UNIVERSITY OF BIRMINGHAM
Southern University of Science and Technology
THE UNIVERSITY OF HONG KONG

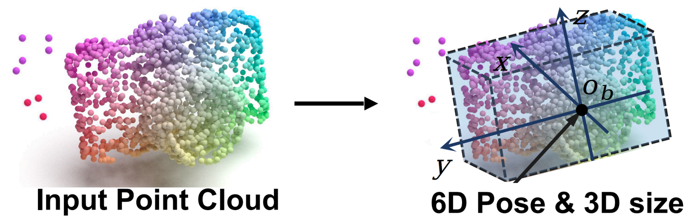
GeoReF: Geometric Alignment Across Shape Variation for Category-level Object Pose Refinement

Linfang Zheng[1,3], Tze Ho Elden Tse*[3], Chen Wang*[1,2], Yinghan Sun[1], Hua Chen[1], Aleš Leonardis[3], Wei Zhang[1], Hyung Jin Chang[3]

[1]Southern University of Science and Technology [2]University of Hongkong [3]University of Birmingham

CVPR JUNE 17-21, 2024 SEATTLE, WA

Problem description



1. Category-level object pose refinement
   Estimate the residual position, rotation and size of the objects within a category based on an initial pose estimation.
2. Application: Augmented reality and autonomous driving etc.

Challenges in category-level object pose refinement

Category-level pose refinement faces challenges due to diverse shape variations within each category. Existing methods, like CATRE, have shown effectiveness but are limited in capturing fine-grained geometric relationships. How to effectively capture the geometric relationships between objects with different shapes is still an open question.

Shape variations of objects within a category

Key contributions

1. **Innovative Architecture for Shape Variations**: Introducing a novel architecture tailored to address shape variation challenges in category-level object pose refinement, resulting in consistent performance gains and superior generalization.
2. **Efficient Cross-Cloud Transformation**: Proposing a unique mechanism for cross-cloud transformation that efficiently integrates information from observed point clouds and shape priors, enhancing pose estimation robustness.
3. **Experimental Validation**: Conducting extensive experiments on two datasets to validate the proposed method. Achieving a significant improvement over SPD on the REAL275 dataset, with a 39.1% increase in the 5°5cm metric, and surpassing CATRE with a 10.5% improvement in the 10°2cm metric.

Our Solution

Towards geometric alignment across shape variation

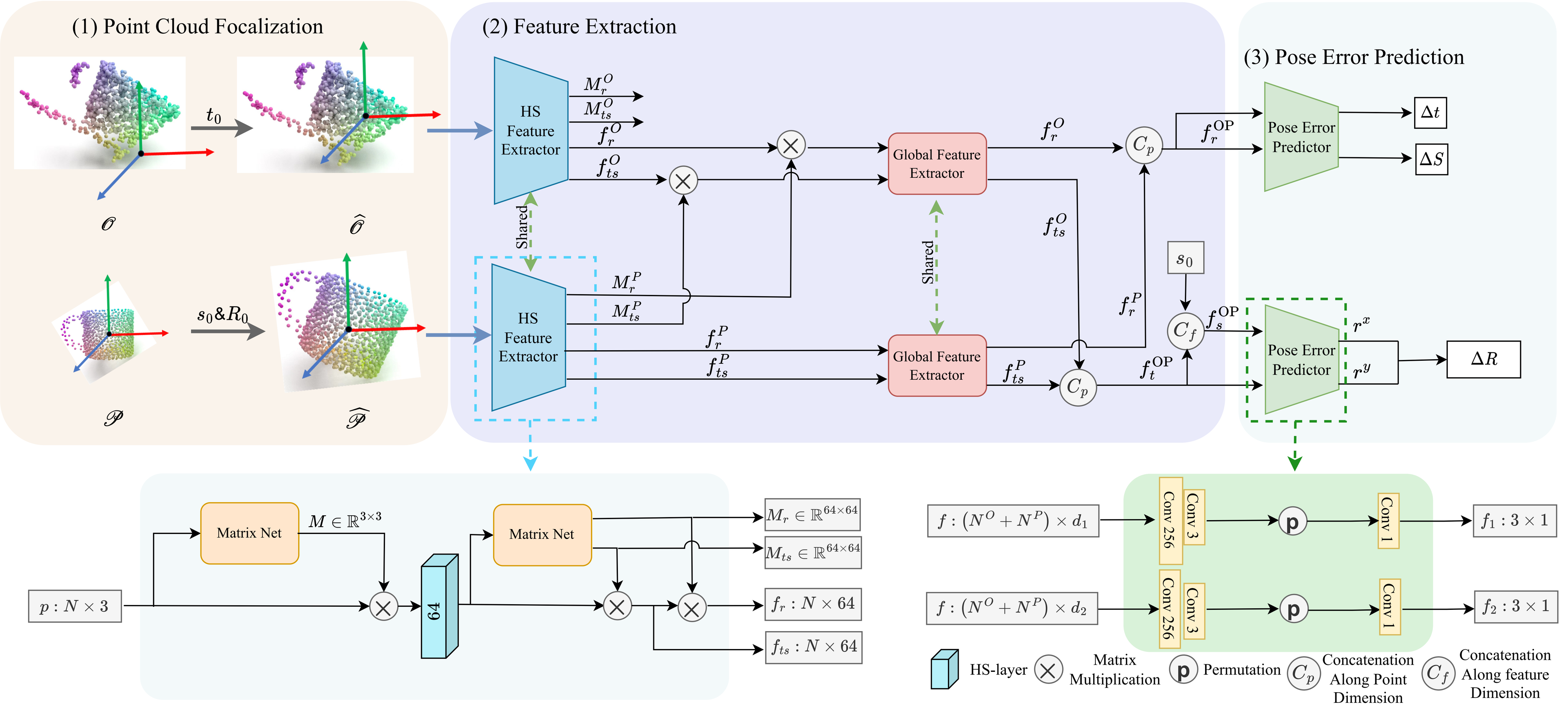
- Graph Convolution with Learnable Affine Transformation (LAT): We incorporates an hybrid-scope graph convolution layer (HS-layer) with learnable affine transformation for improved local and global geometric information extraction and addressing geometric discrepancies.
- Cross-Cloud Transformation (CCT): We propose a cross-cloud transformation mechanism to efficiently merge information from observed point clouds and shape priors.
- Integrating Shape Prior in Pose Estimation: We push the limit of our method by incorporating the shape prior information into the scale and translation branch.

Contact

- Linfang Zheng (lxz948@student.bham.ac.uk)
- Control and Learning for Robotics and Autonomy Lab (https://www.wzhanglab.site)
- Human-Centred Visual Learning Group at University of Birmingham (https://www.cs.bham.ac.uk/~changhj)

GeoReF overall framework

Input: a point cloud, a shape prior, and an initial pose **Output:** 6D pose and 3D size of the target object



Propose method overview: Our method comprises three main modules. It starts with point cloud focalization using the initial estimation, followed by geometric-based feature extraction. The extracted features are then used for rotation, translation, and size error estimation. The HS Feature Extractor includes learnable affine transformations (LATs) for adaptive point and feature adjustment

Ablation study

✓ Effectiveness of each proposed components

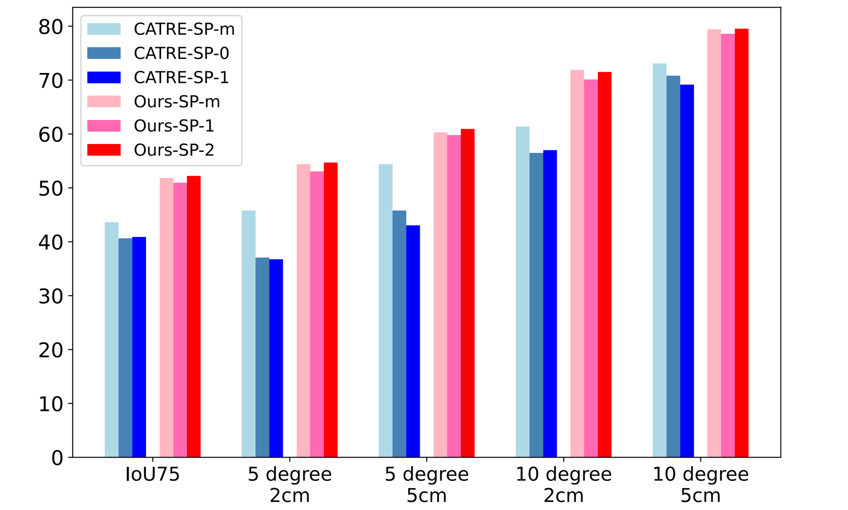
i) Cross-Cloud Transformation
ii) Prior in pose estimation branch
iii) Learnable affine transformation

Row	Method	IoU_{50}	IoU_{75}	5°2cm	5°5cm	10°2cm	10°5cm
A0	CATRE[41] (baseline)	77.0	43.6	45.8	54.4	61.4	73.1
B0	**Ours**: E0 + Cross-cloud Transformation	**79.2** 2.2↑	**51.8** 8.2↑	**54.4** 8.6↑	**60.3** 5.9↑	**71.9** 10.5↑	**79.4** 6.3↑
D0	A0 + prior in ST branch	77.1	45.8	48.0	54.6	63.8	72.5
E0	D0: PointNet → HS-layer+LAT	79.4	51.0	52.4	58.6	69.4	77.7

✓ Generalisability test on CAMERA25

Row	Method	Train Data Size	IoU_{75}	5°2cm	5°5cm	10°2cm	10°5cm
A0	CATRE	275K	76.1	<u>75.4</u>	80.3	83.3	89.3
B0	CATRE	5K	63.2	66.4	72.3	79.4	87.4
B1	**Ours**	5K	<u>77.5</u>	<u>75.4</u>	<u>81.1</u>	<u>83.4</u>	<u>90.0</u>
C0	CATRE	10K	66.5	69.7	75.5	81.8	89.1
C1	**Ours**	10K	**79.2**	**77.9**	**84.0**	**83.8**	**90.5**

✓ Robustness to different shape priors



Quantitative results

Table 3. **Comparison with other methods on CAMERA25.**

Method	IoU_{75}	5°2cm	5°5cm	10°2cm	10°5cm
NOCS [9]	37.0	32.3	40.9	48.2	64.6
DualPoseNet [4]	71.7	64.7	70.7	77.2	84.7
CR-Net [10]	75.0	72.0	76.4	81.0	87.7
SGPA [1]	69.1	70.7	74.5	82.7	88.4
SAR-Net [3]	62.6	66.7	70.9	75.3	80.3
SSP-Pose [12]	-	64.7	75.5	-	87.4
RBP-Pose [11]	-	73.5	79.6	82.1	89.5
GPV-Pose [2]	-	72.1	79.1	-	89.0
HS-Pose [13]	-	73.3	80.5	80.4	89.4
SPD* [8]	46.9	54.1	58.8	73.9	82.1
SPD*+CATRE [6]	76.1	75.4	80.3	83.3	89.3
SPD*+**Ours**	**79.2**	**77.9**	**84.0**	**83.8**	**90.5**

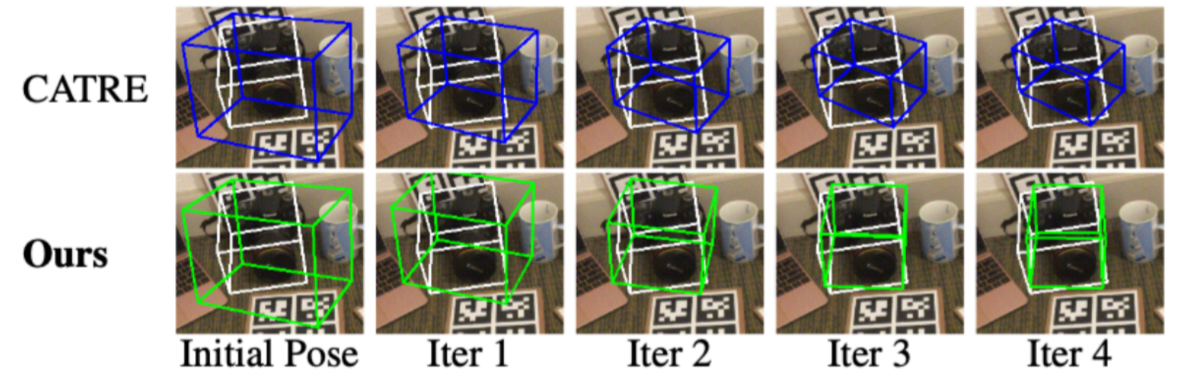
Table 4. **Comparison with other methods on REAL275.**

Method	IoU_{75}	5°2cm	5°5cm	10°2cm	10°5cm
NOCS [45]	9.4	7.2	10.0	13.8	25.2
DualPoseNet [20]	30.8	29.3	35.9	50.0	66.8
CR-Net [46]	33.2	27.8	34.3	47.2	60.8
SGPA [4]	37.1	35.9	39.6	61.3	70.7
RBP-Pose [53]	24.5	38.2	48.1	63.1	79.2
GPV-Pose [7]	23.1	32.0	42.9	55.0	73.3
HS-Pose [55]	39.1	<u>46.5</u>	<u>55.2</u>	<u>68.6</u>	**82.7**
SPD* [41]	27.0	19.1	21.2	43.5	54.0
SPD*+CATRE [25]	<u>43.6</u>	45.8	54.4	61.4	73.1
SPD*+**Ours**	**51.8**	**54.4**	**60.3**	**71.9**	<u>79.4</u>

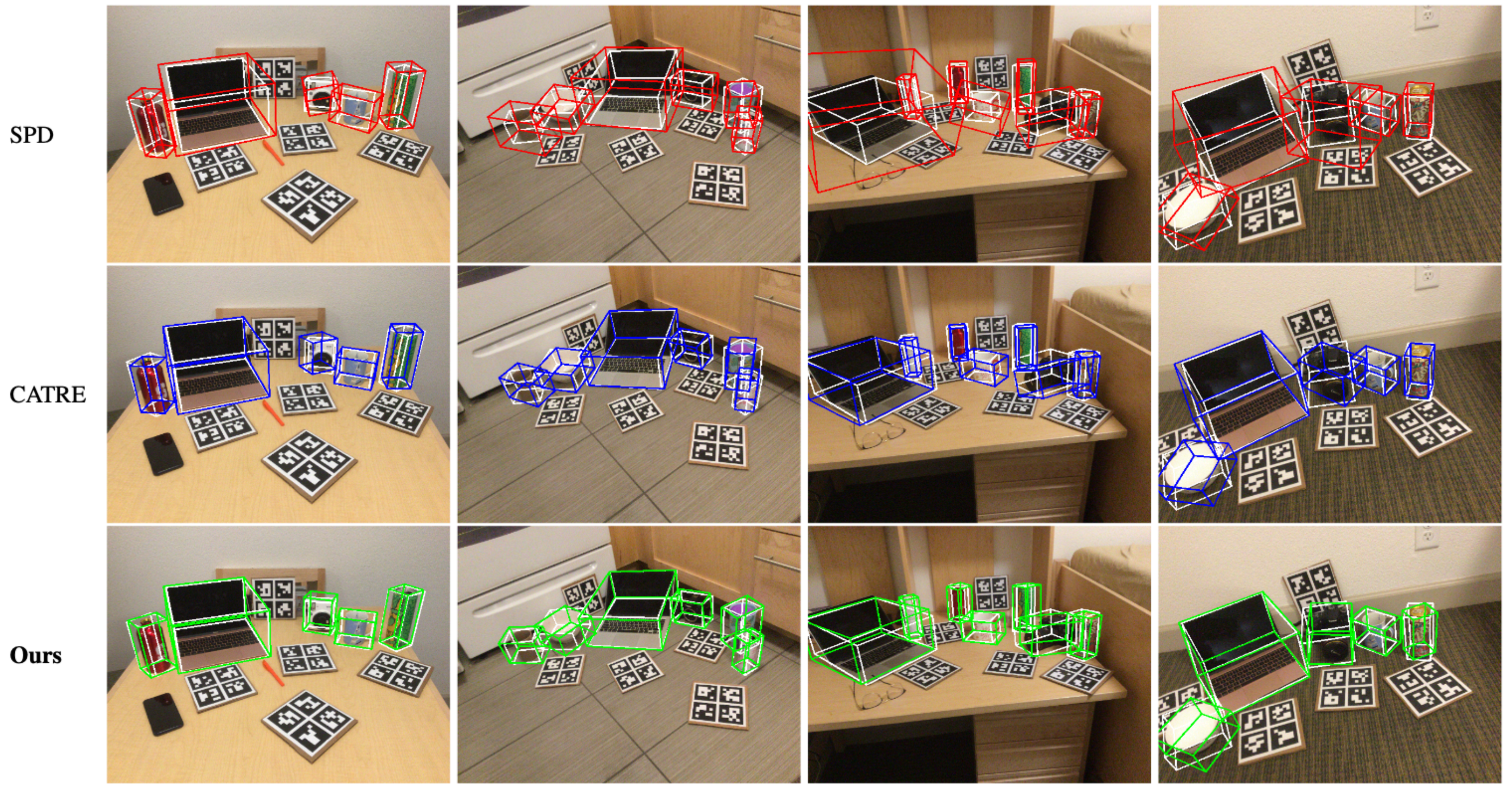
Qualitative examples

➢ *Refinement Iterations*

Comparison of proposed method (row \#2) and CATRE (row \#1) during a complete refinement iteration, both using SPD as initial estimation.



➢ *Comparison on REAL275*



Qualitative comparison of proposed (row \#3) and baseline [] (row \#2) methods using SPD (row \#1) as initial estimation. Ground truth shown with white lines. Note that the estimated rotations of symmetric objects (\eg bowl, bottle, and can) are considered correct if the symmetry axis is aligned.

Key references:

[6] Liu et al. ECCV, 2022: CATRE: Iterative point clouds alignment for category-level object pose refinement.
[8] Tian et al. ECCV, 2020: Shape prior deformation for categorical 6d object pose and size estimation.