SGTAPose: Robot Structure Prior Guided Temporal Attention for Camera-to-Robot Pose Estimation from Image Sequence

Yang Tian*, Jiyao Zhang*, Zekai Yin*, Hao Dong†
CFCS Peking University

Tag : WED-AM-066

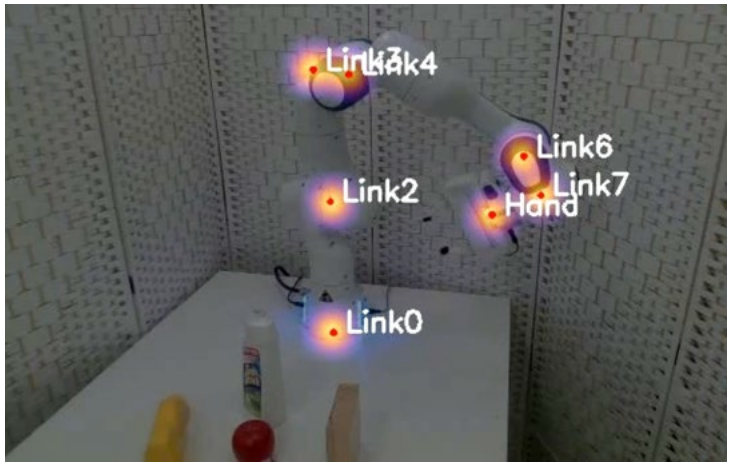
Camera-to-Robot Pose Estimation from Image Sequence

Inputs:

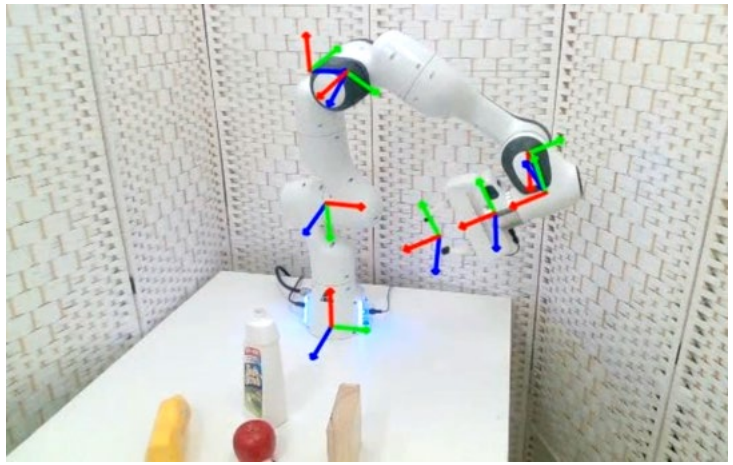
Image Sequence

Outputs:

Belief Map

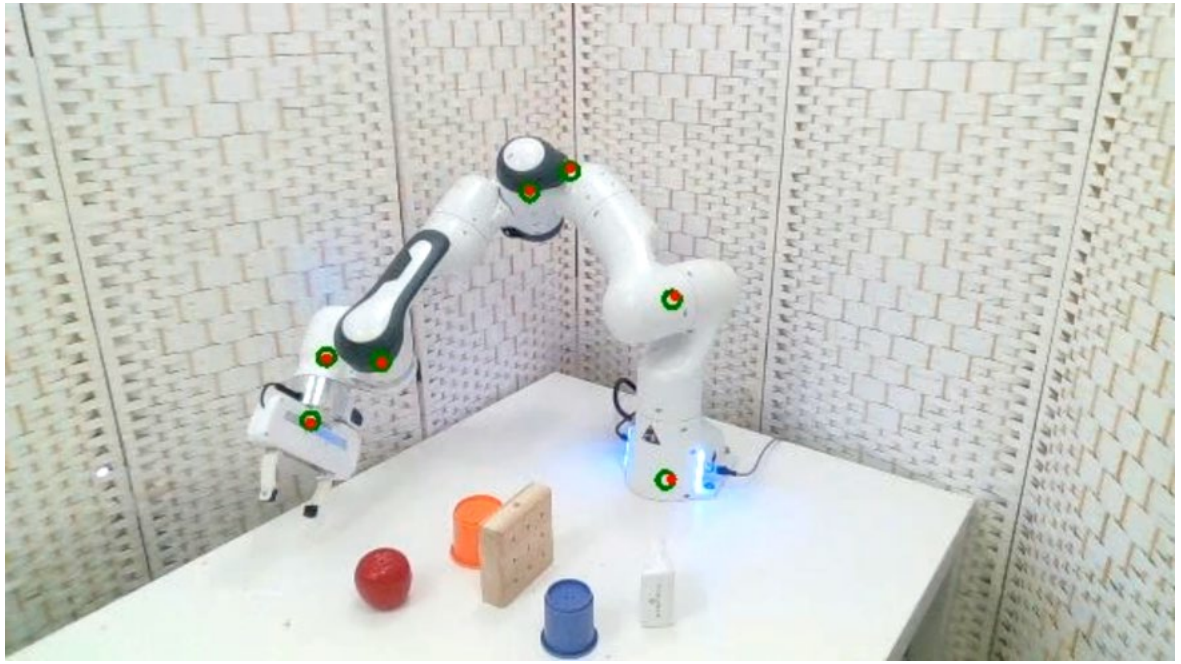


Camera-to-Robot 6D Pose

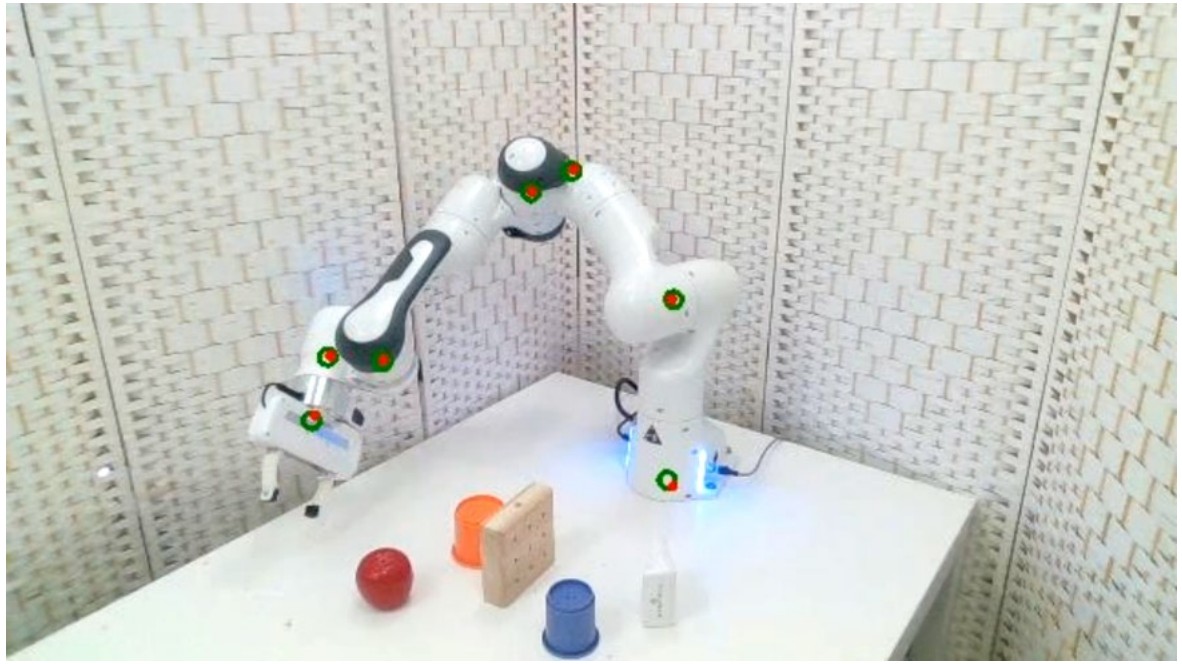


Results: Camera-to-Robot Pose Estimation on Real Data

Our Predictions



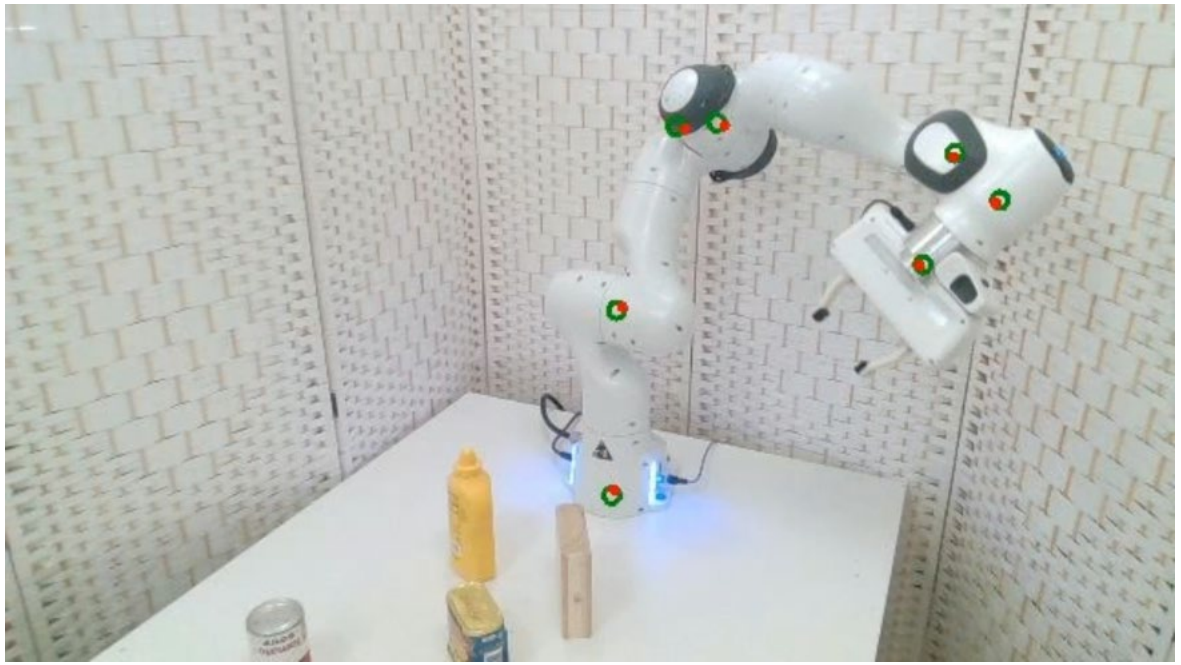
Dream

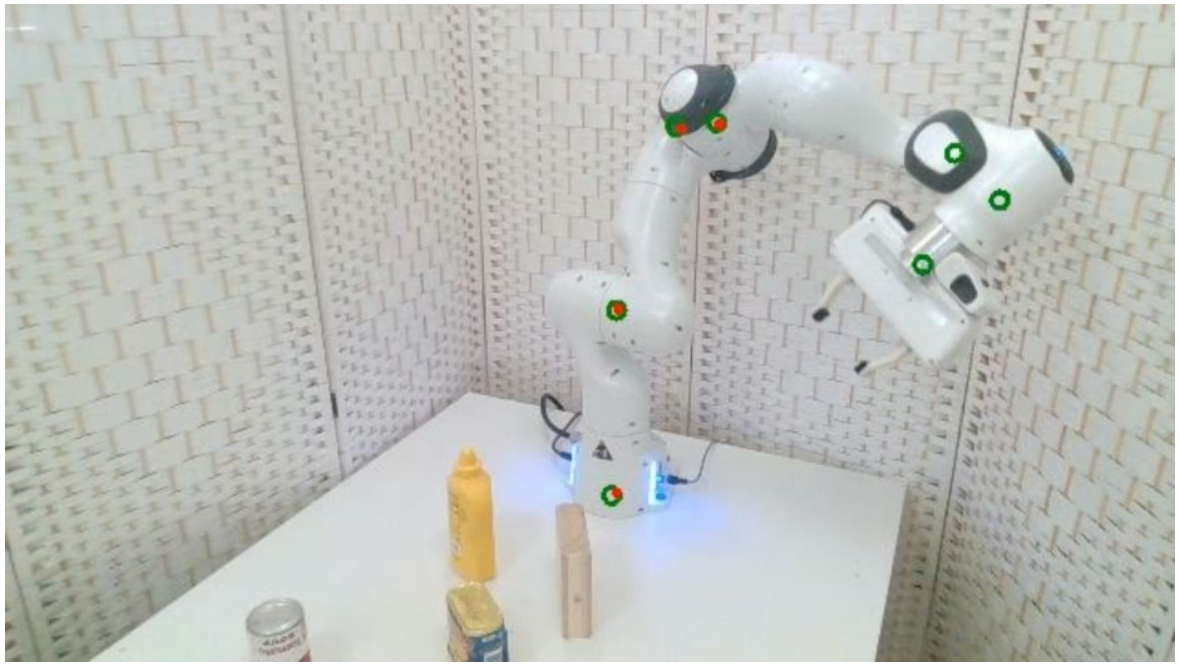


Results: Camera-to-Robot Pose Estimation on Real Data

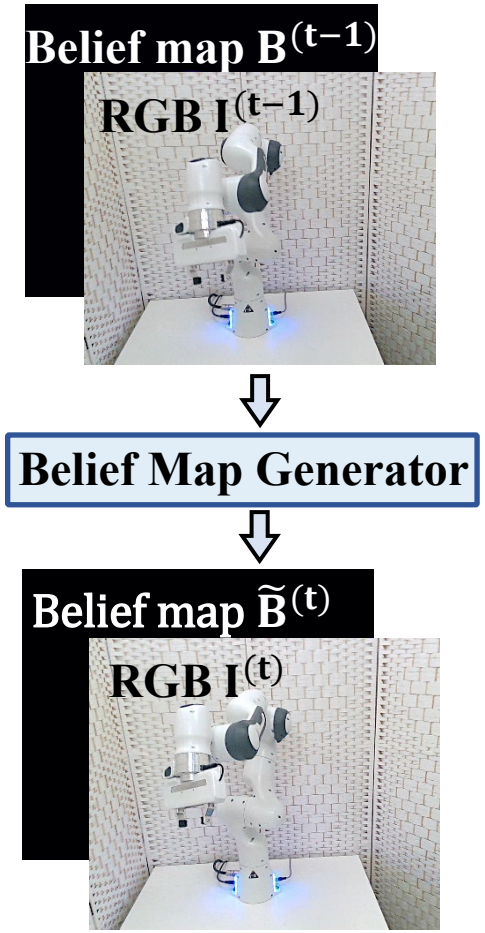
Our Predictions

Dream

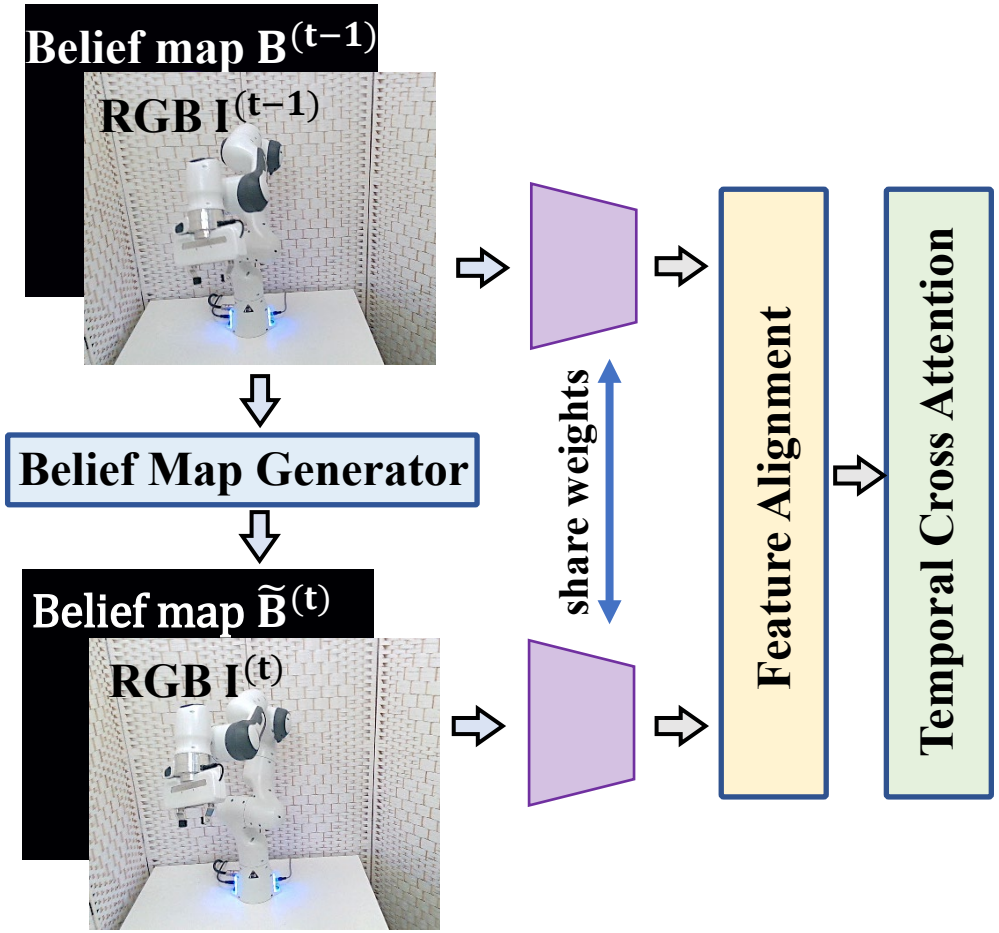




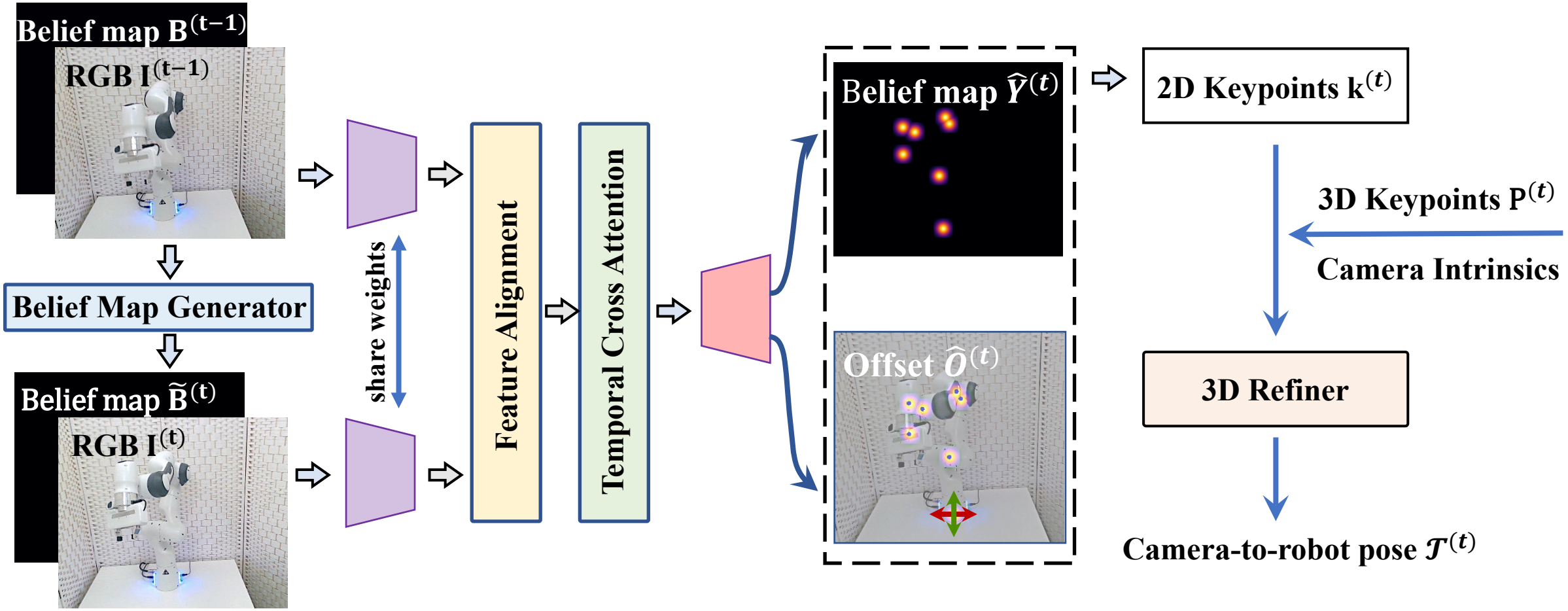
Camera-to-Robot Pose Estimation Scheme



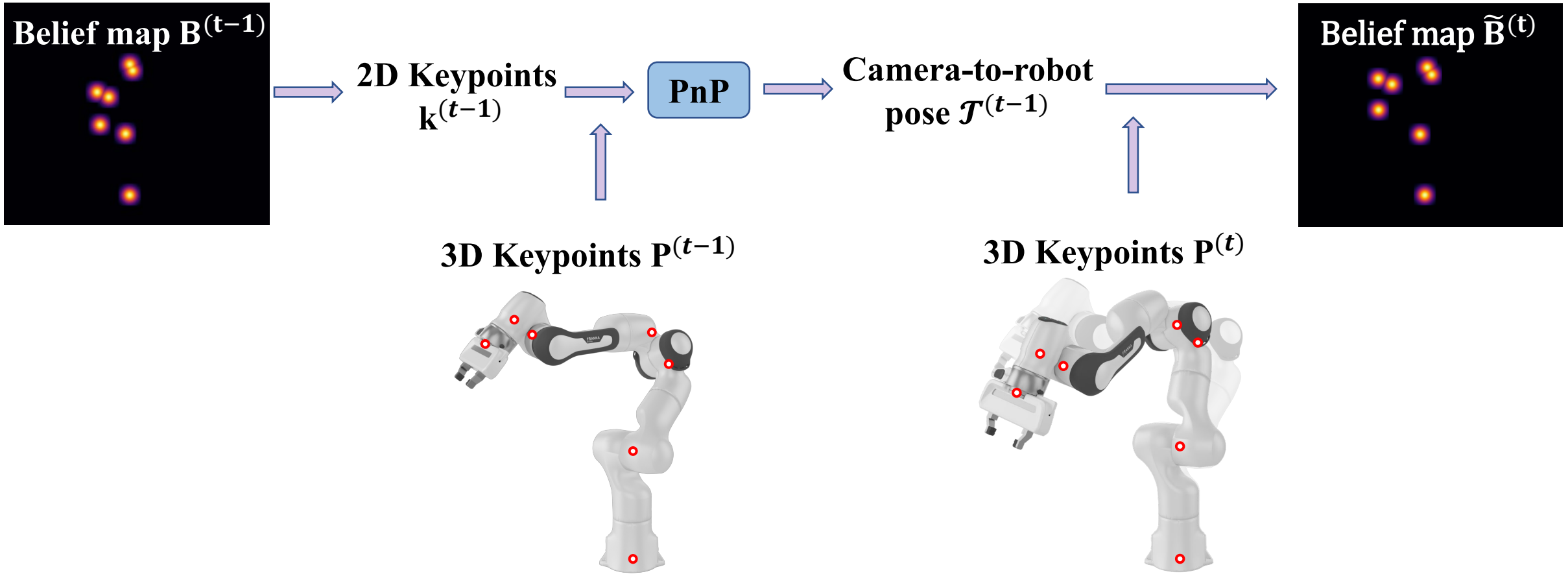
Camera-to-Robot Pose Estimation Scheme



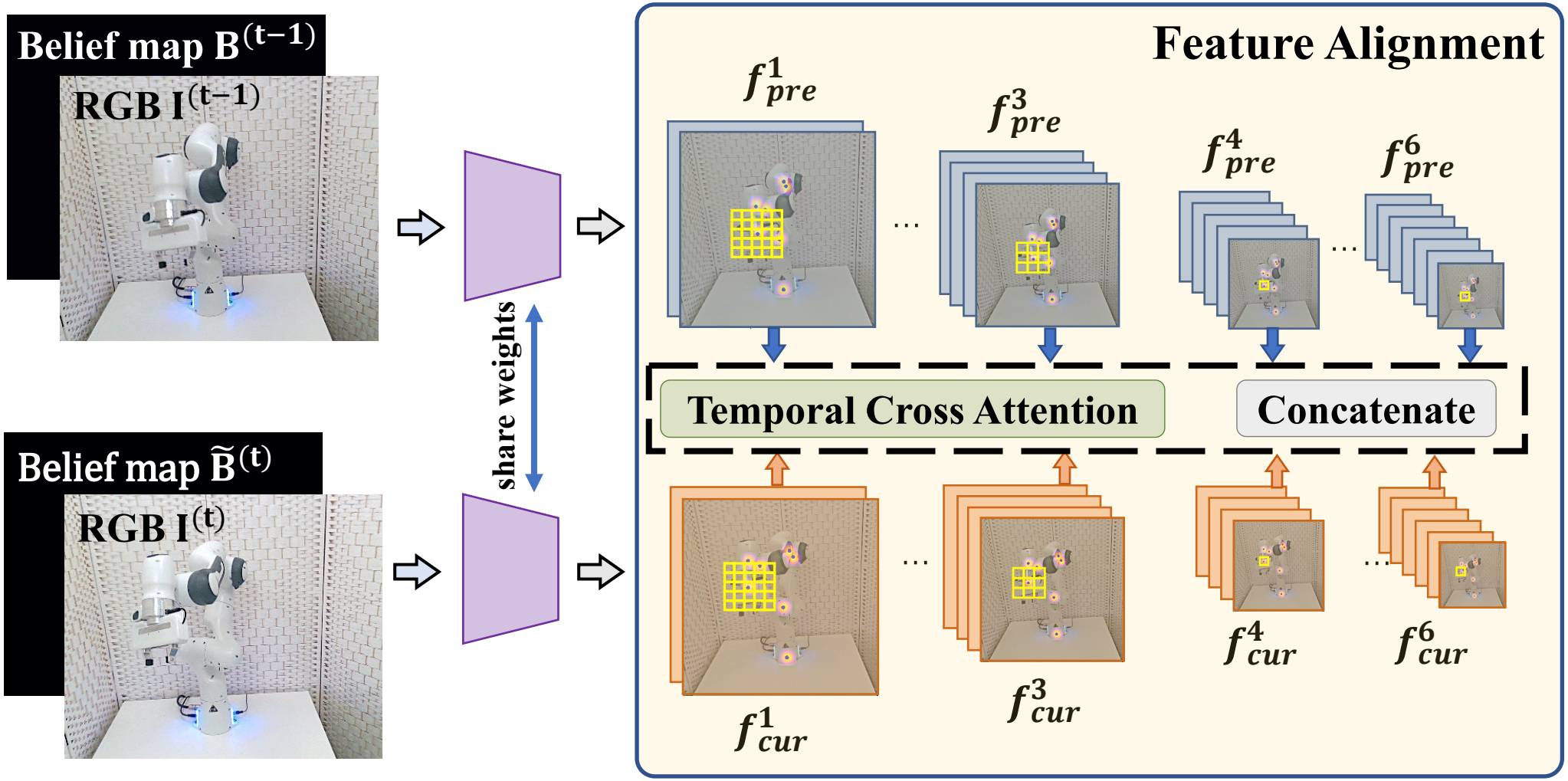
Camera-to-Robot Pose Estimation Scheme



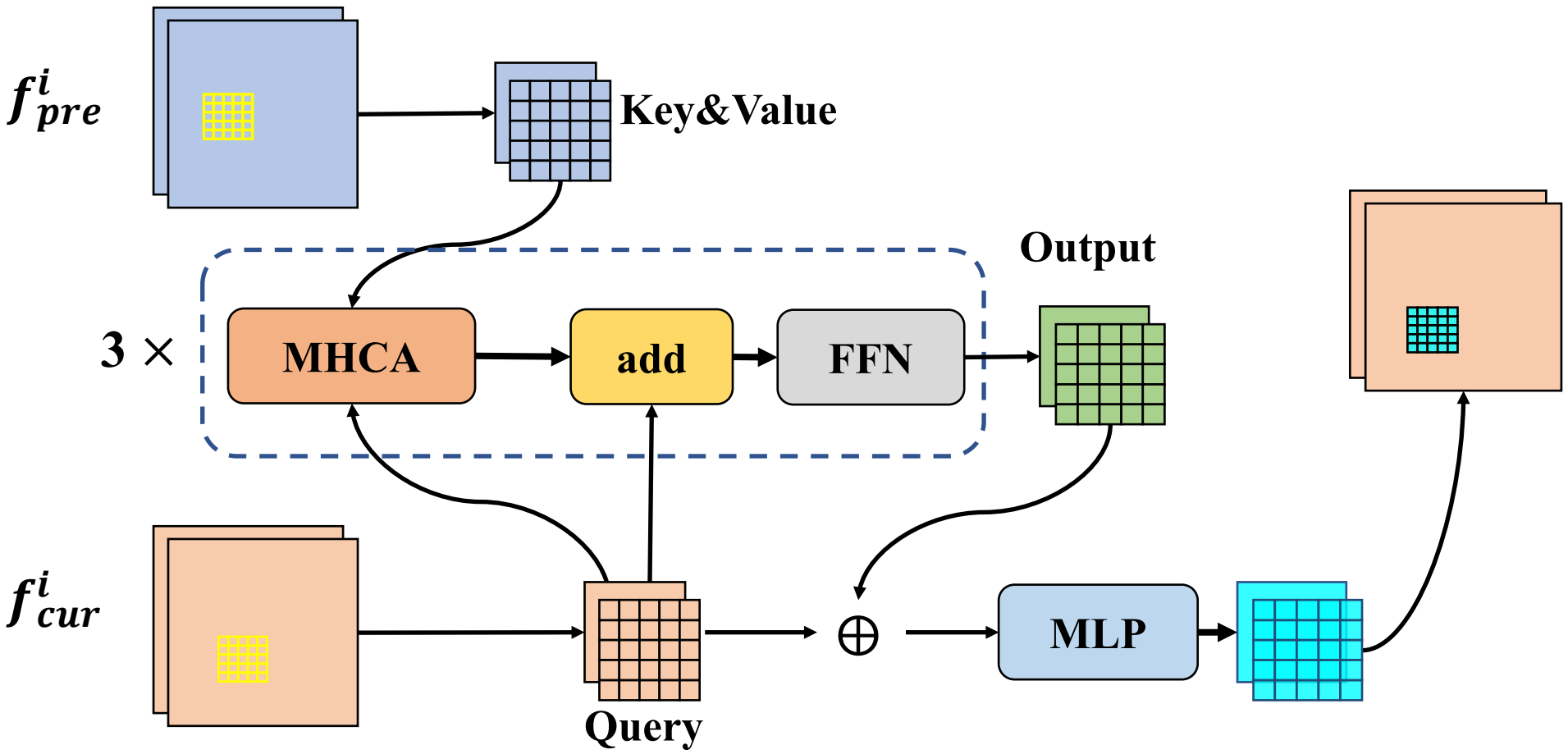
Framework: Belief Map Generator



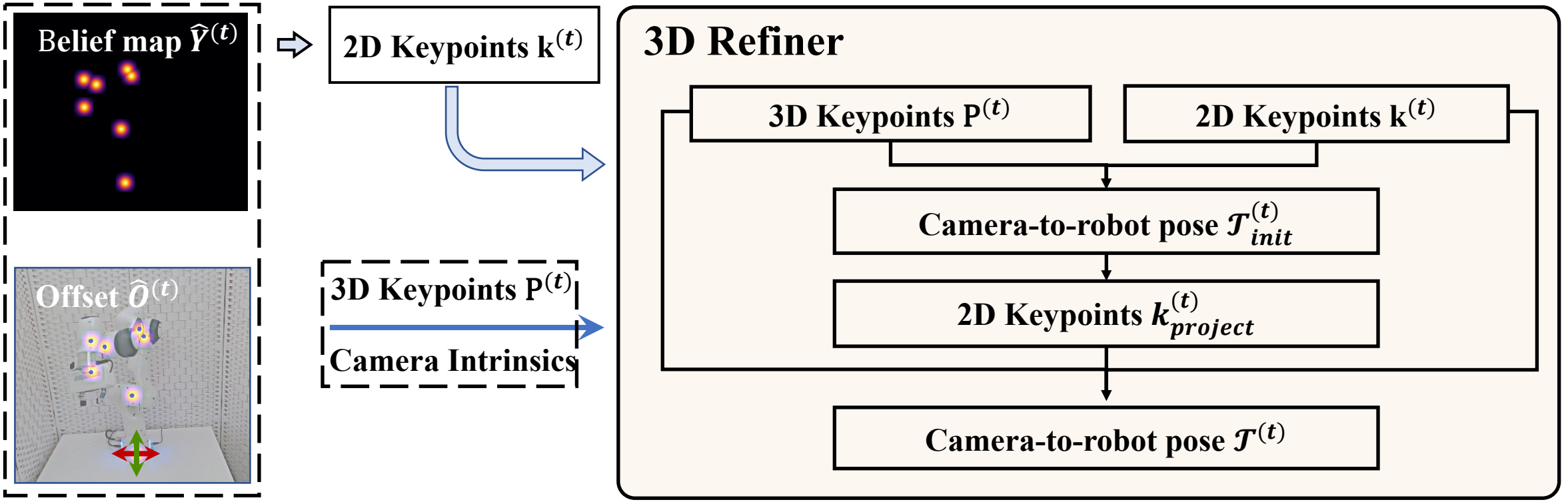
Framework: Feature Alignment



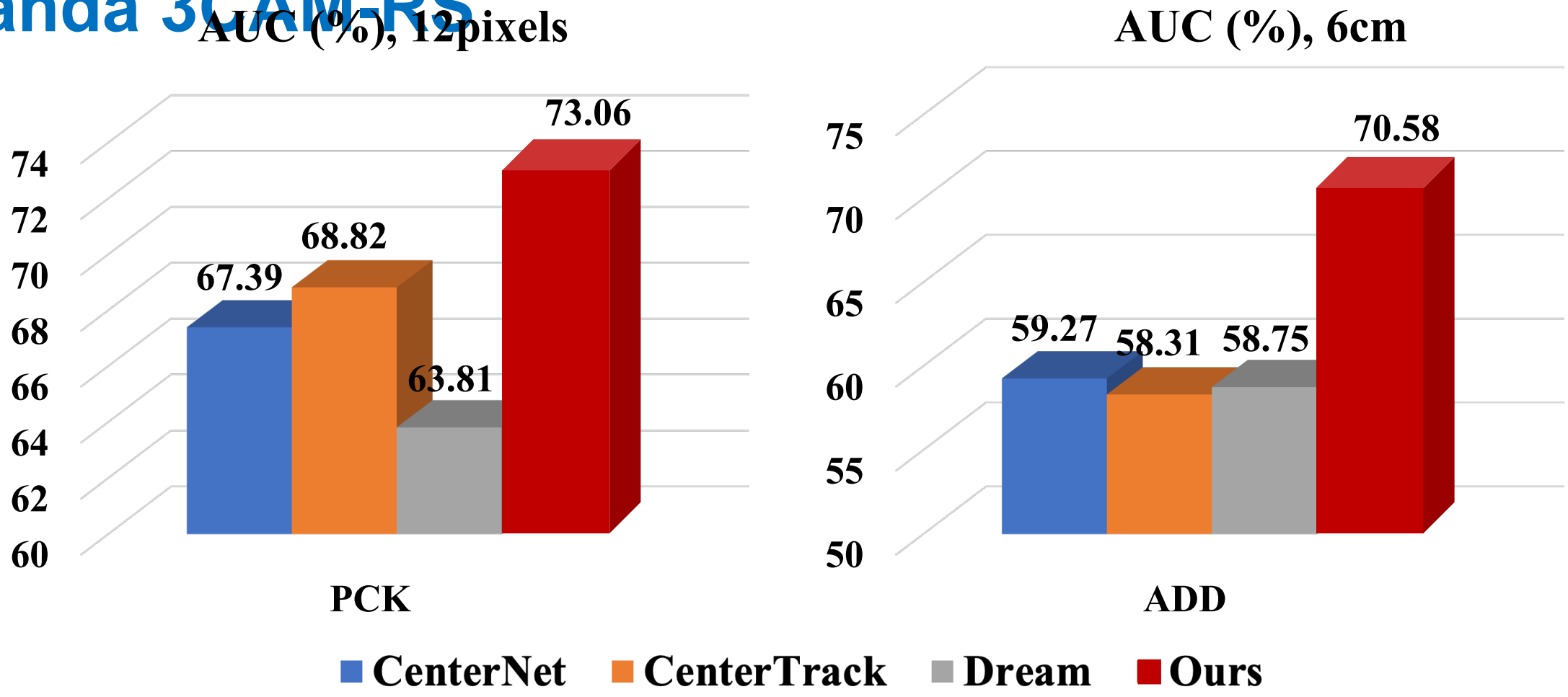
Framework: Temporal Cross Attention



Framework: 3D Refiner



Results: Camera-to-Robot Pose Estimation on Panda 3CAM-RS

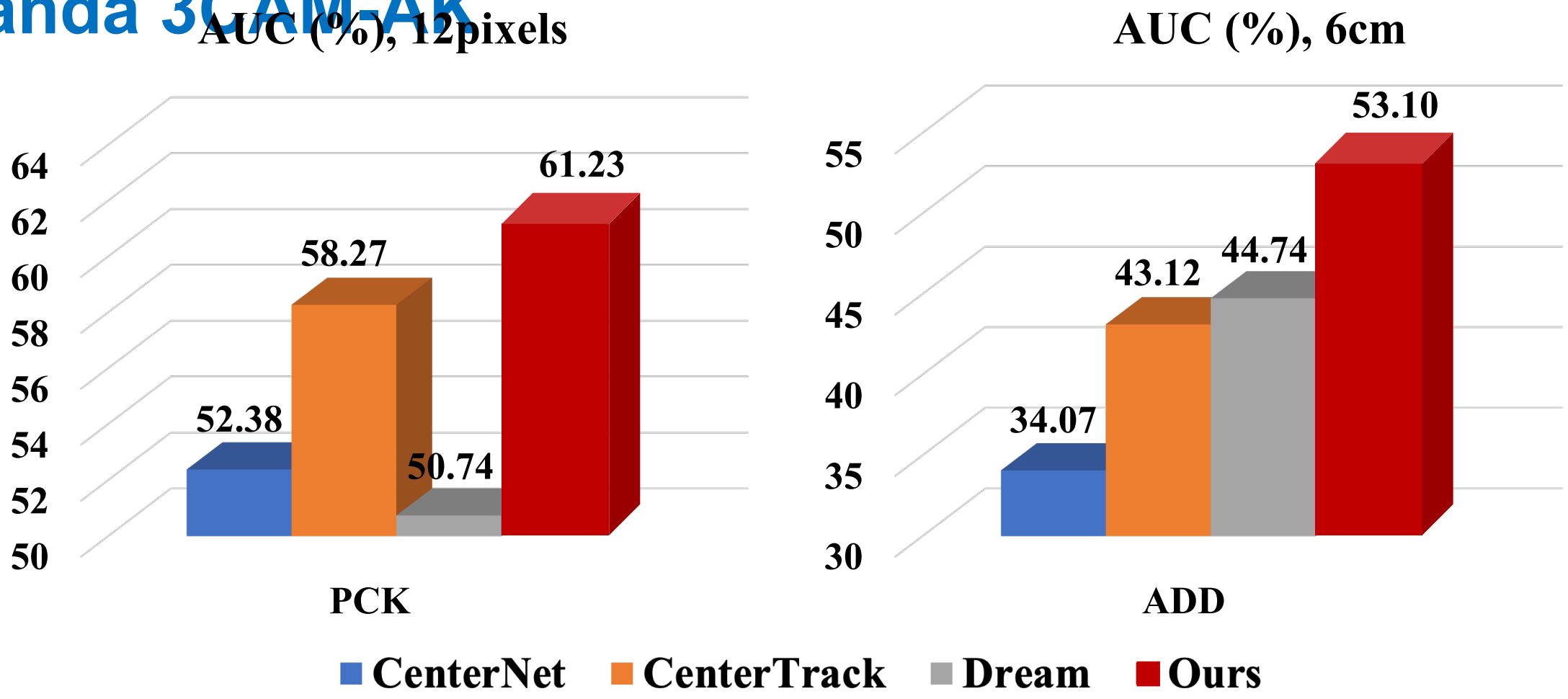


Duan, Kaiwen, et al. "Centernet: Keypoint triplets for object detection." *ICCV* 2019

Zhou, Xingyi, et al. "Tracking objects as points." *ECCV* 2020

Lee, Timothy E., et al. "Camera-to-robot pose estimation from a single image." *ICRA* 2020

Results: Camera-to-Robot Pose Estimation on Panda 3CAM-AK

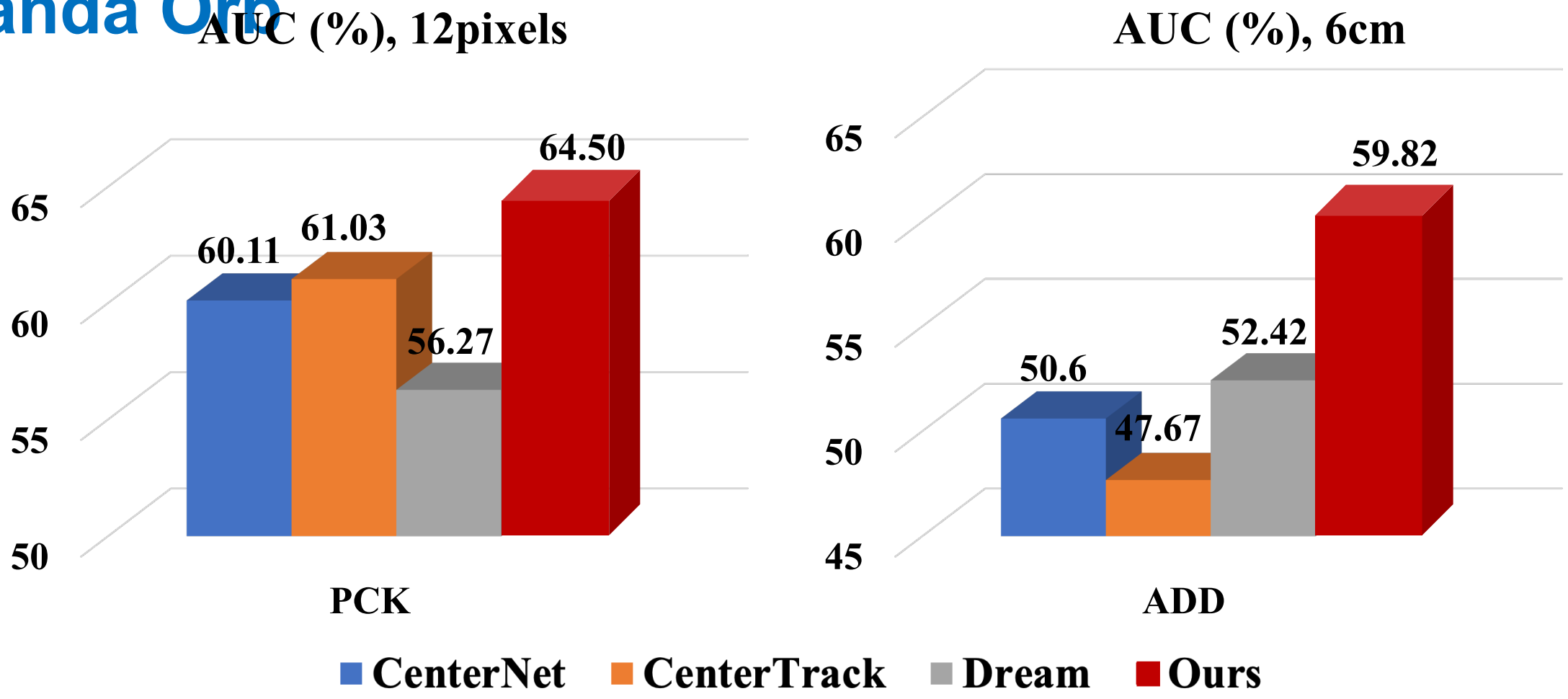


Duan, Kaiwen, et al. "Centernet: Keypoint triplets for object detection." *ICCV* 2019

Zhou, Xingyi, et al. "Tracking objects as points." *ECCV* 2020

Lee, Timothy E., et al. "Camera-to-robot pose estimation from a single image." *ICRA* 2020

Results: Camera-to-Robot Pose Estimation on Panda Orb

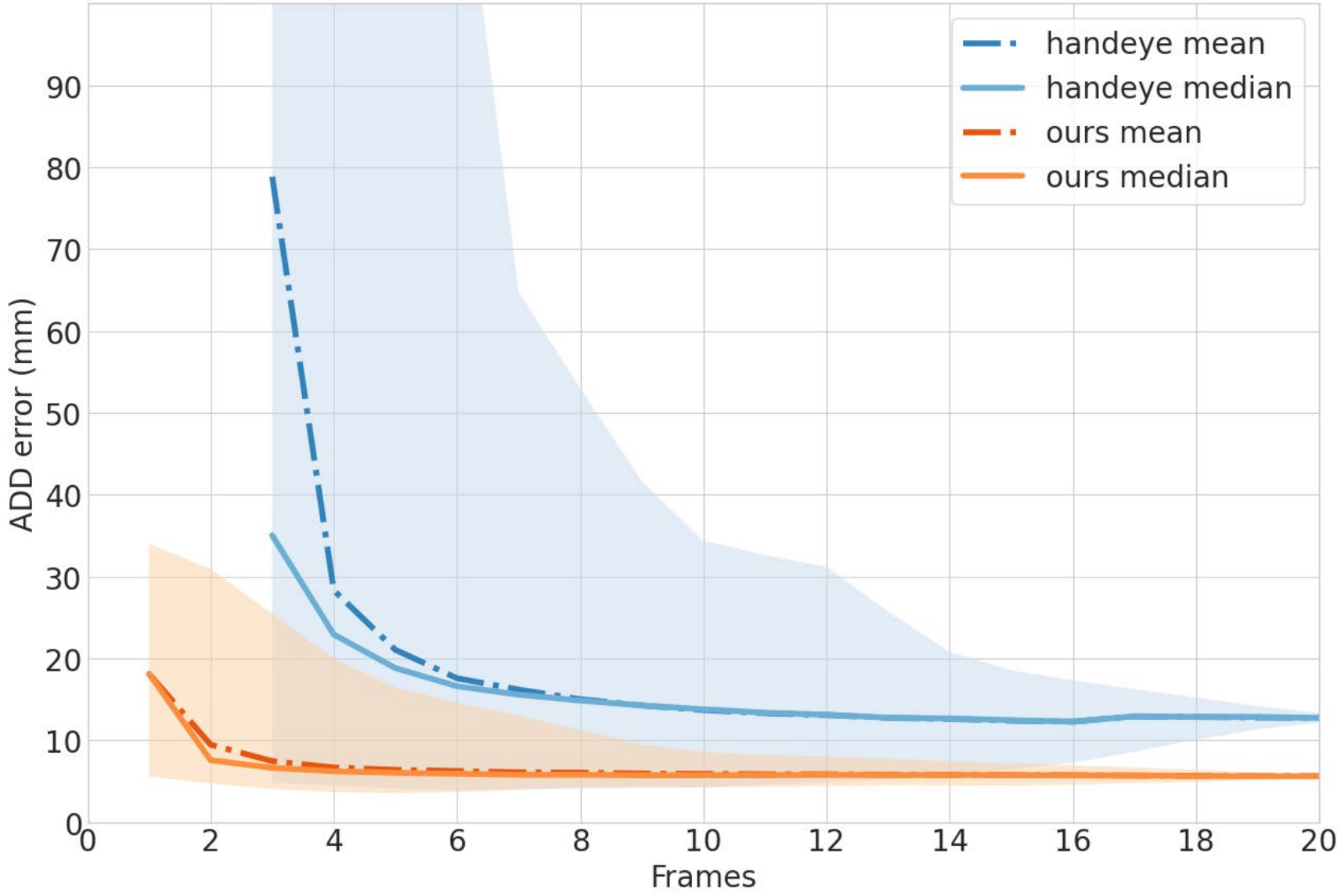


Duan, Kaiwen, et al. "Centernet: Keypoint triplets for object detection." *ICCV* 2019

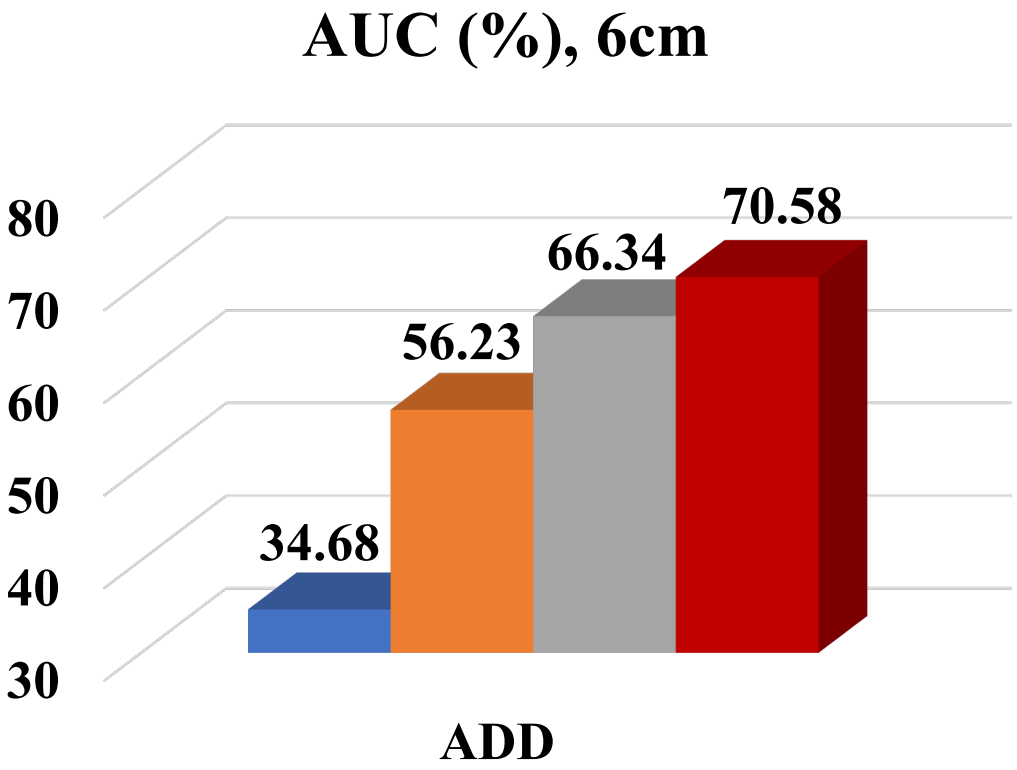
Zhou, Xingyi, et al. "Tracking objects as points." *ECCV* 2020

Lee, Timothy E., et al. "Camera-to-robot pose estimation from a single image." *ICRA* 2020

Results: Compare with Classic Hand-Eye Calibration



Ablation Study on Camera-to-Robot Pose Estimation



- Feature Alignment
- Temporal Cross Attention
- 3D Refiner

Downstream Task: Robotic Grasping

Static Experiment

Downstream Task: Robotic Grasping

Dynamic Experiment #1

Downstream Task: Robotic Grasping

Dynamic Experiment #2

Downstream Task: Robotic Grasping

Dynamic Experiment #3

Conclusion

- We propose a novel pipeline for **camera-to-robot pose estimation** from **single-view successive frames**.

- With the **robot structure prior** guidance, our method can efficiently fuse the **temporal feature** from different frames.

- Our method demonstrates significant improvements over several datasets, strong dominance compared with traditional hand-eye calibration, and high accuracy and stability in downstream grasping tasks.