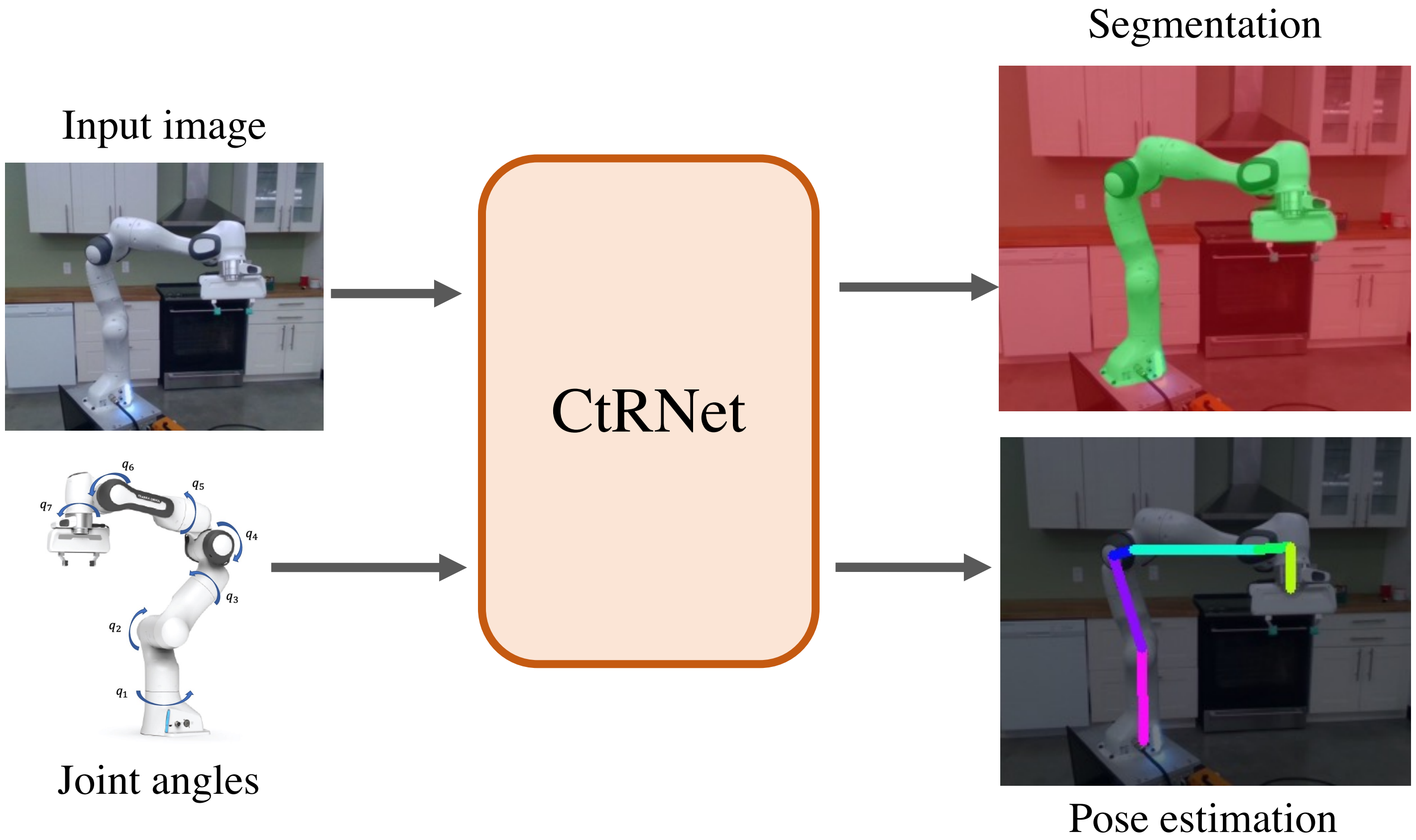
University of California, San Diego

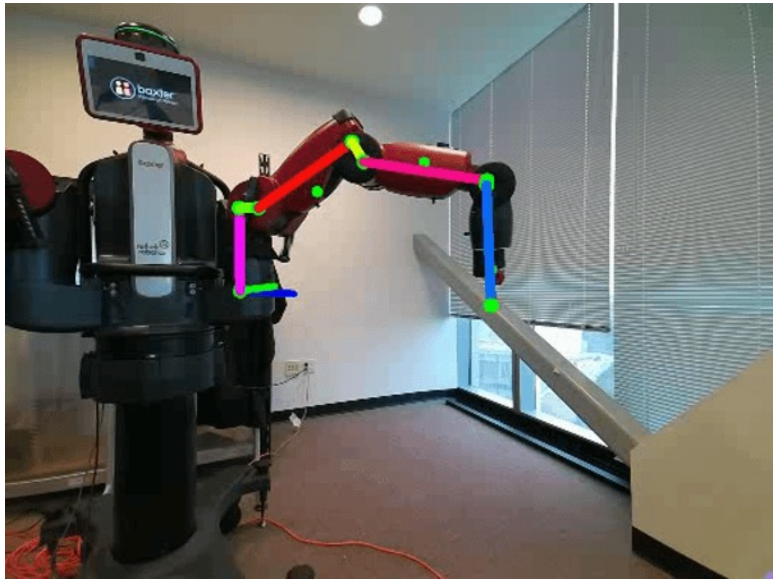
UC San Diego
JACOBS SCHOOL OF ENGINEERING
Electrical and Computer Engineering

ARClab
UCSD Advanced Robotics and Controls

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Camera-to-Robot Pose Estimation Network

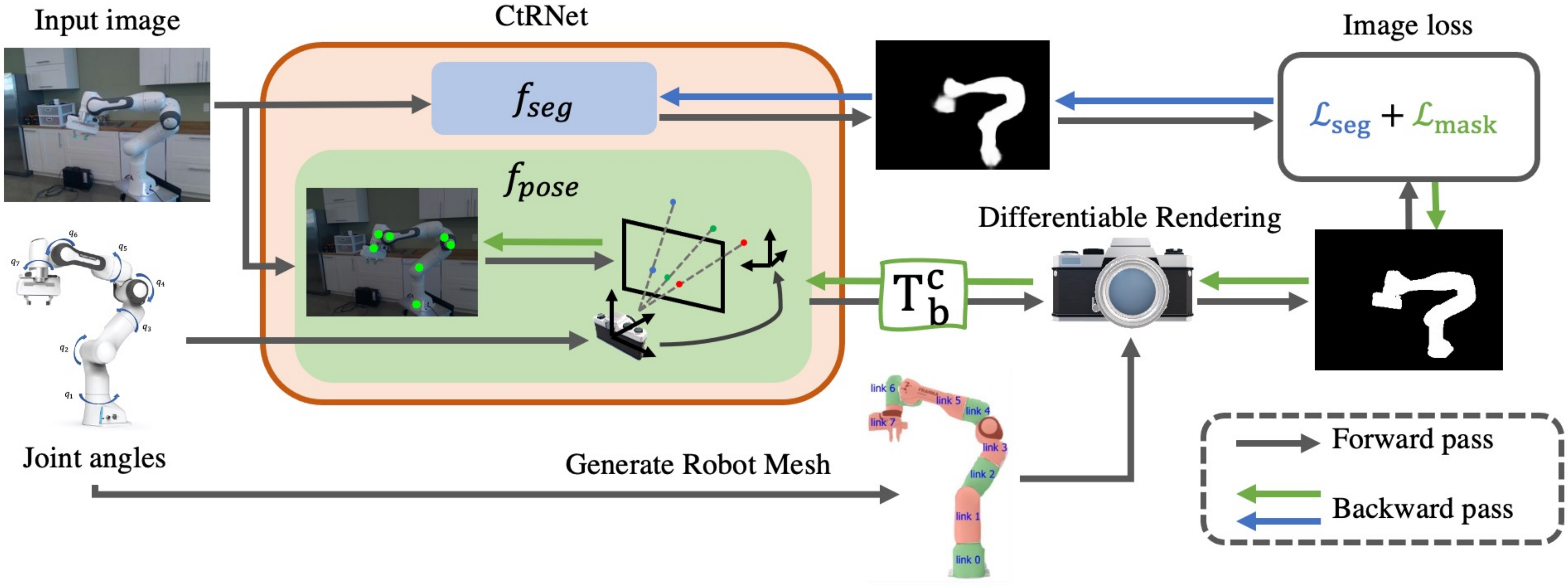


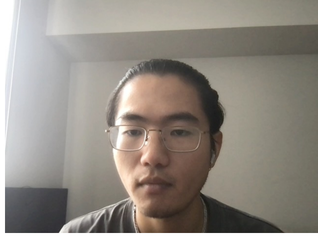


CtRNet inference in real-time

CtRNet segments the robot and estimates the robot pose

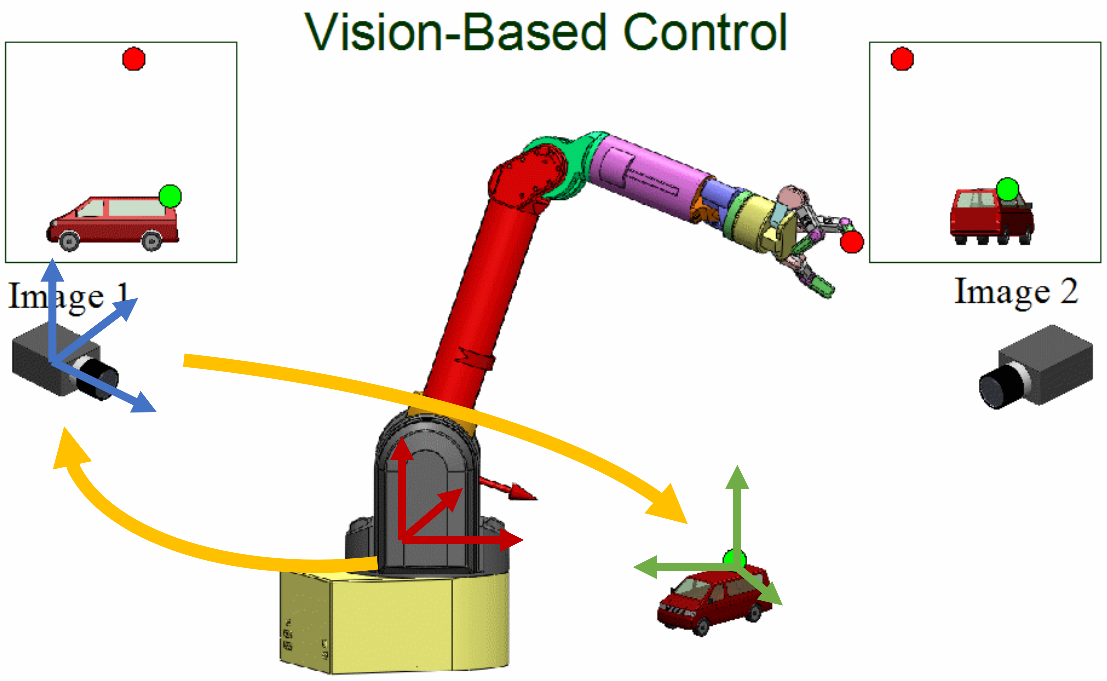
Self-training for sim-to-real transfer





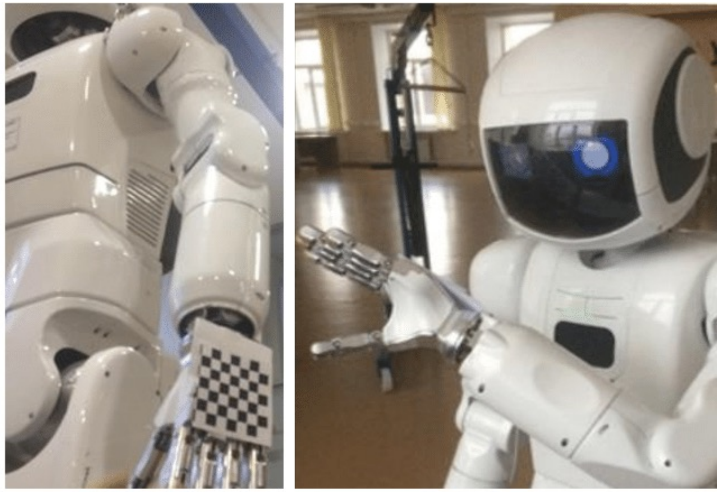
Background

- Camera-to-robot calibration is a fundamental procedure for vision-based robot control.

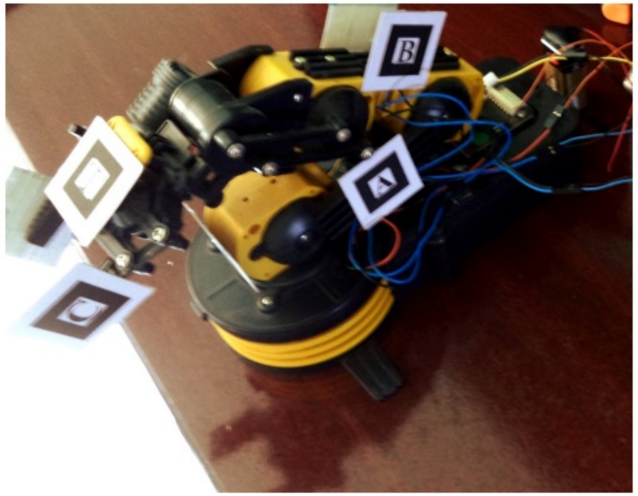


http://webdocs.cs.ualberta.ca/~vis/robotics/uncalib/uncalibvs.htm

Marker-based Calibration



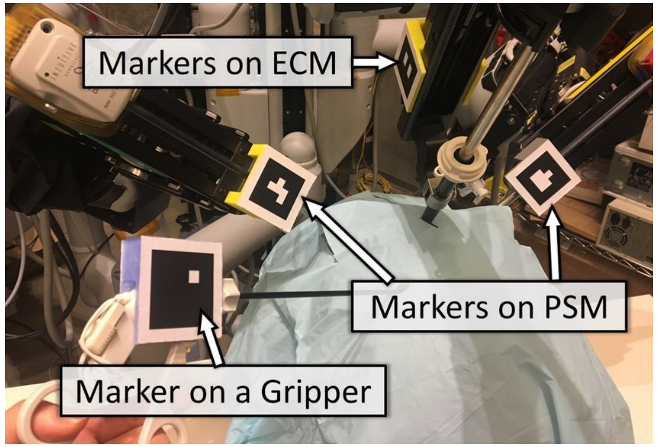
Zakiev, Aufar, et al. "Pilot virtual experiments on aruco and artag systems comparison for fiducial marker rotation resistance." *Proceedings of 14th International Conference on Electromechanics and Robotics "Zavalishin's Readings"*. Springer, Singapore, 2020.



Vieira, Igor Pereira, Armando Alves Neto, and Leonardo Amaral Mozelli. "Fiducial markers applied for pose tracking of a robotic manipulator: application in visual servoing control." *2016 XIII Latin American Robotics Symposium and IV Brazilian Robotics Symposium (LARS/SBR)*. IEEE, 2016.

Ranganathan, Pradeep, et al. "Coordinating a team of robots for urban reconnaisance." *Proceedings of the land warfare conference (LWC)*. 2010.

Lightbody, Peter, Tomáš Krajník, and Marc Hanheide. "A versatile high-performance visual fiducial marker detection system with scalable identity encoding." *Proceedings of the symposium on applied computing*. 2017.

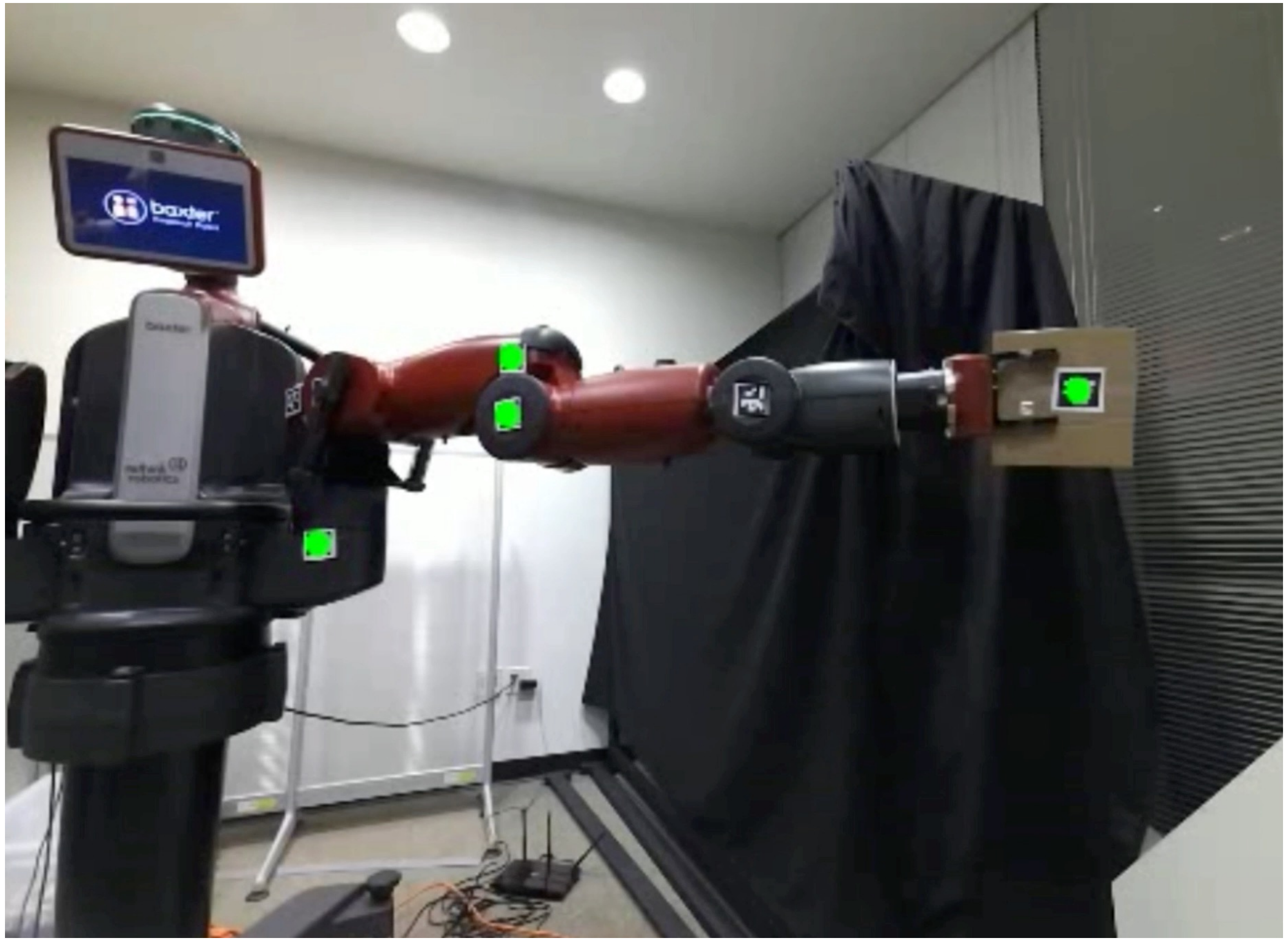


Qian, Long, Anton Deguet, and Peter Kazanzides. "ARssist: augmented reality on a head-mounted display for the first assistant in robotic surgery." *Healthcare technology letters* 5.5 (2018): 194-200.

The calibration is cumbersome in reality…

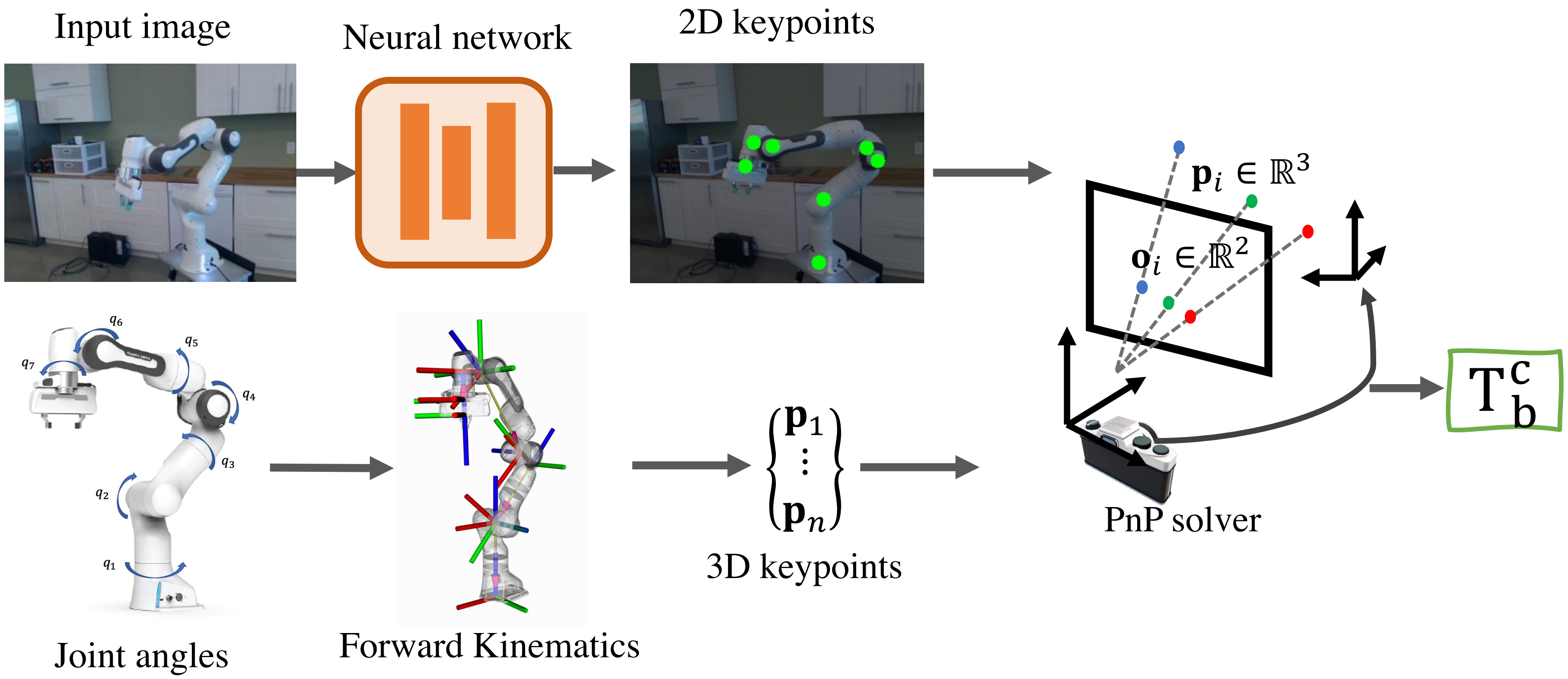
- Physically modify the robot
- Measurement error
- Collect multiple images
- Offline calibration procedures
- Removing the markers

☹

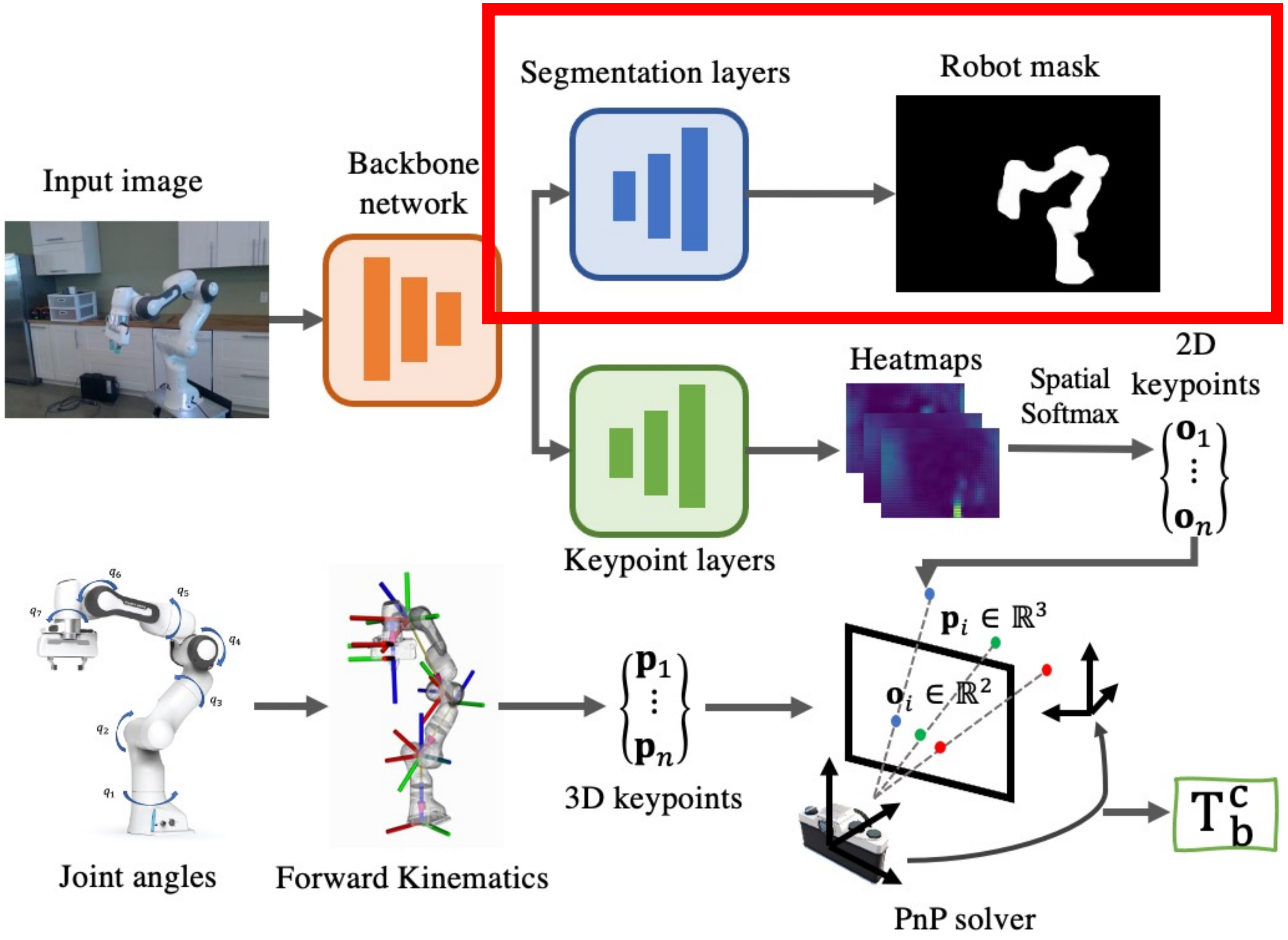


Marker-less camera-to-robot pose estimation

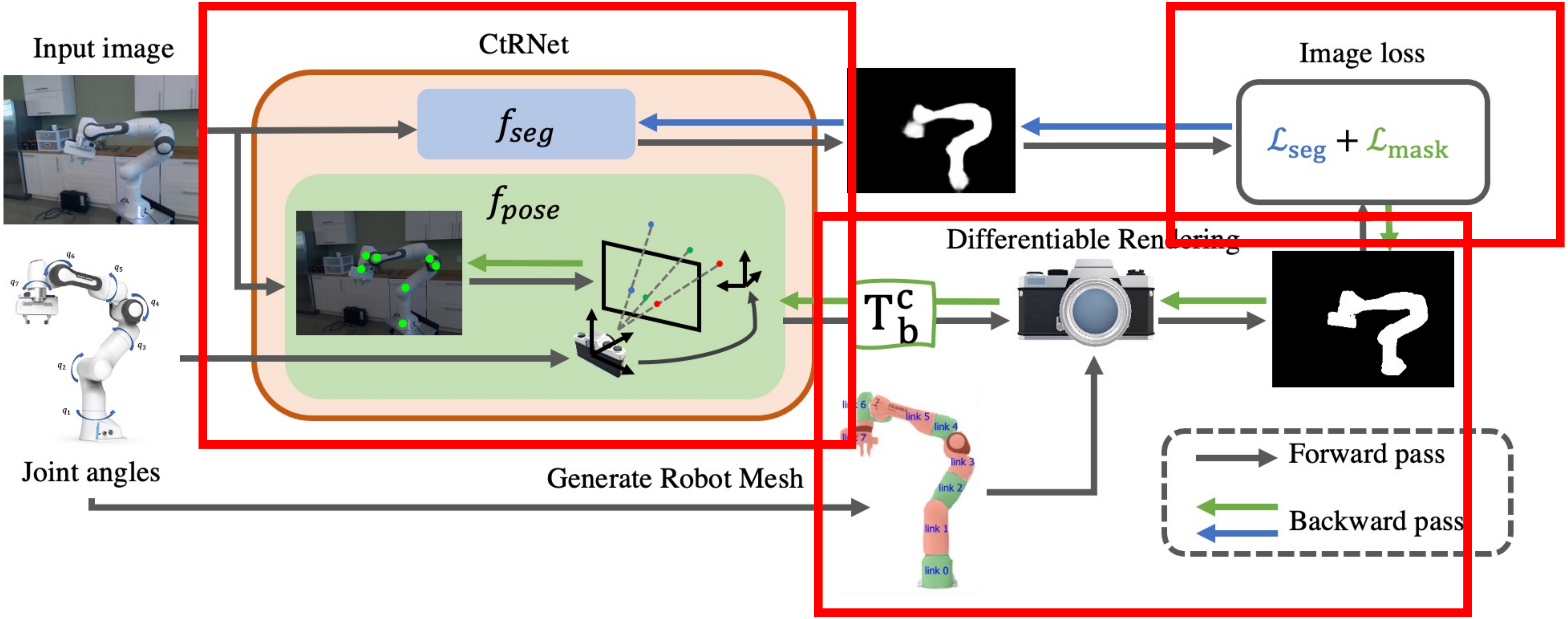
- CtRNet using a neural network to detect 2D keypoints for pose estimation



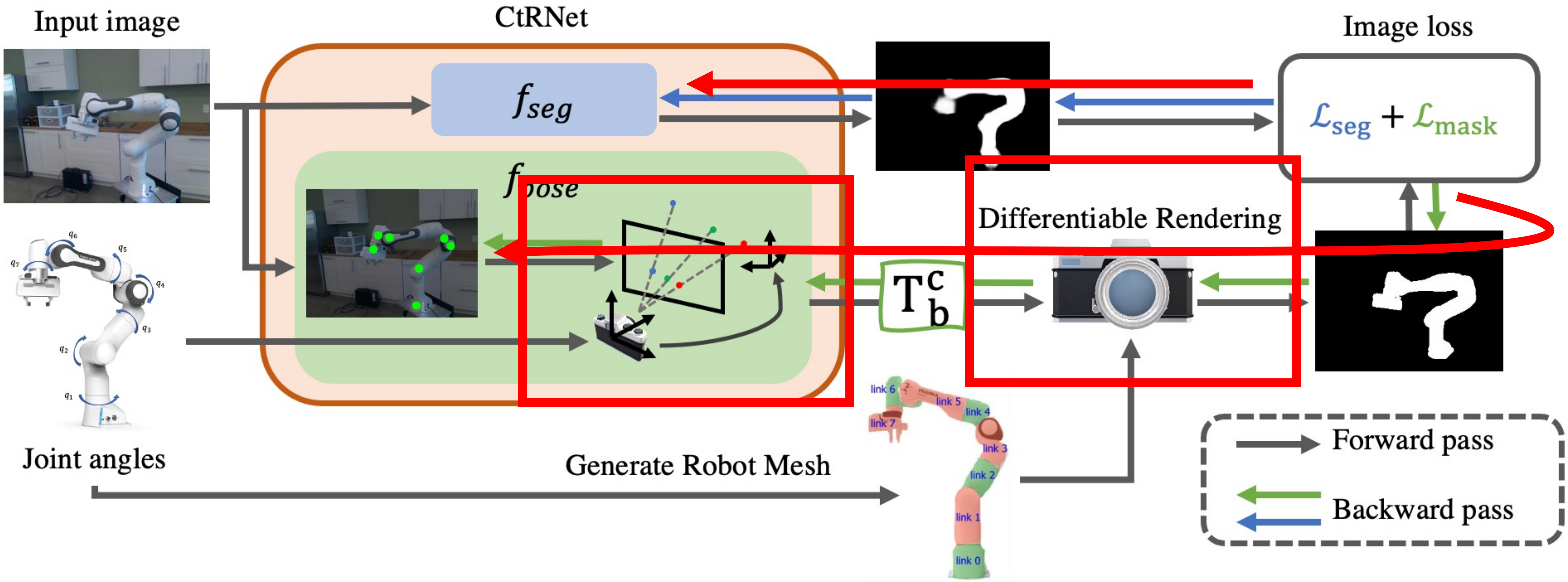
Camera-to-robot pose estimation network

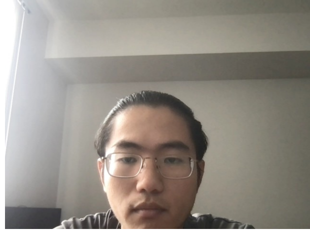


Self-supervised training

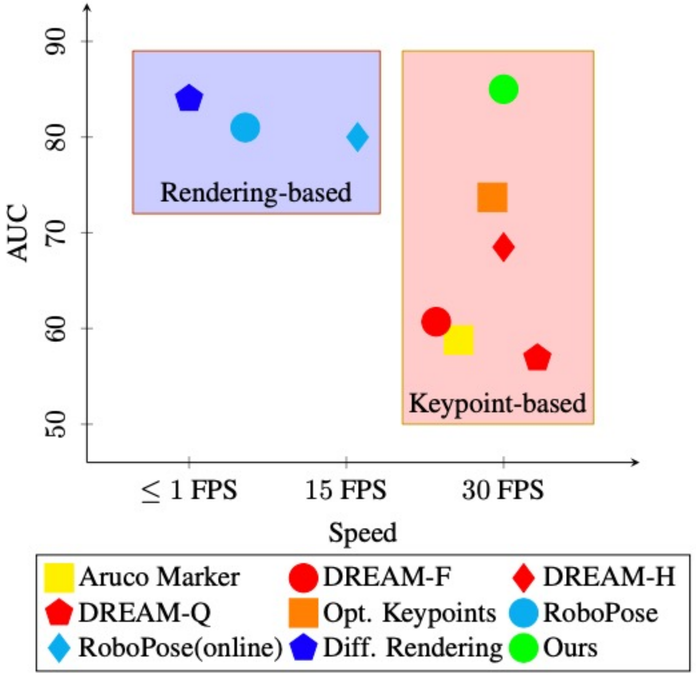


Self-supervised training





Compare with the SOTA

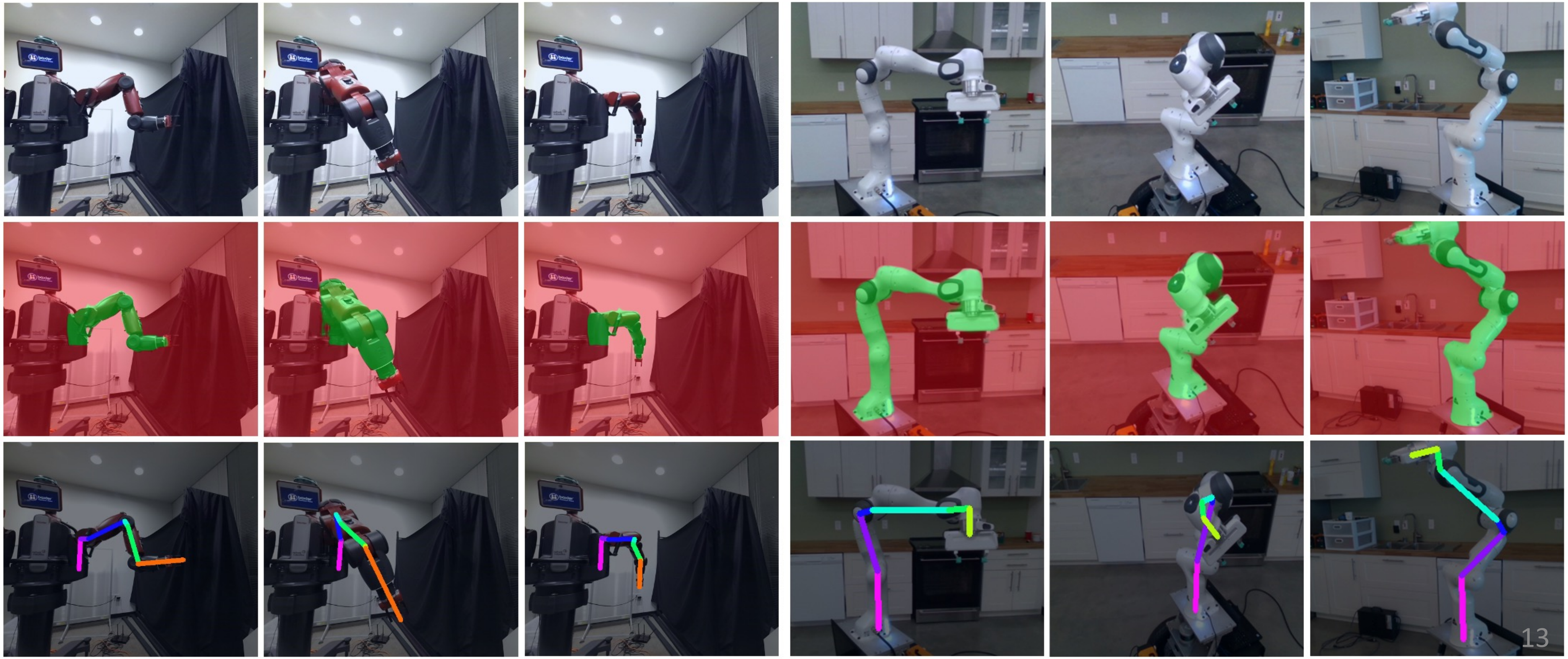


Method	Category	Backbone	Panda 3CAM-AK		Panda 3CAM-XK		Panda 3CAM-RS		Panda ORB		All	
			AUC ↑	Mean (m) ↓	AUC ↑	Mean (m) ↓	AUC ↑	Mean (m) ↓	AUC ↑	Mean (m) ↓	AUC ↑	Mean (m) ↓
DREAM-F [29]	Keypoint	VGG19	68.912	11.413	24.359	491.911	76.130	2.077	61.930	95.319	60.740	113.029
DREAM-Q [29]	Keypoint	VGG19	52.382	78.089	37.471	54.178	77.984	0.027	57.087	67.248	56.988	59.284
DREAM-H [29]	Keypoint	ResNet101	60.520	0.056	64.005	7.382	78.825	0.024	69.054	25.685	68.584	17.477
RoboPose [26]	Rendering	ResNet34	76.497	0.024	**85.926**	**0.014**	76.863	0.023	80.504	**0.019**	80.094	**0.020**
CtRNet	Keypoint	ResNet50	**89.928**	**0.013**	79.465	0.032	**90.789**	**0.010**	**85.289**	0.021	**85.962**	**0.020**

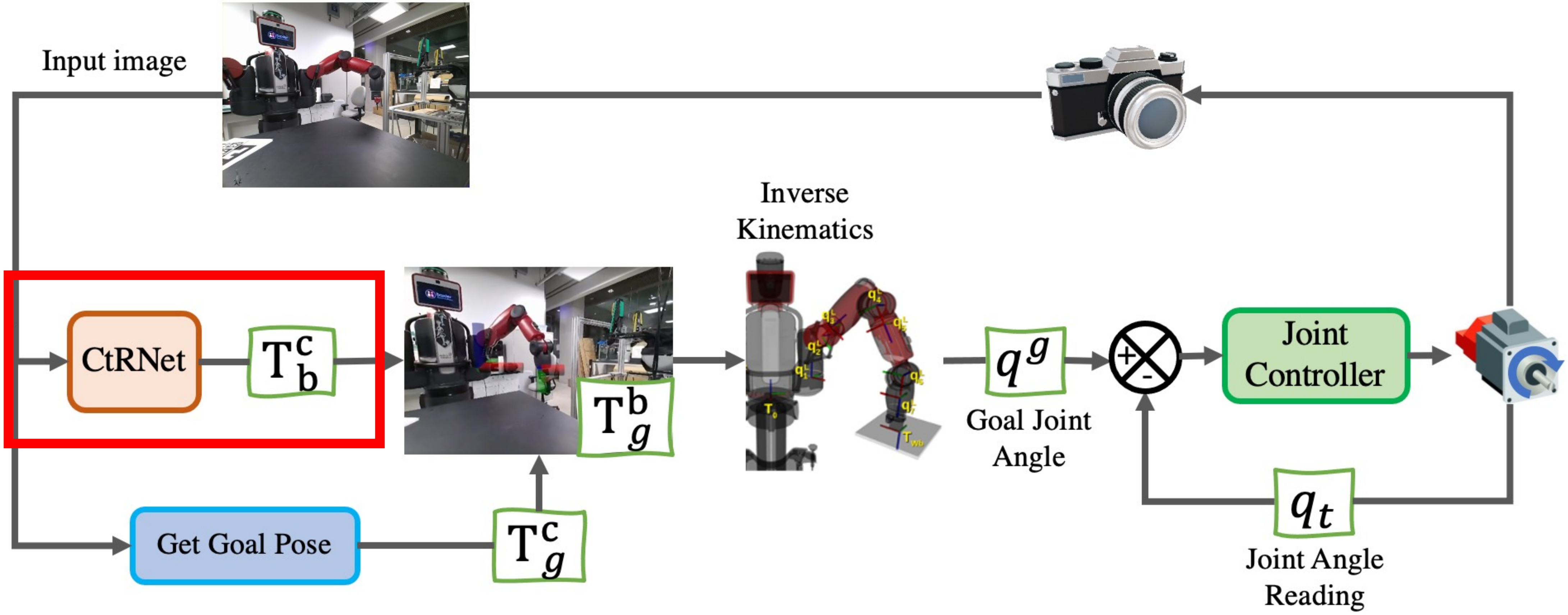
Table 1. Comparison of our methods with the state-of-the-art methods on DREAM-real datasets using ADD metric. We report the mean and AUC of the ADD on each dataset and the overall accuracy.

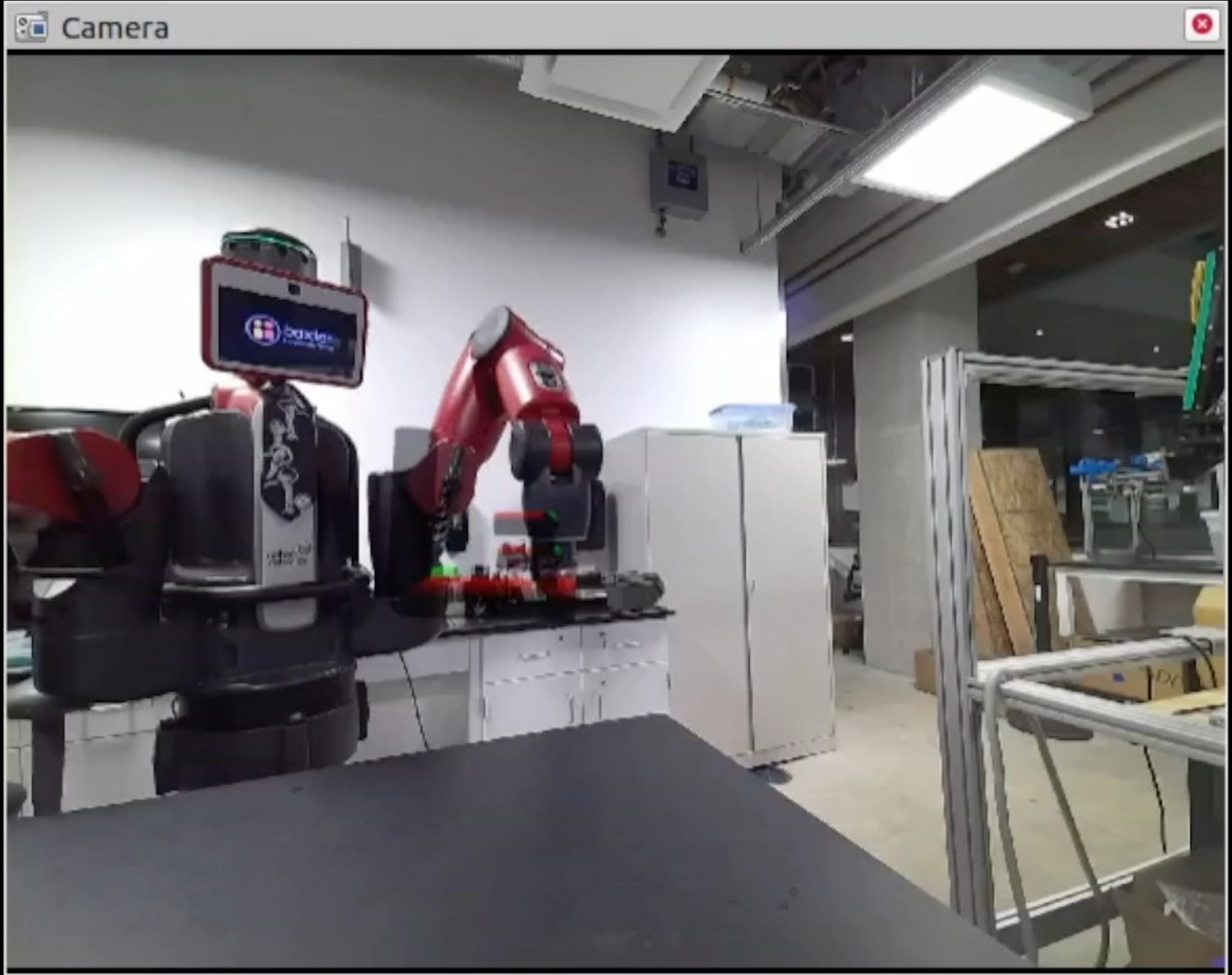
Method	Category	PCK (2D)		Mean 2D Err.	ADD (3D)		Mean 3D Err.
		@50 pixel ↑	AUC ↑	(pixel) ↓	@100 mm ↑	AUC ↑	(mm) ↓
Aruco Marker [14]	Keypoint	0.49	57.15	286.98	0.30	43.45	2447.34
DREAM-Q [29]	Keypoint	0.33	44.01	1277.33	0.32	40.63	386.17
Opt. Keypoints [34]	Keypoint	0.69	75.46	49.51	0.47	65.66	141.05
Diff. Rendering	Rendering	0.74	78.60	42.30	0.78	81.15	74.95
CtRNet	Keypoint	**0.99**	**93.94**	**11.62**	**0.88**	**83.93**	**63.81**

Qualitative results



Visual servoing with moving camera




Camera

If you have any question…

UC San Diego

JACOBS SCHOOL OF ENGINEERING

Electrical and Computer Engineering

ARClab

UCSD Advanced Robotics and Controls

https://github.com/ucsdarclab/CtRNet-robot-pose-estimation