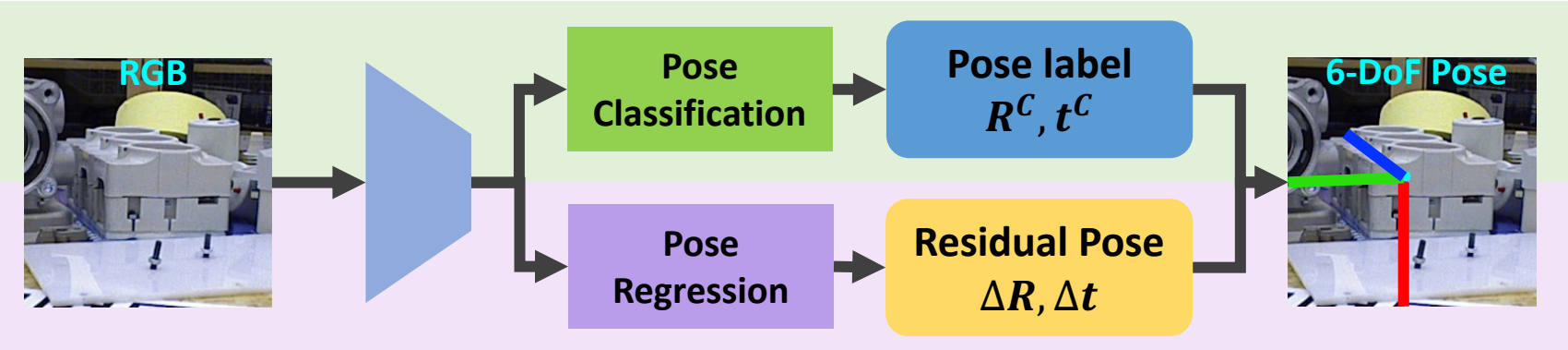
MRC-Net: 6-DoF Pose Estimation with MultiScale Residual Correlation

Yuelong Li*, Yafei Mao*, Raja Bala, Sunil Hadap

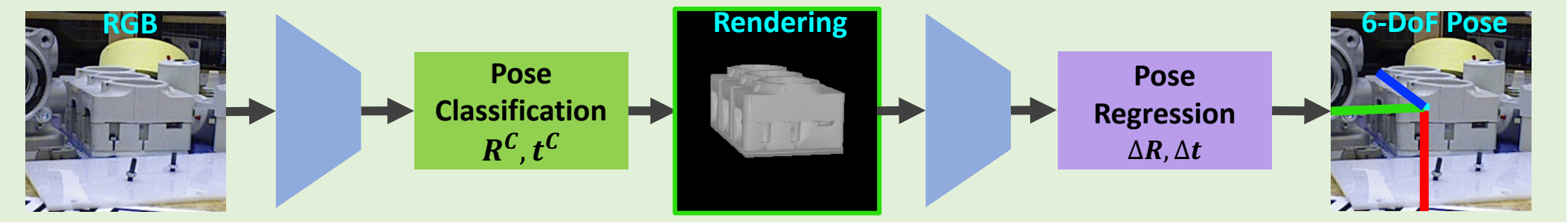
Motivation

- Previous methods combine classification and regression to leverage their complementary merits.
- This is typically achieved via multitask learning of coarse pose classification and residual regression.



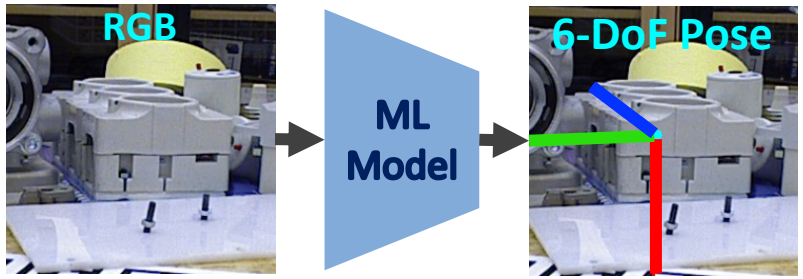
Motivation

- Previous methods combine classification and regression to leverage their complementary merits.
- This is typically achieved via multitask learning of coarse pose classification and residual regression.
- **In contrast**, we show that a **sequential** design can be more effective.

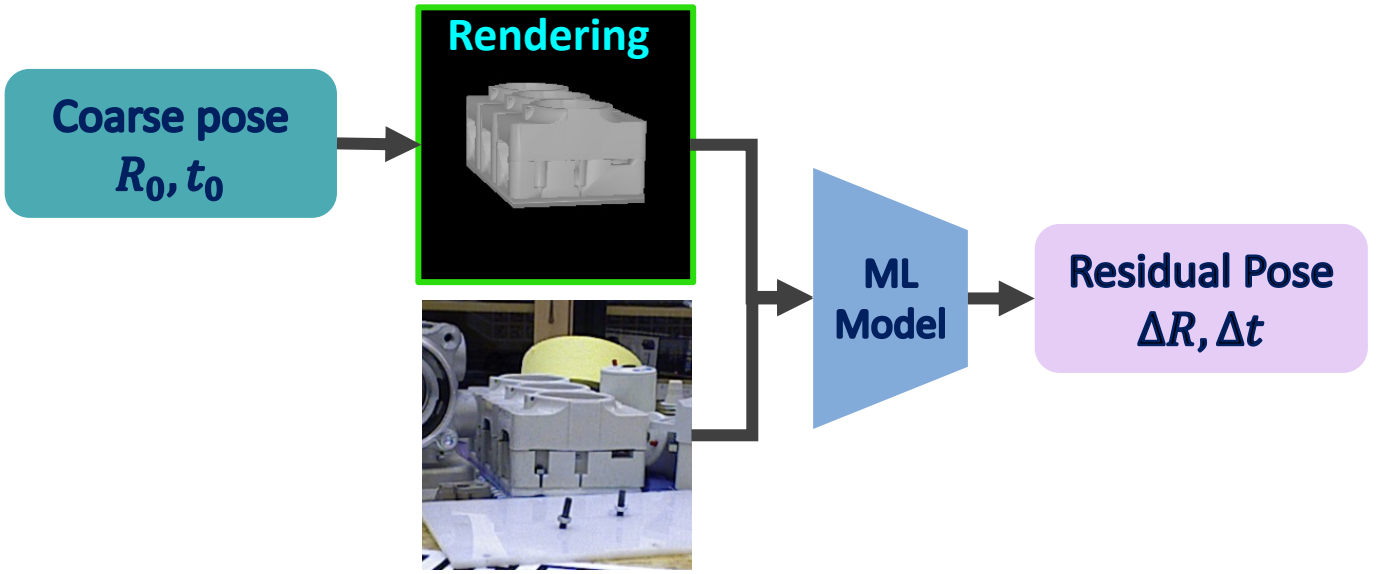


Previous Work

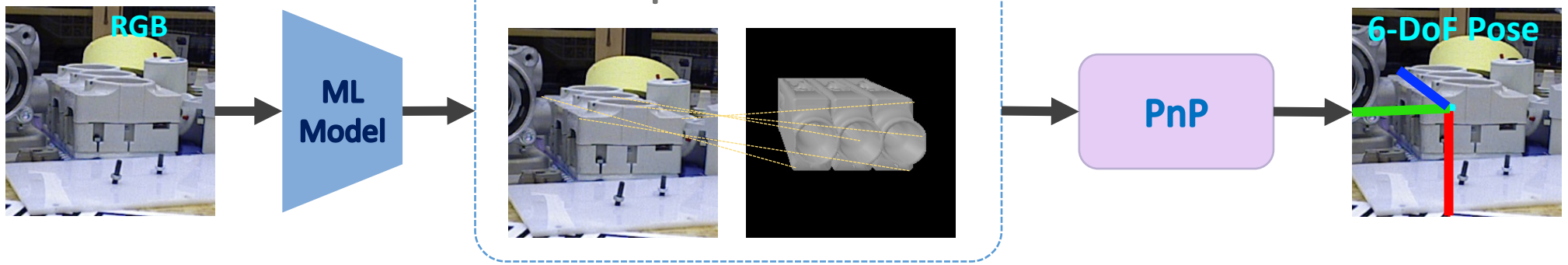
Direct pose estimation



Render-and-compare



Correspondence-based methods



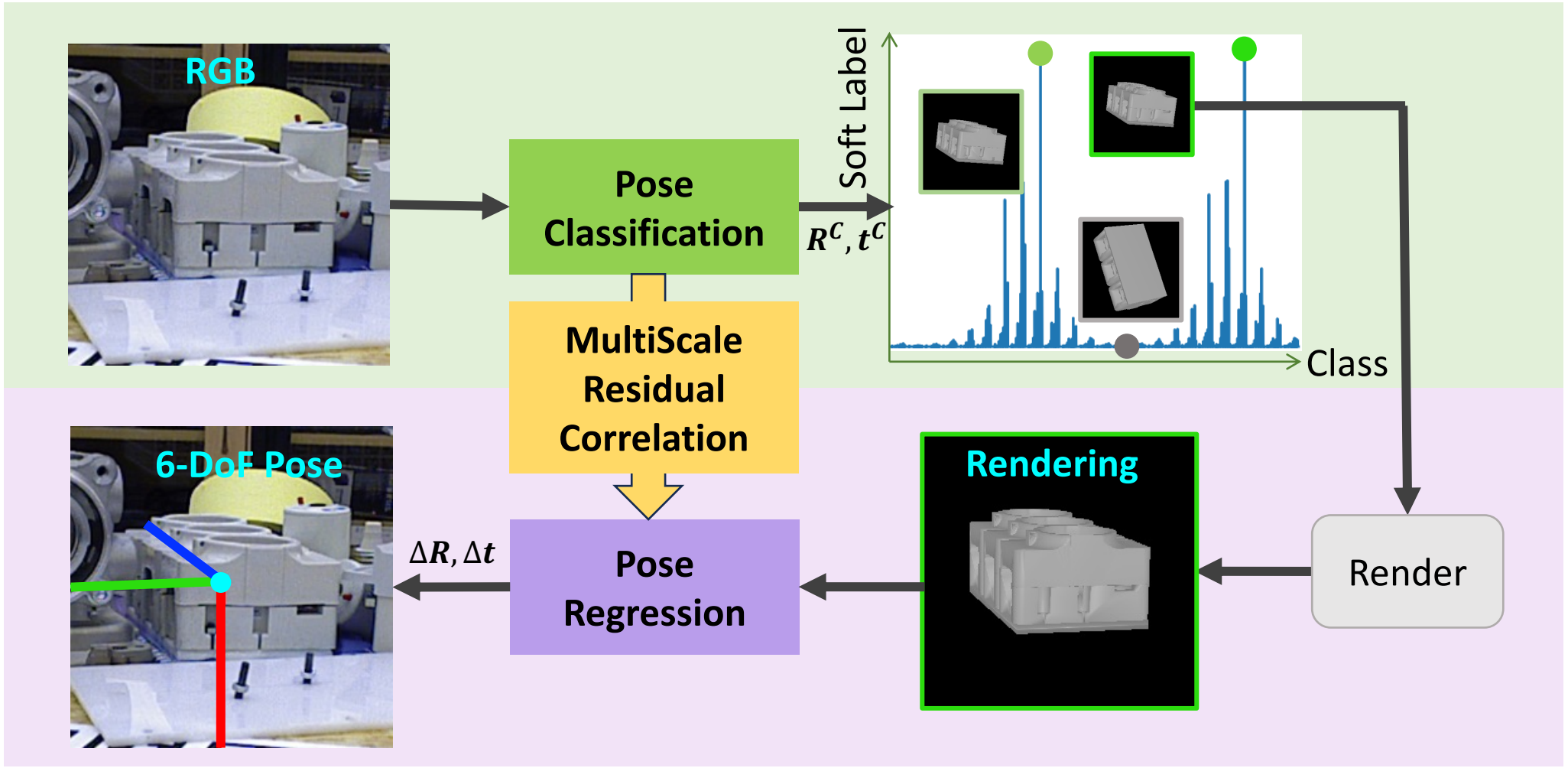
Our Contributions

- A **single shot** baseline model without pre-initialization, post-processing and iterative refinement.

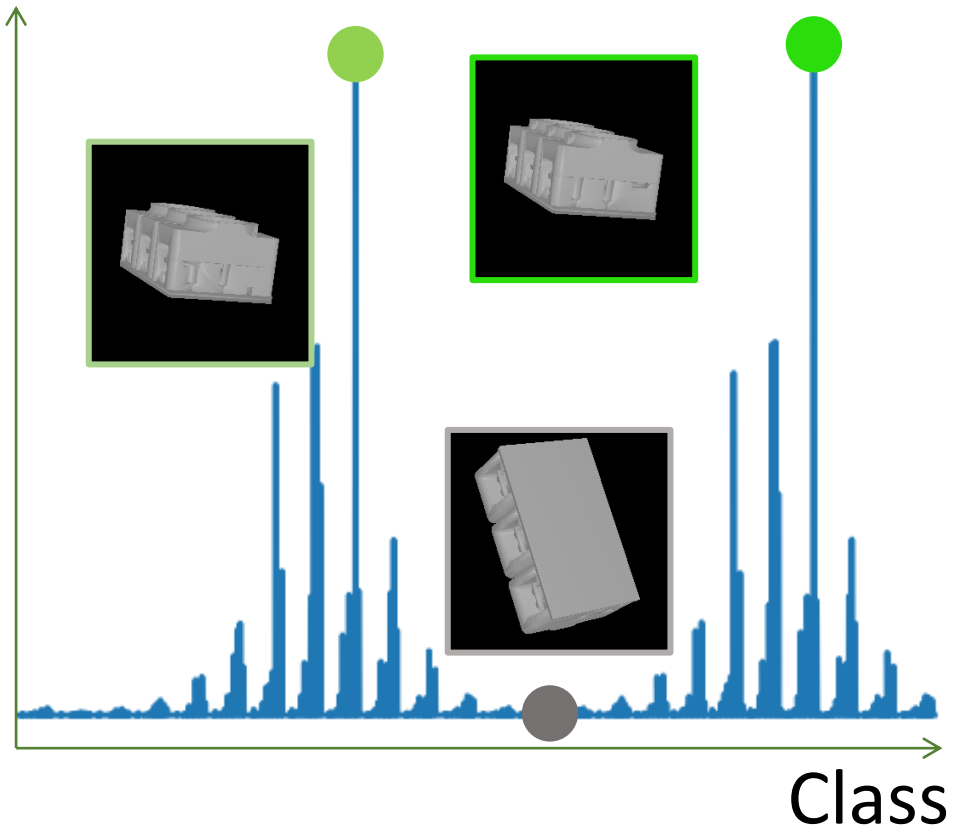
- Propose **sequential design** with **soft labels** and **multiscale residual correlation (MRC)** layer.
  - Together gives **+12.2%** average recall over the parallel baseline.

- SOTA performance with near real-time efficiency.
  - Achieves **+2.4%** average recall improvement over reported results.

Proposed Method



Soft Rotation Labels

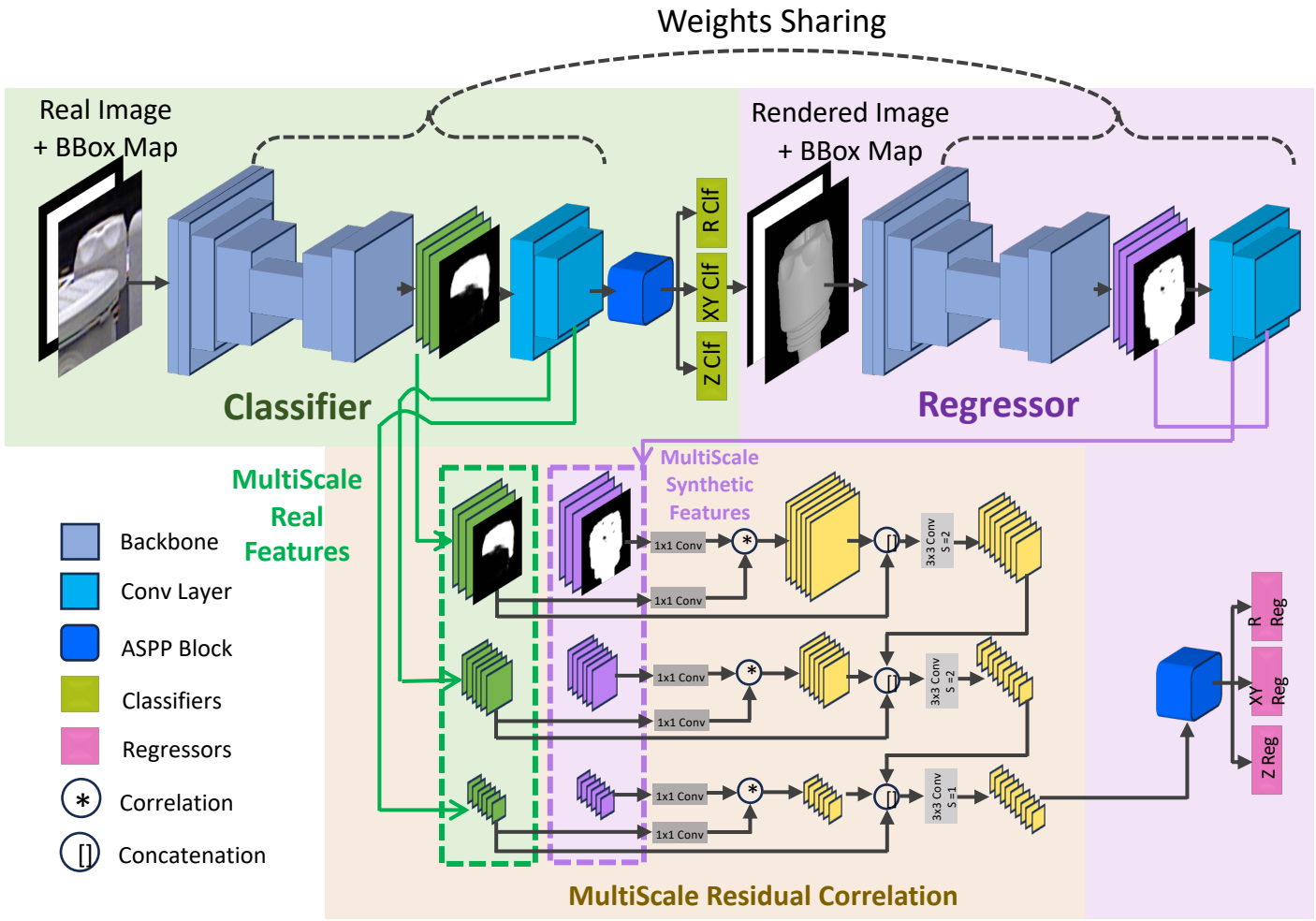


- Conventional hard labels do not account for object symmetry
- Solution: use soft probabilistic labels

$$l_{\mathrm{k}}^{\mathrm{R}} = \exp\left\{-\frac{\rho_{\mathrm{sym}}(R^*, t^*; R_k, t^*)}{\sigma}\right\}$$

Symmetry aware distance

Network Architecture



Comparison with SOTA

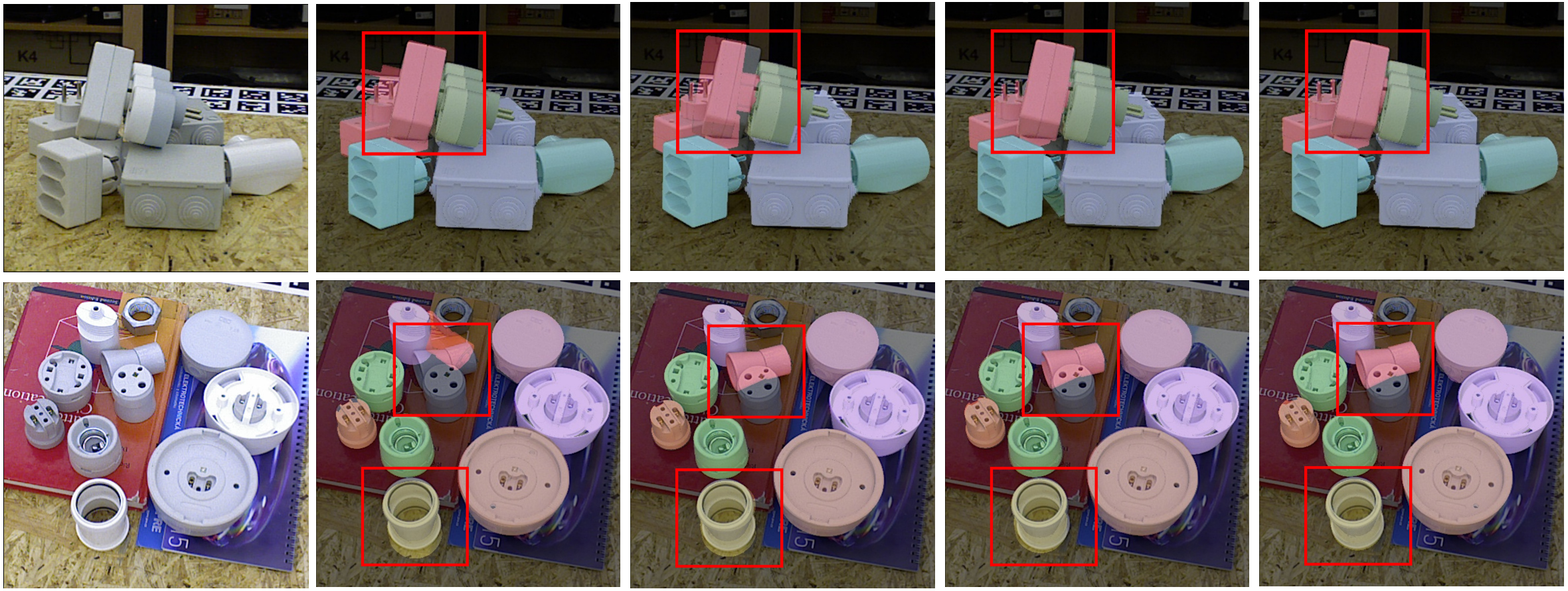
Method	T-LESS	ITODD	YCB-V	LM-O	Avg.
EPOS	46.7	18.6	49.9	54.7	42.5
CDPNv2	40.7	10.2	39.0	62.4	38.1
CosyPose	64.0	21.6	57.4	63.3	51.6
SurfEmb	74.1	38.7	65.3	65.6	60.9
SC6D	73.9	30.3	61.0	-	-
SCFlow	-	-	65.1	68.2	-
NCF	-	-	67.3	63.2	-
Ours	**77.1**	**39.3**	**68.1**	**68.5**	**63.3**

Training on synthetic PBR samples

Method	T-LESS	YCB-V
CDPNv2	47.8	53.2
CosyPose [1]	72.8	**82.1**
SurfEmb [2]	77.0	71.8
SC6D [3]	78.0	78.8
CIR	71.5	82.4
Ours	**79.8**	81.7

Fine-tuning on real samples

Qualitative Samples



Input image PFA SC6D ZebraPose Ours

Ablation Studies

Method	AR_{VSD}	AR_{MSSD}	AR_{MSPD}	AR
Classifier Only	55.6	62.8	77.4	65.3
Parallel design	55.3	62.7	76.6	64.9
Hard Label	67.8	72.2	84.6	74.8
No MRC	69.4	73.5	85.7	76.2
Full Model	70.6	74.7	86.0	77.1

Effectiveness of soft labels

Effectiveness of sequential design

Effectiveness of MRC layer

Conclusions

- Our sequential design brings significant performance improvement.
- MRC layer more effectively captures feature correspondences.
- Entire pipeline outperforms SOTA method while maintaining near real time efficiency.