JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Knowledge Distillation for 6D Pose Estimation by Aligning Distributions of Local Predictions

Shuxuan Guo[1] Yinlin Hu[2] Jose M. Alvarez[3] Mathieu Salzmann[1,4]

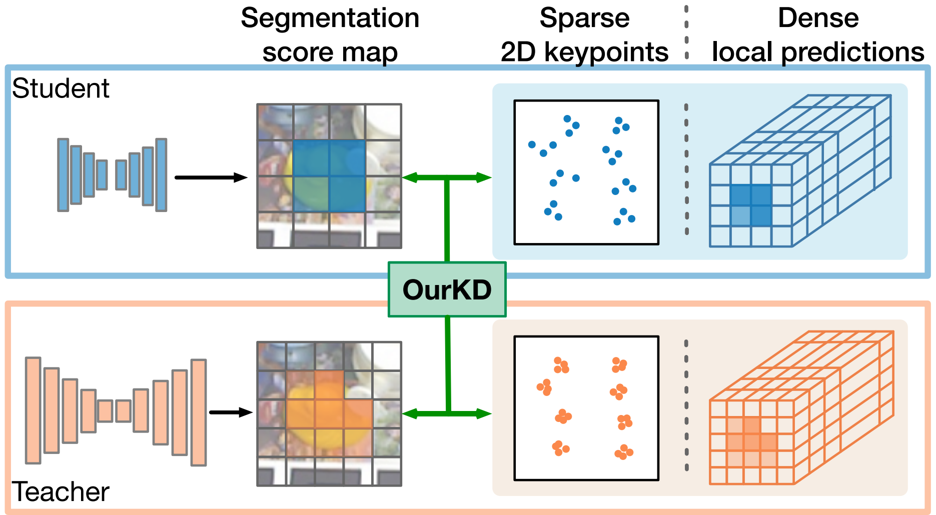
[1]CVLab, EPFL [2]MagicLeap [3]NVIDIA [4]ClearSpace

THU-AM-206

CVPR 2023

Preview

Task-driven KD for 6D pose estimation



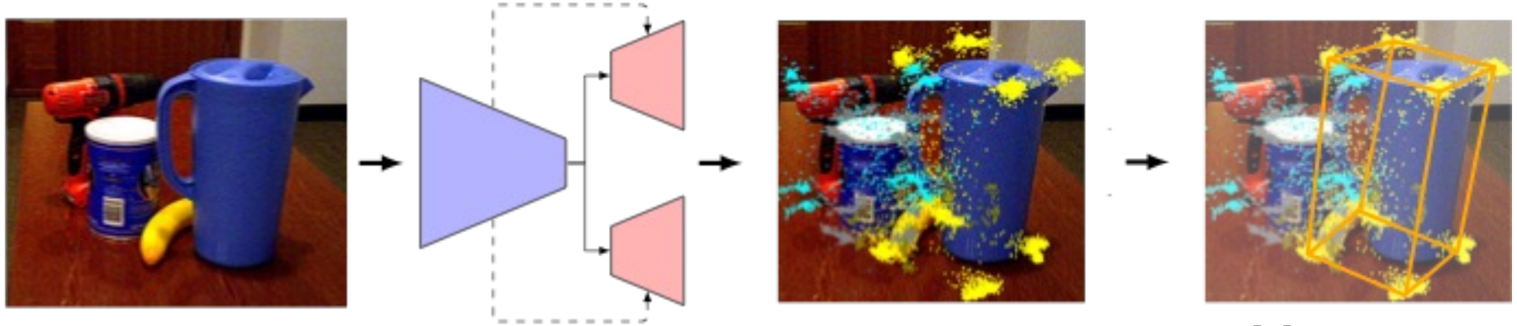
jointly distill ***local prediction distribution***
+ ***segmentation score map***


Code

- Motivation
  - Compact student networks typically struggle to predict local sparse/dense predictions precisely.
- Method
  - Align local prediction distributions and segmentation score maps based on optimal transport algorithm.
- Take home messages
  - The **first** knowledge distillation in the context of 6D pose estimation.
  - Our KD **generalizes** to both sparse keypoints and dense predictions 6D pose estimation frameworks.
  - Our KD can be used **in conjunction with** feature distillations to further boost the student's performance.

Introduction

- Sparse keypoint-based
  - 8 corners of the 3D object bounding box

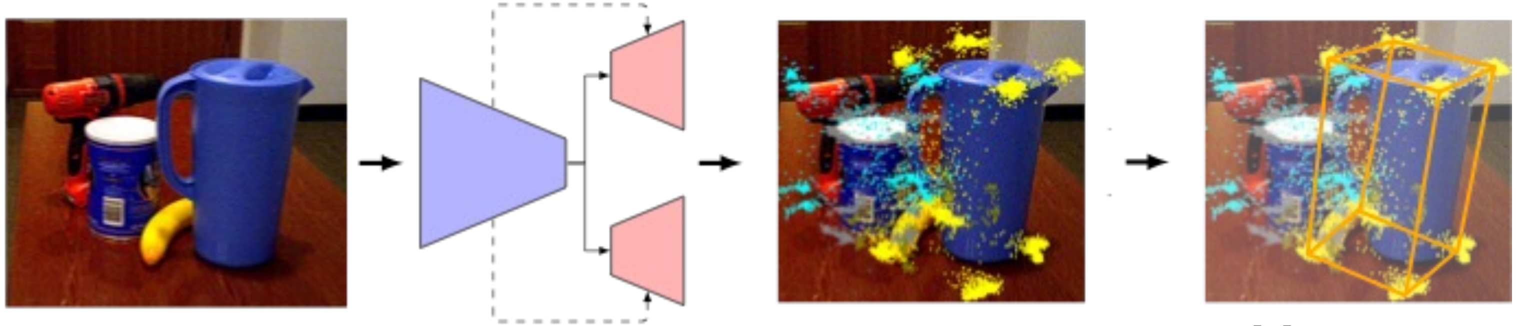


Segmentation-driven 6D Object Pose Estimation [1]

[1] Hu et al. "Segmentation-driven 6d object pose estimation." CVPR2019.
[2] Wang et al. "GDR-Net: Geometry-Guided Direct Regression Network for Monocular 6D Object Pose Estimation." CVPR2021.
[3] Su et al. "ZebraPose: Coarse to Fine Surface Encoding for 6DoF Object Pose Estimation." CVPR2022.

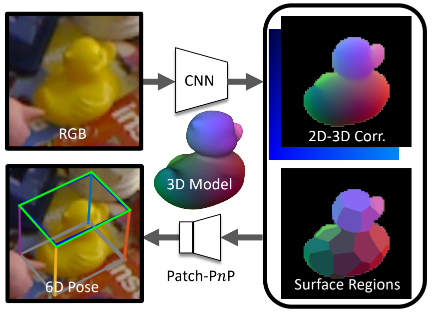
Introduction

- Sparse keypoint-based
  - 8 corners of the 3D object bounding box

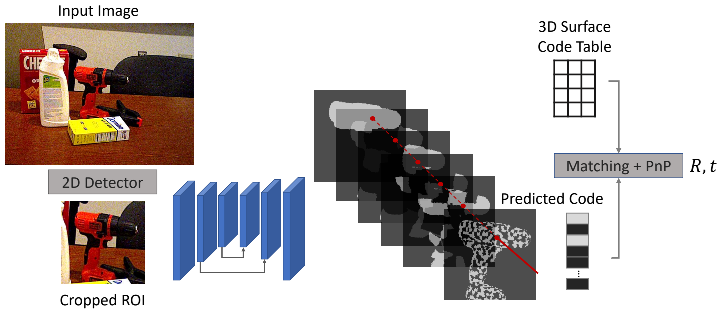


Segmentation-driven 6D Object Pose Estimation [1]

- Dense local prediction-based
  - Intermediate dense representations
    - Pixel-wise 2D-to-3D correspondence
    - Extra geometry features



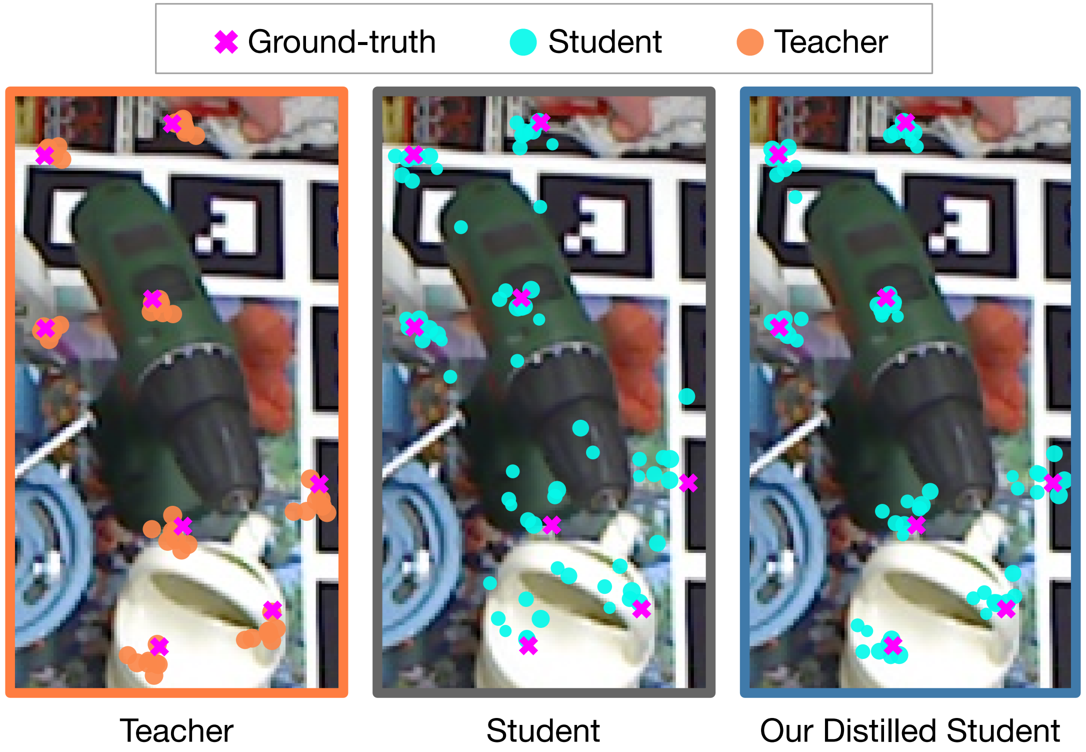
GDR-Net [2]



ZebraPose [3]

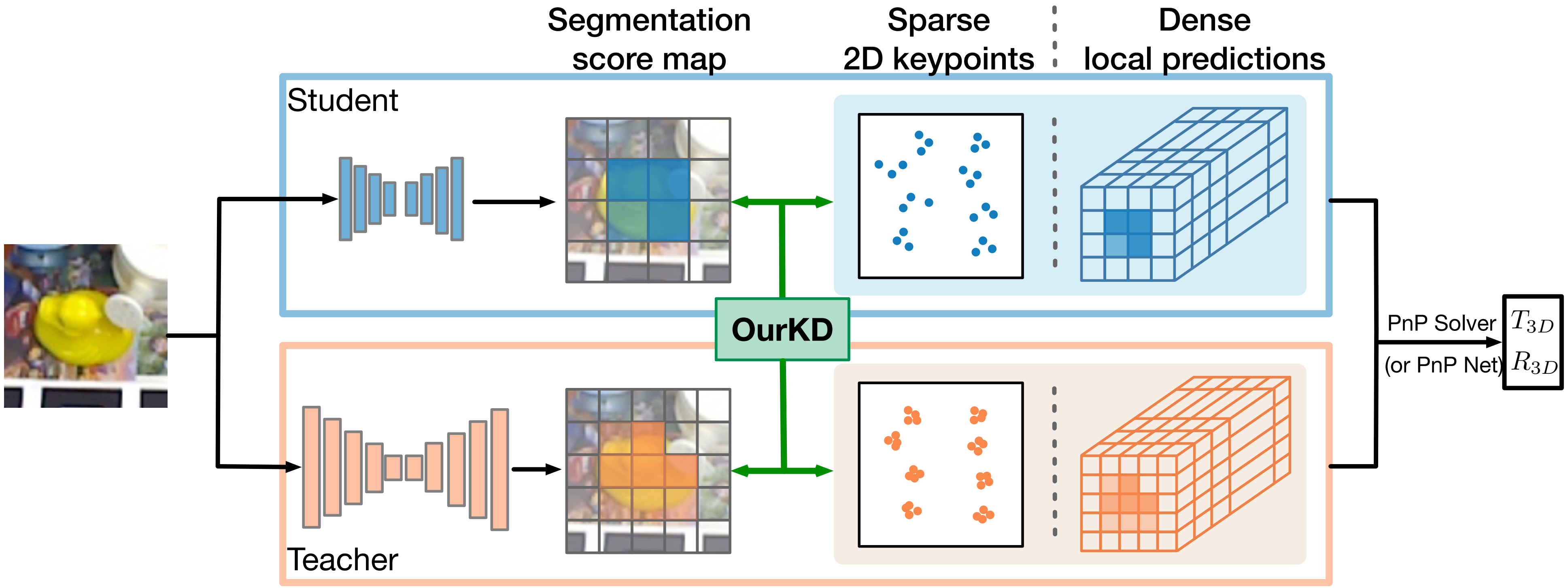
[1] Hu et al. "Segmentation-driven 6d object pose estimation." CVPR2019.
[2] Wang et al. "GDR-Net: Geometry-Guided Direct Regression Network for Monocular 6D Object Pose Estimation." CVPR2021.
[3] Su et al. "ZebraPose: Coarse to Fine Surface Encoding for 6DoF Object Pose Estimation." CVPR2022.

Motivation



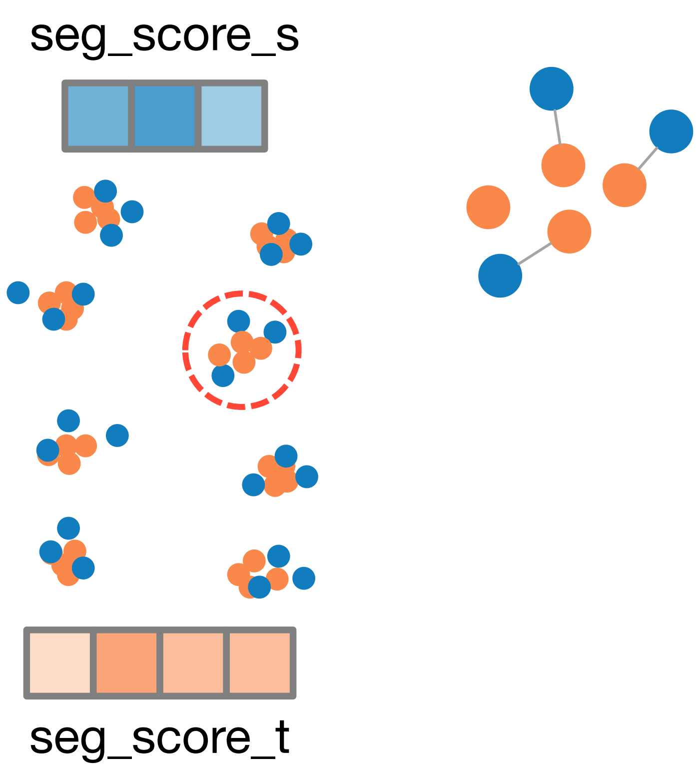
- Compact student network struggles predicting ***precise*** 2D keypoint locations as the teacher can do.
- ***Local predictions***, such as sparse 2D keypoints or dense predictions, are important to 6D pose estimation.

Method: Aligning Distributions of Local Predictions



Our KD is based on ***optimal transport*** that jointly distills the teacher's ***local prediction distribution*** + ***segmentation score map*** into the student.

Method: Aligning Distributions of Local Predictions



- KD on regressed location
  - All predictions are equal
  - Optimal transport

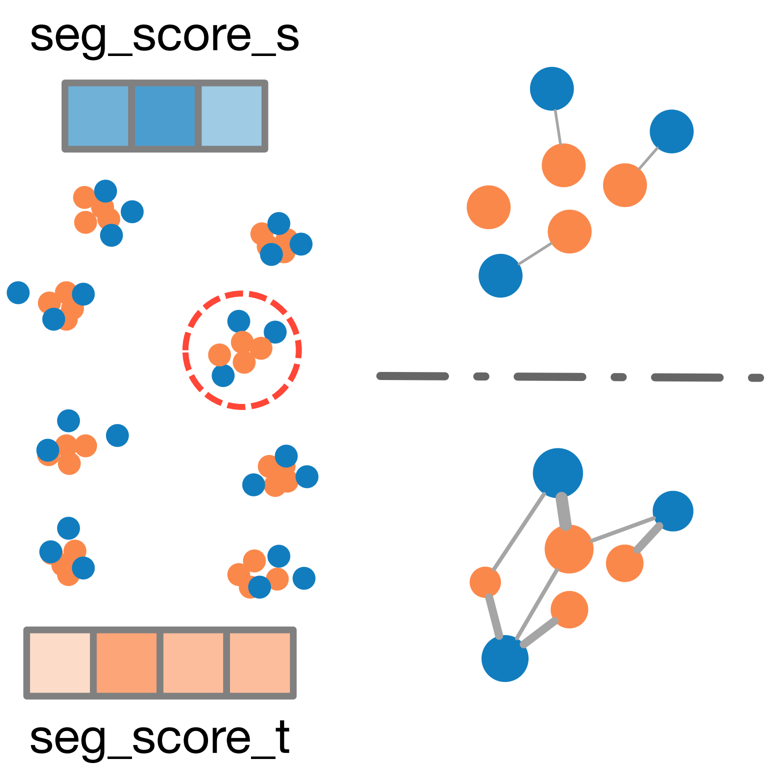
P_i^s, P_i^t : the student's / teacher's local predictions

N^s, N^t : the number of student / teacher local predictions

$$\bar{\mathcal{L}}_{kd}(P^s, P^t; \pi) = \min_{\pi} \sum_{i=1}^{N^s} \sum_{j=1}^{N^t} \pi_{ij} \|P_i^s - P_j^t\|_p$$

$$\text{s.t.} \quad \forall i, \ \sum_{j=1}^{N^t} \pi_{ij} = \frac{1}{N^s}, \quad \forall j, \ \sum_{i=1}^{N^s} \pi_{ij} = \frac{1}{N^t}$$

Method: Aligning Distributions of Local Predictions



- KD on regressed location
  - All predictions are equal
  - Optimal transport

    P_i^s, P_i^t : the student's / teacher's local predictions

    N^s, N^t : the number of student / teacher local predictions

- KD on regressed location + segmentation scores
  - Predictions are ***NOT*** equal
  - Weighted optimal transport
  - Softer alignment

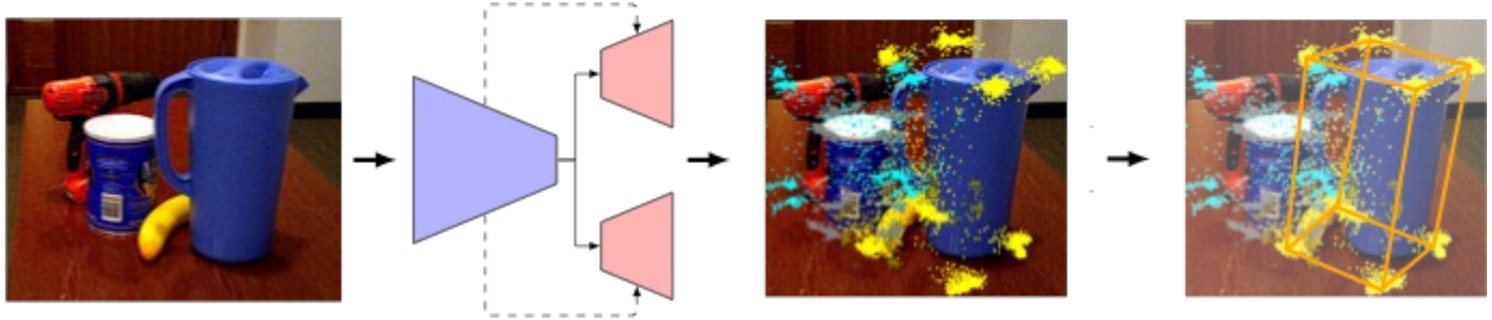
    α_i^s, α_j^t : segmentation score for cell i / j in student / teacher networks

$$\tilde{\mathcal{L}}_{kd}(P^s, P^t; \alpha^s, \alpha^t; \pi) = \min_{\pi} \sum_{i=1}^{N^s} \sum_{j=1}^{N^t} \pi_{ij} \| P_i^s - P_j^t \|_2$$

$$\text{s.t.} \quad \forall i, \ \sum_{j=1}^{N^t} \pi_{ij} = \alpha_i^s \ , \quad \forall j, \ \sum_{i=1}^{N^s} \pi_{ij} = \alpha_j^t$$

Method: Keypoint Distribution Alignment

- WDRNet+ -- SOTA sparse keypoint-based approach
  - Predict the 2D locations of the 8 object bounding box corners
  - WDRNet* + Cropped ROI



- Loss: Separate losses for the 8 individual keypoints clusters

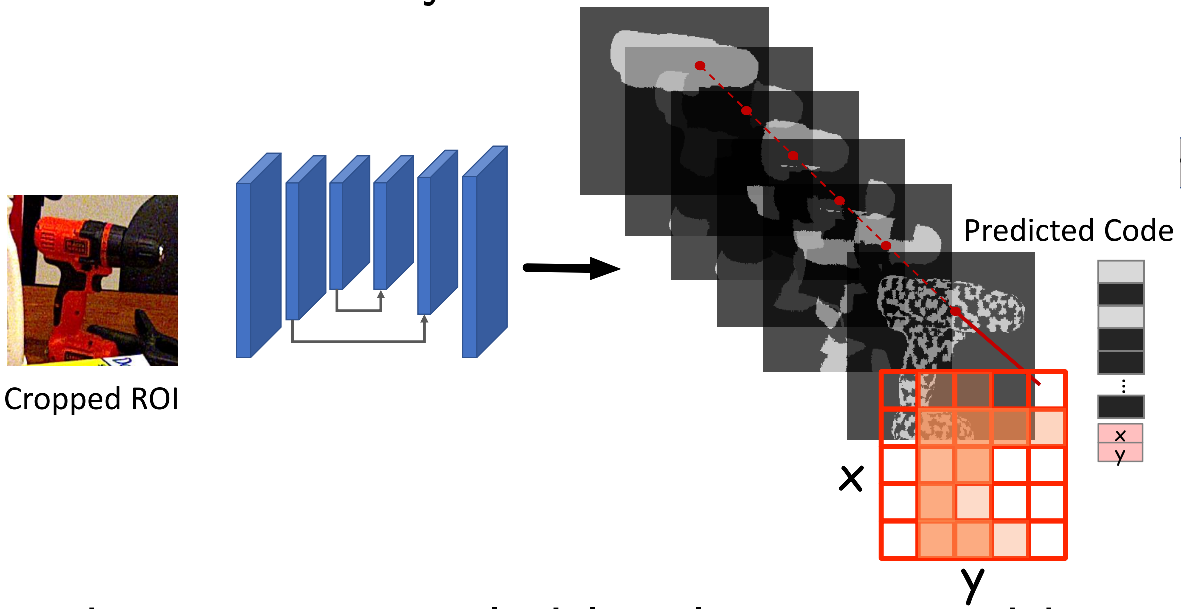
$$\mathcal{L}_{kd}^{kp}(\{C_k^s\},\{C_k^t\};\alpha^s,\alpha^t;\{\pi^k\}) = \sum_{k=1}^{8} \mathcal{L}_{kd}(C_k^s,C_k^t;\alpha^s,\alpha^t;\pi^k).$$

C_k^s, C_k^t: predictions for the kth 2D keypoint location; α^s, α^t: segmentation scores

* WDRNet: Hu et al. "Wide-Depth-Range 6D Object Pose Estimation in Space." CVPR2021.

Method: Dense Binary Code Distribution Alignment

- ZebraPose* -- SOTA dense local prediction-based approach
  - Predict a 16D binary code probability vector at each cell
  - Concatenate the x- and y-coordinate in the feature map



  - Loss: over the average-pooled local augmented binary code probabilities

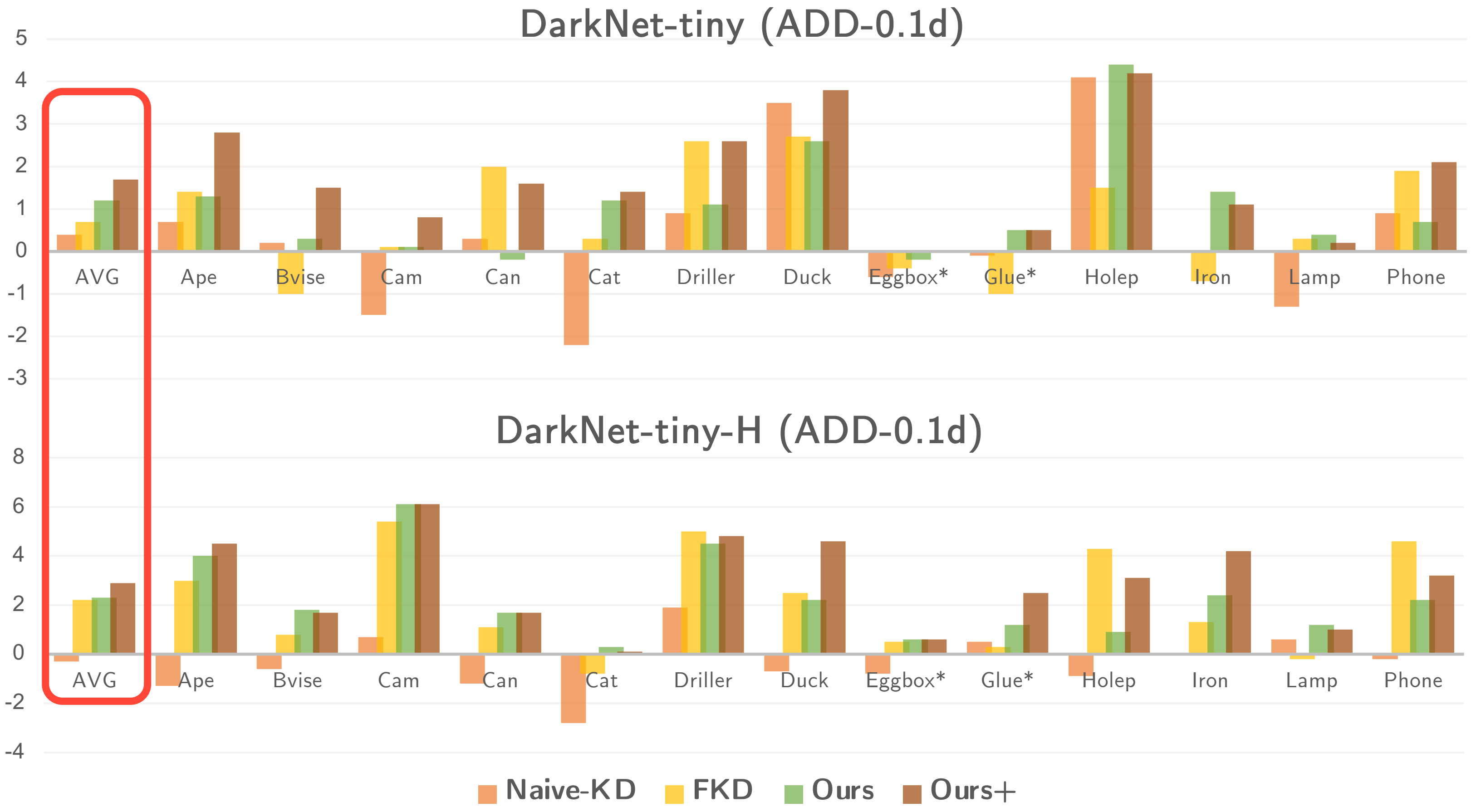
$$\mathcal{L}_{kd}^{bc}(B^s, B^t; \alpha^s, \alpha^t; \pi) = \mathcal{L}_{kd}(B^s, B^t; \alpha^s, \alpha^t; \pi).$$

$B^s, B^t;\ \alpha^s, \alpha^t$: average-pooled local predictions and segmentation scores

* ZebraPose: Su et al. "ZebraPose: Coarse to Fine Surface Encoding for 6DoF Object Pose Estimation." CVPR2022.

Experiments & Results

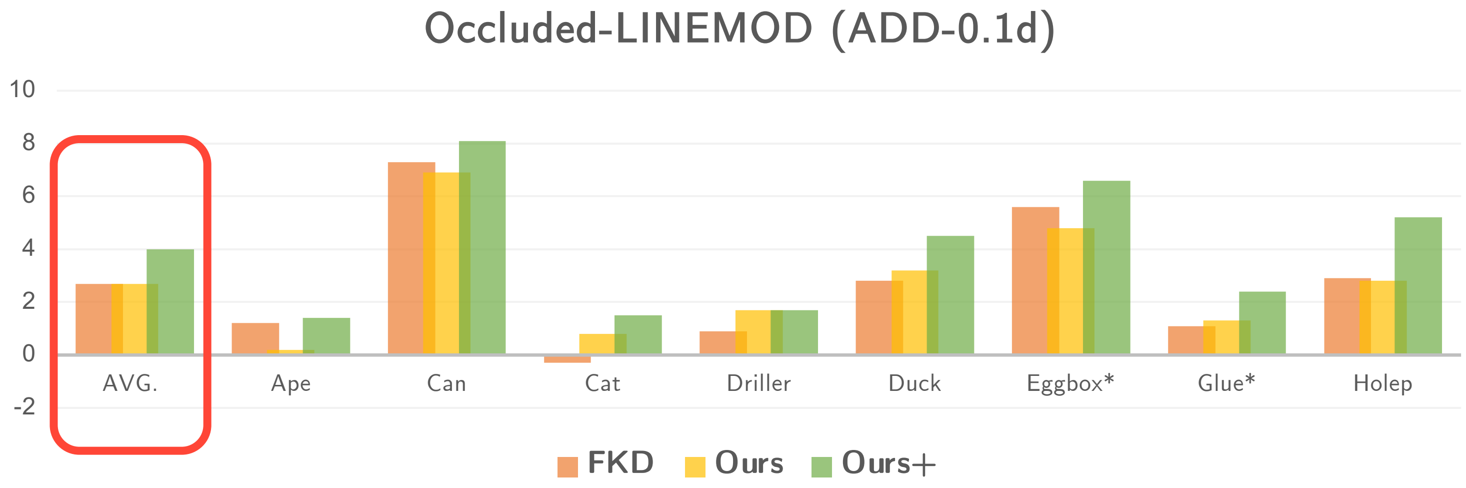
- WDRNet+ -- SOTA sparse keypoint-based approach
  - LINEMOD

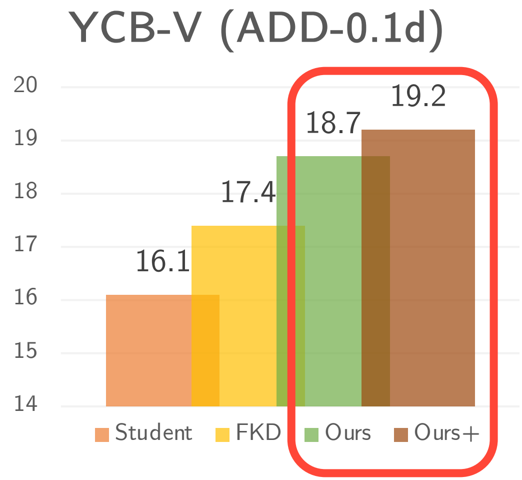


* FKD: Zhang et al. "Improve object detection with feature-based knowledge distillation: Towards accurate and efficient detectors." ICLR2021.

Experiments & Results

- WDRNet+ -- SOTA sparse keypoint-based approach
  - Occluded-LINEMOD
  - YCB-V

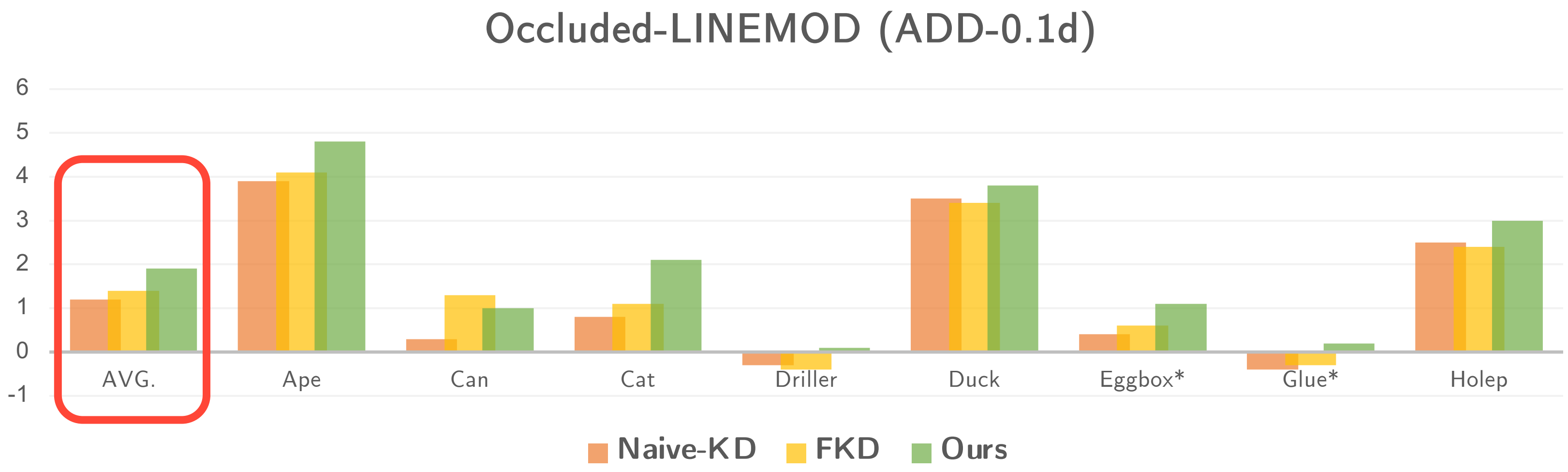




* FKD: Zhang et al. "Improve object detection with feature-based knowledge distillation: Towards accurate and efficient detectors." ICLR2021.

Experiments & Results

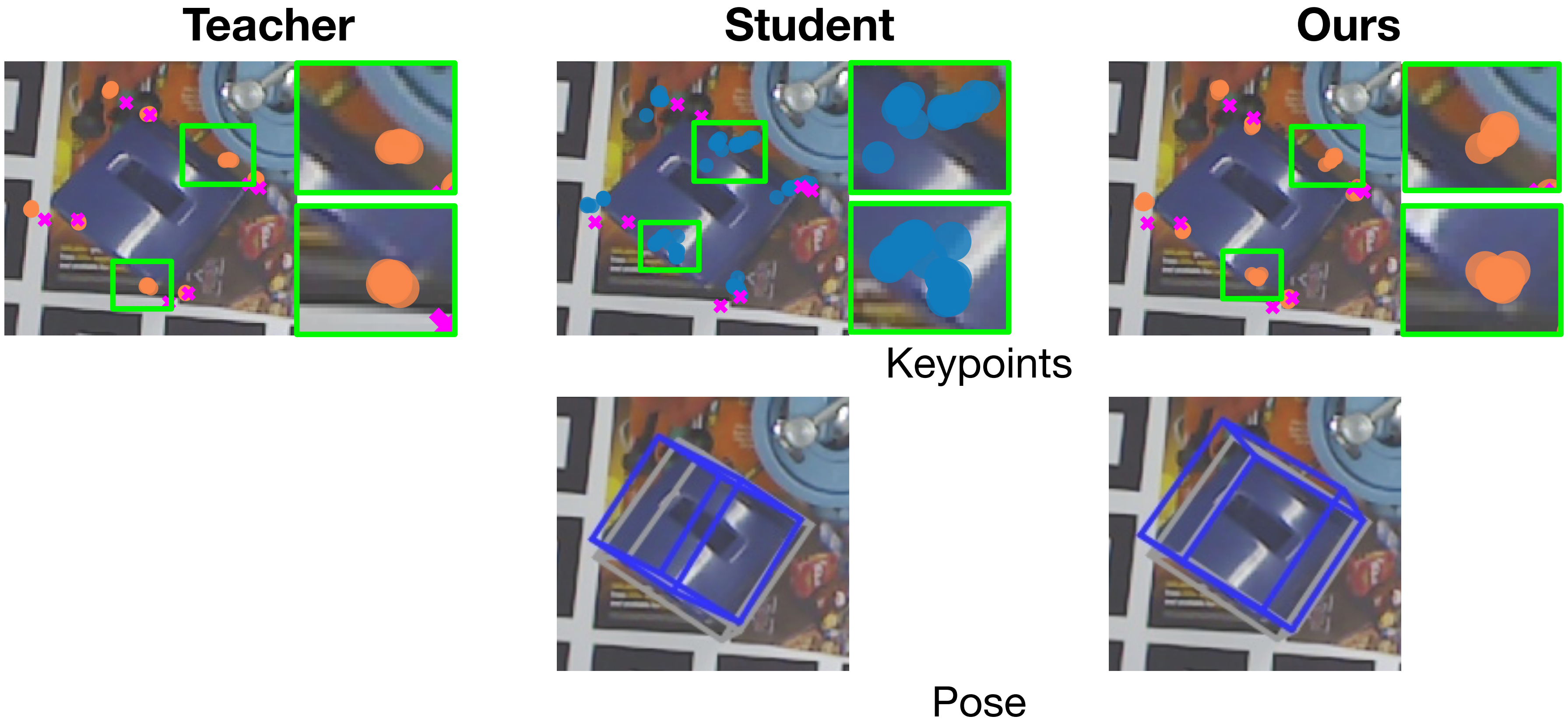
- ZebraPose -- SOTA dense local prediction-based approach
    - Occluded-LINEMOD



* FKD: Zhang et al. "Improve object detection with feature-based knowledge distillation: Towards accurate and efficient detectors." ICLR2021.

Discussion

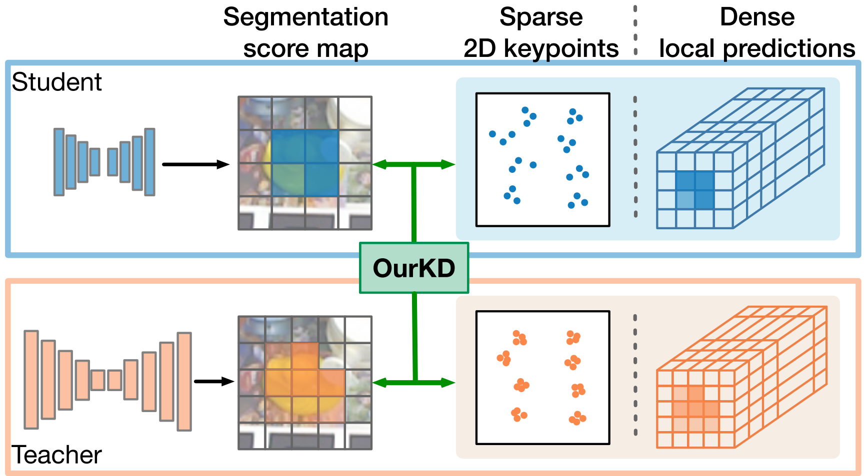
- Qualitative Analysis



- ***Mimic*** the teacher's keypoints distributions & Predict ***tighter*** keypoints clusters
- ***More accurate*** 6D pose estimation

Summary

- The first knowledge distillation in the context of 6D pose estimation.
- Our KD is driven by the 6D pose estimation task
  - Align the teacher and student local distributions together with their segmentation scores.
- Our KD generalizes to both sparse keypoints and dense predictions 6D pose estimation frameworks.
- Our KD can be used in conjunction with feature distillation to further boost the student's performance.



Code is available @
https://github.com/GUOShuxuan/kd-6d-pose-adlp
Please check our paper for more details.