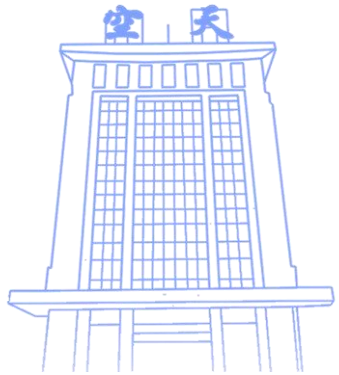
ICCV OCT 19-23, 2025 HONOLULU HAWAII

ICCV2025

Deterministic Object Pose Confidence Region Estimation

Jinghao Wang[1,2] Zhang Li[1,2] Zi Wang[1,2,*]
Banglei Guan[1,2]Yang Shang[1,2] Qifeng Yu[1,2]

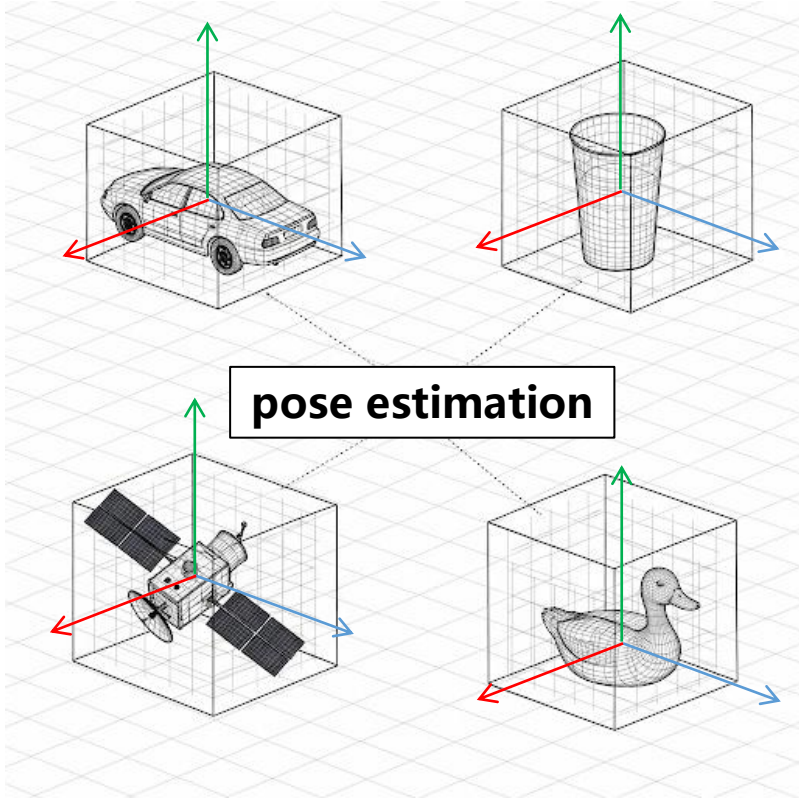
1 National University of Defense Technology
2 Hunan Provincial Key Laboratory of Image Measurement and Vision Navigation



October 22, 2025

1 Object Pose Estimation

- **Instance-Level Object Pose Estimation**
  - **Template-based methods**
    - RGB-based template
    - Point cloud-based template
  - **Voting-based methods**
    - Indirect voting
    - Direct voting
  - **Regression-based methods**
    - Geometry-guided regression
    - Direct regression
  - **Correspondence-based methods**
    - Sparse correspondence
    - Dense correspondence



Instance-Level Object Pose Estimation is the task of determining the precise 3D position (translation) and 3D orientation (rotation) of a specific, known object instance within a given image or 3D scene.

[1] Liu, Jian, et al. "Deep learning-based object pose estimation: A comprehensive survey." arXiv preprint arXiv:2405.07801 (2024).

1.1 Template-based methods

Template-based methods

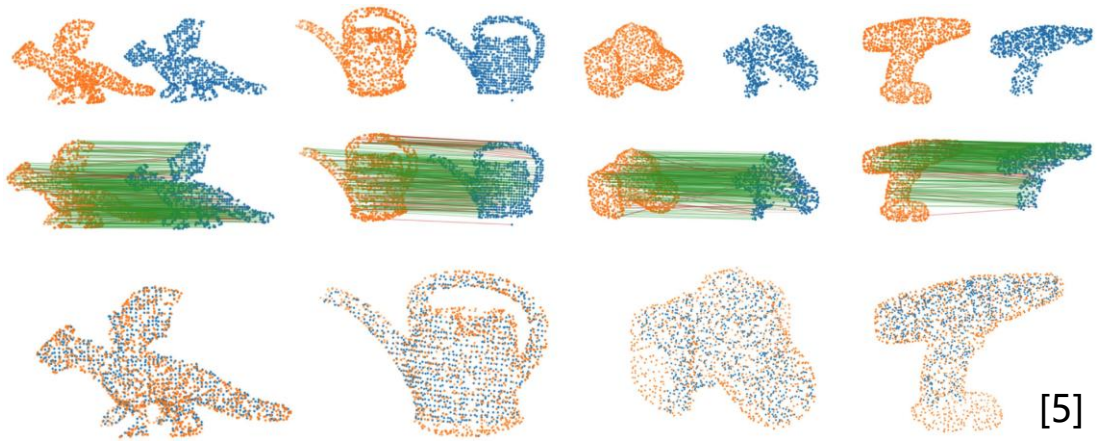
- **RGB-based template**
- **Point cloud-based template**

[2] X. Liu and J. Zhang, "6dof pose estimation with object cutout based on a deep autoencoder," in ISMAR-Adjunct, 2019.

[3] Y. Zhang and C. Zhang, "6d object pose estimation algorithm using preprocessing of segmentation and keypoint extraction," in I2MTC, 2020.

[4] H. Jiang and M. Salzmann, "Se(3) diffusion model-based point cloud registration for robust 6d object pose estimation," in NeurIPS, 2023.

[5] Z. Dang and L. Wang, "Match normalization: Learning-based point cloud registration for 6d object pose estimation in the real world," IEEE TPAMI, 2024.



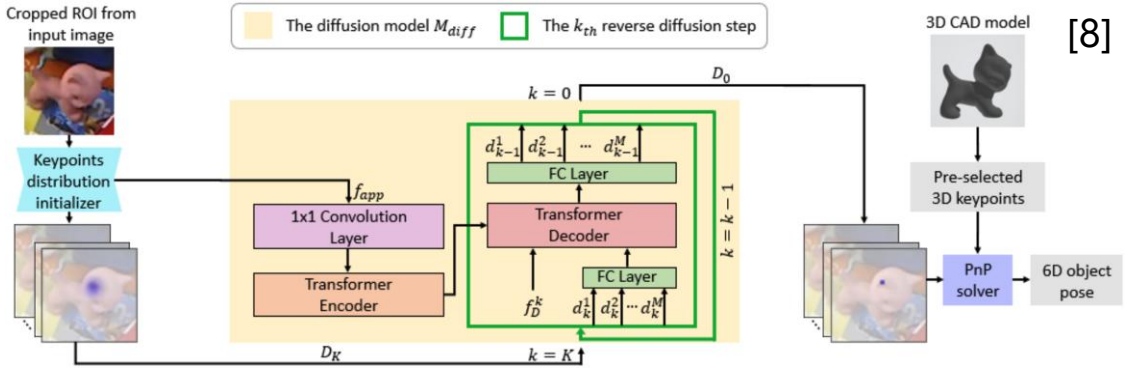
[5]

Template-based methods involve identifying the most similar template from a set of templates labeled with ground-truth object poses.

1.2 Voting-based methods

Voting-based methods

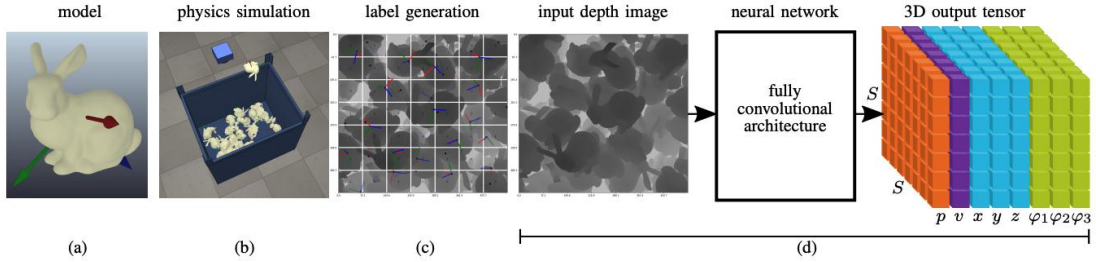
- **Indirect voting**
  - [6] X. Liu and S. Iwase, "Kdfnet: Learning keypoint distance field for 6d object pose estimation," in IROS, 2021.
  - [7] P. Liu and Q. Zhang, "Bdr6d: Bidirectional deep residual fusion network for 6d pose estimation," IEEE TASE, 2023.
  - [8] L. Xu and H. Qu, "6d-diff: A keypoint diffusion framework for 6d object pose estimation," in CVPR, 2024.
- **Direct voting**
  - [9] G. Zhou and Y. Yan, "A novel depth and color feature fusion framework for 6d object pose estimation," IEEE TMM, 2020.
  - [10] X. Liu and X. Yuan, "A depth adaptive feature extraction and dense prediction network for 6-d pose estimation in robotic grasping," IEEE TII, 2023.
  - [11] F. Mu and R. Huang, "Temporalfusion: Temporal motion reasoning with multi-frame fusion for 6d object pose estimation," in IROS, 2021.



[8]

Voting-based methods determine object pose through a pixel-level or point-level voting scheme.

1.3 Regression-based methods



[15]

Regression-based methods

- **Geometry-guided regression**
- **Direct regression**

[12] G. Gao and M. Lauri, "6d object pose regression via supervised learning on point clouds," in ICRA, 2020.

[13] M. Lin and V. Murali, "6d object pose estimation with pairwise compatible geometric features," in ICRA, 2021.

[14] Z. Liu and Q. Wang, "Pa-pose: Partial point cloud fusion based on reliable alignment for 6d pose tracking," PR, 2024.

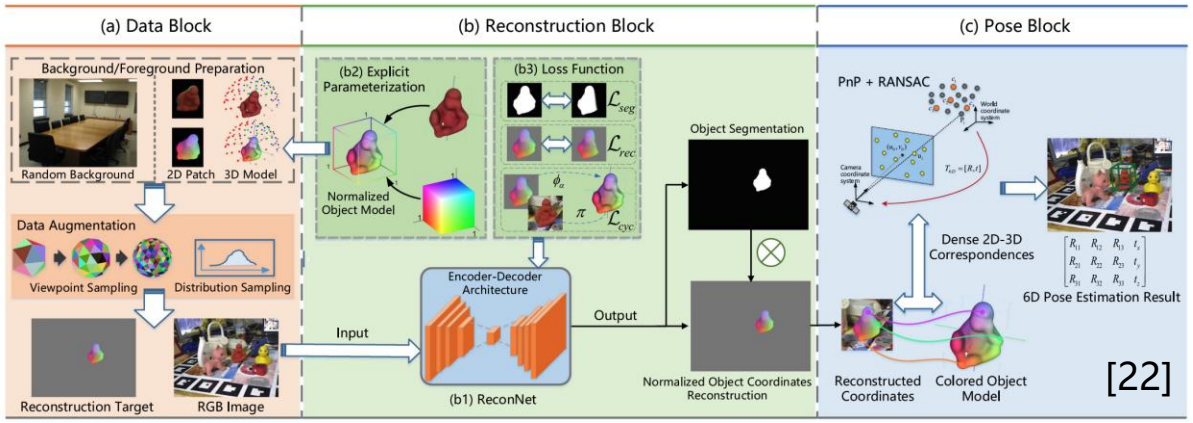
[15] K. Kleeberger and M. F. Huber, "Single shot 6d object pose estimation," in ICRA, 2020.

[16] V. Sarode and X. Li, "Pcrnet: Point cloud registration network using pointnet encoding," arXiv preprint arXiv:1908.07906, 2019.

[17] S. H. Bengtson and H. Astrom, "Pose estimation from rgb images of highly symmetric objects using a novel multi-pose loss and differential rendering," in IROS, 2021.

Regression-based methods aim to directly obtain the object pose from the learned features.

1.4 Correspondence-based methods



[18] M. Rad and V. Lepetit, "Bb8: A scalable, accurate, robust to partial occlusion method for predicting the 3d poses of challenging objects without using depth," in ICCV, 2017.

[19] B. Tekin and S. N. Sinha, "Real-time seamless single shot 6d object pose prediction," in CVPR, 2018.

[20] B. Doosti and S. Naha, "Hope-net: A graph-based model for hand-object pose estimation," in CVPR, 2020.

[21] H. Chen and P. Wang, "Epro-pnp: Generalized end-to-end probabilistic perspective-n-points for monocular object pose estimation," in CVPR, 2022.

[22] D. Wang and G. Zhou, "Geopose: Dense reconstruction guided 6d object pose estimation with geometric consistency," IEEE TMM, 2021.

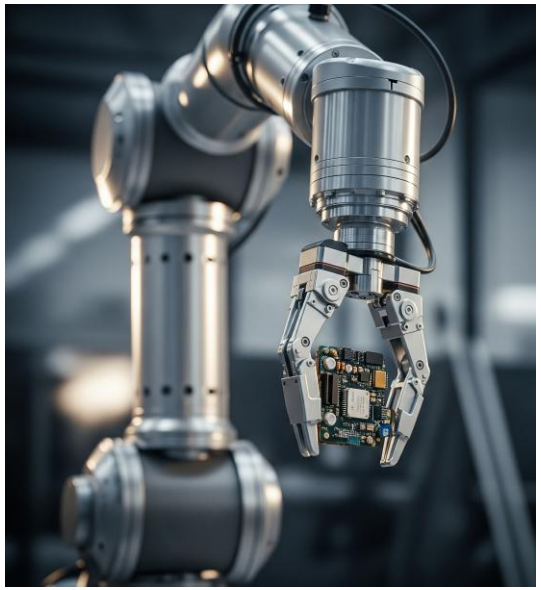
[23]L. Huang and T. Hodan, "Neural correspondence field for object pose estimation," in ECCV, 2022.

Correspondence-based object pose estimation refers to techniques that involve identifying correspondences between the input data and the object CAD model.

Correspondence -based methods

- **Sparse correspondence**
- **Dense correspondence**

2.1 Motivation



Will this grasp succeed?

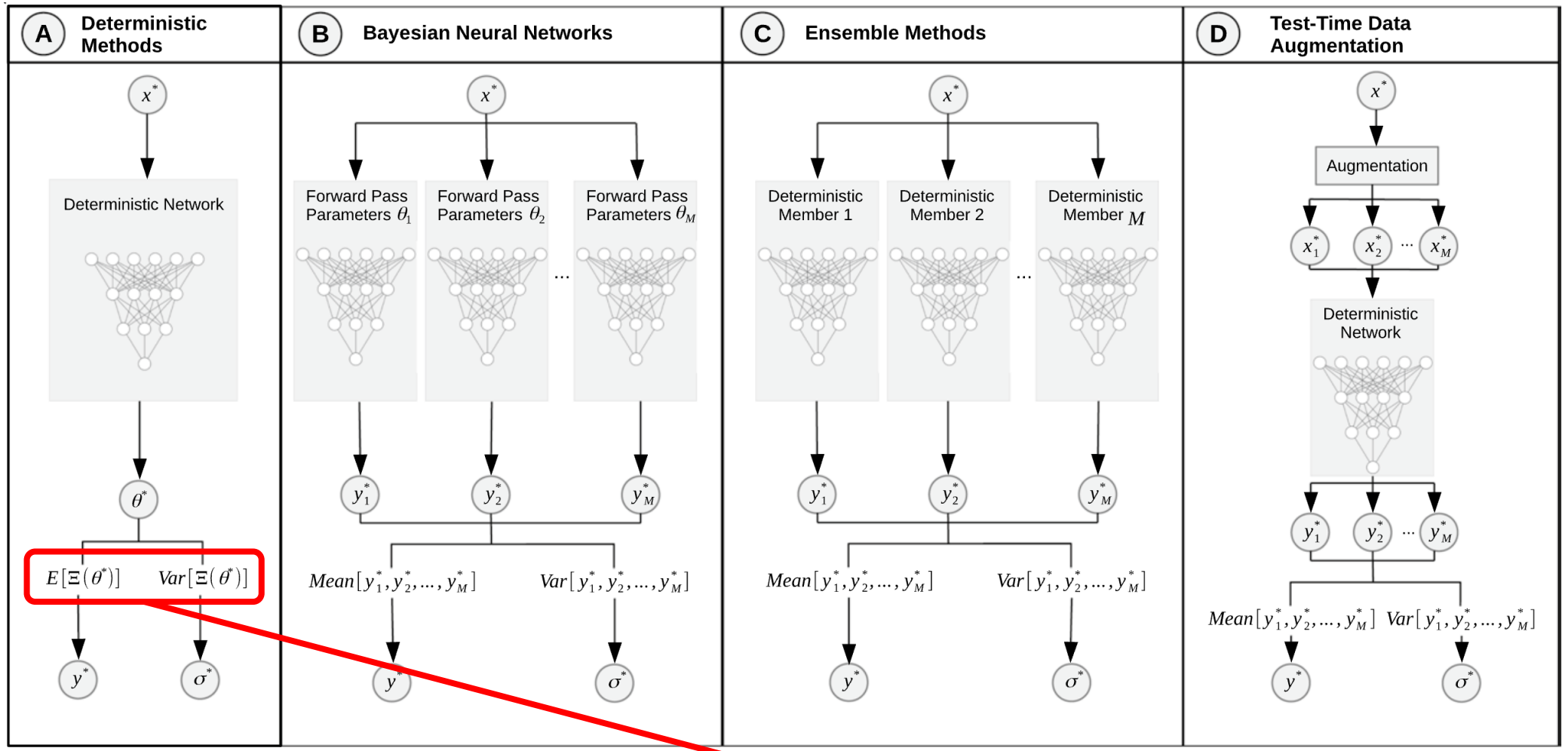
Is the car's perception reliable?

Will this maneuver be safe?

Answering this requires a statistically valid confidence region.

For safety-critical tasks, a single pose estimate is not enough.
We need to quantify the uncertainty.

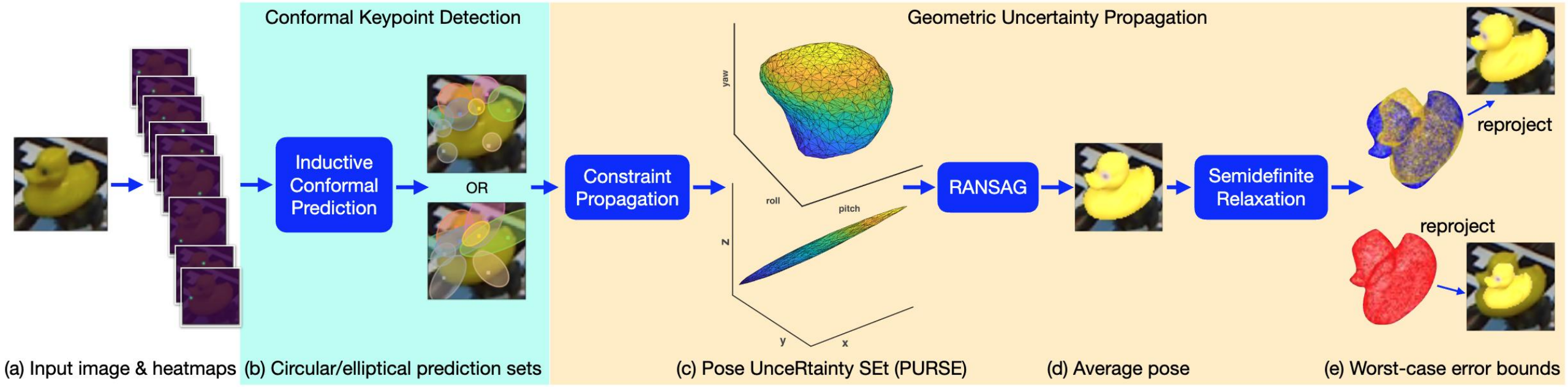
2.2 Uncertainty Quantification



a statistically guaranteed coverage probability (e.g., 90%)?

[24] Gawlikowski, Jakob, et al. "A survey of uncertainty in deep neural networks." Artificial Intelligence Review 56.Suppl 1 (2023): 1513-1589.

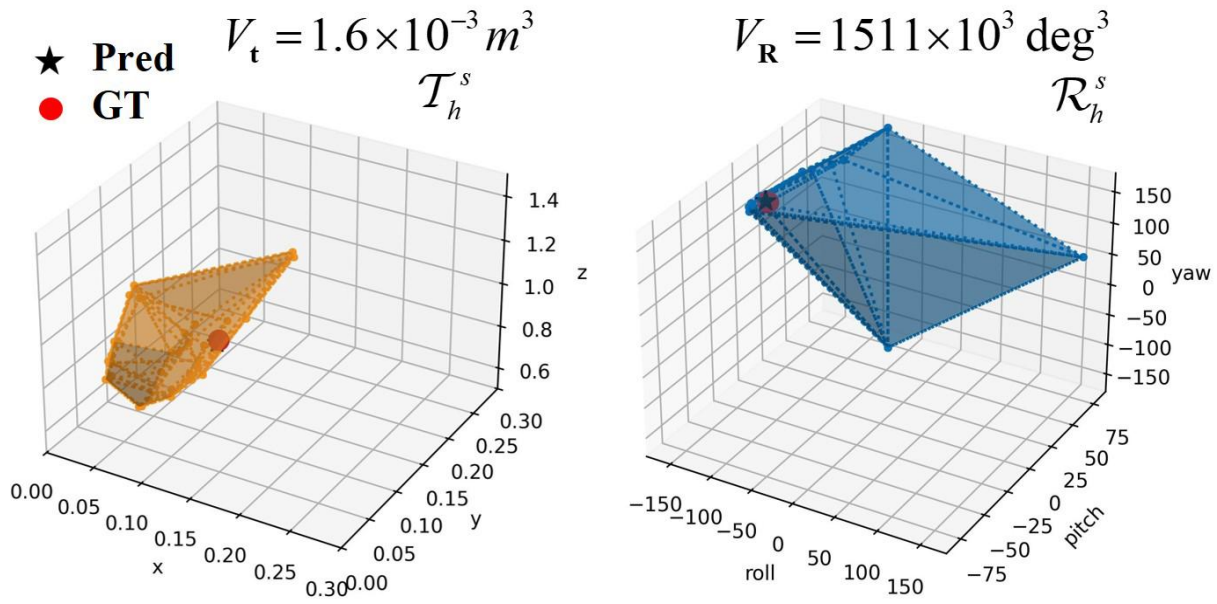
2.3 Pose Confidence Region Estimation

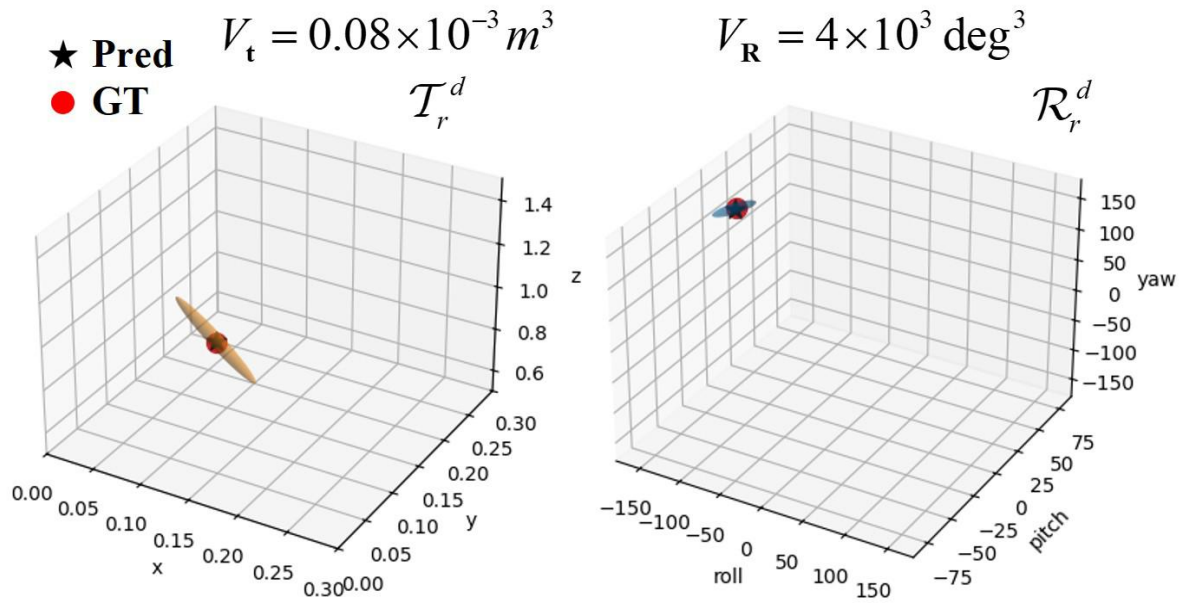


Inductive conformal prediction is a framework that wraps any machine learning model to produce prediction sets with a statistically guaranteed coverage probability (e.g., 90%).

[25] Yang, H., & Pavone, M. (2023). Object pose estimation with statistical guarantees: Conformal keypoint detection and geometric uncertainty propagation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8947-8958).

2.4 Narrower Confidence Region





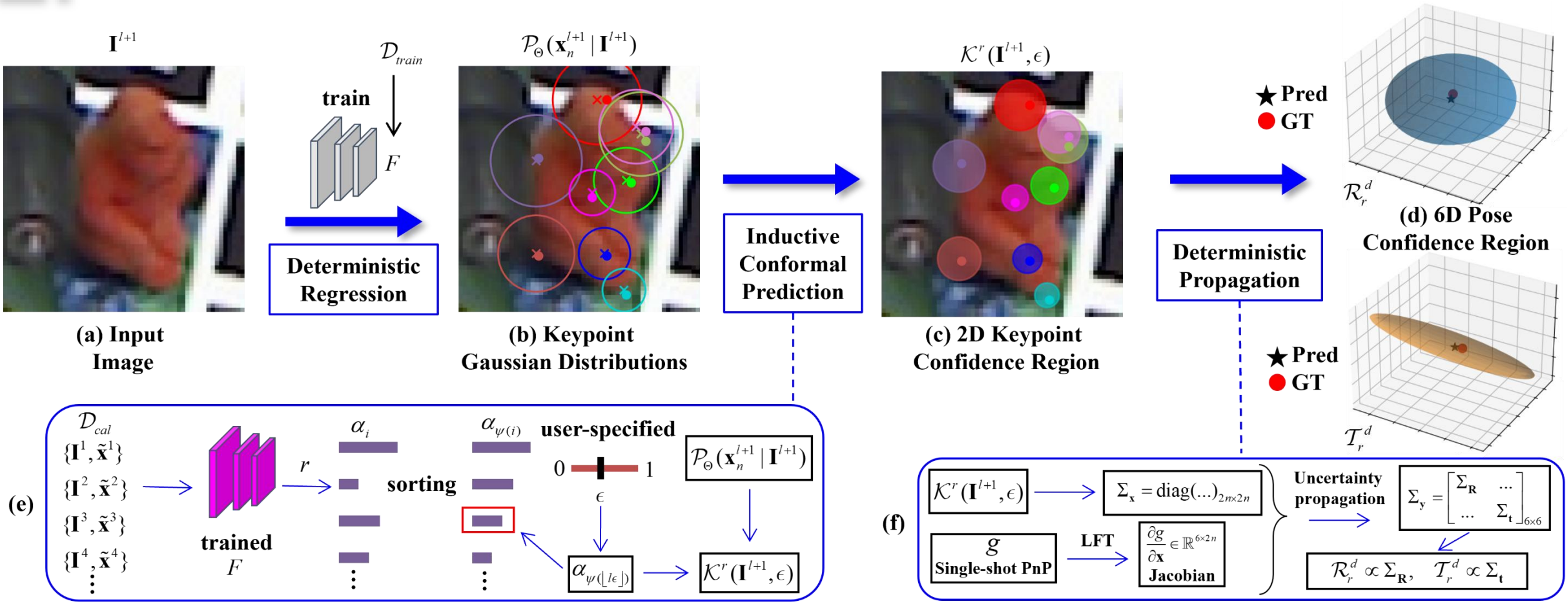
Sampling method [25]

Deterministic method

Narrower confidence regions of our deterministic approach compared to the sampling method

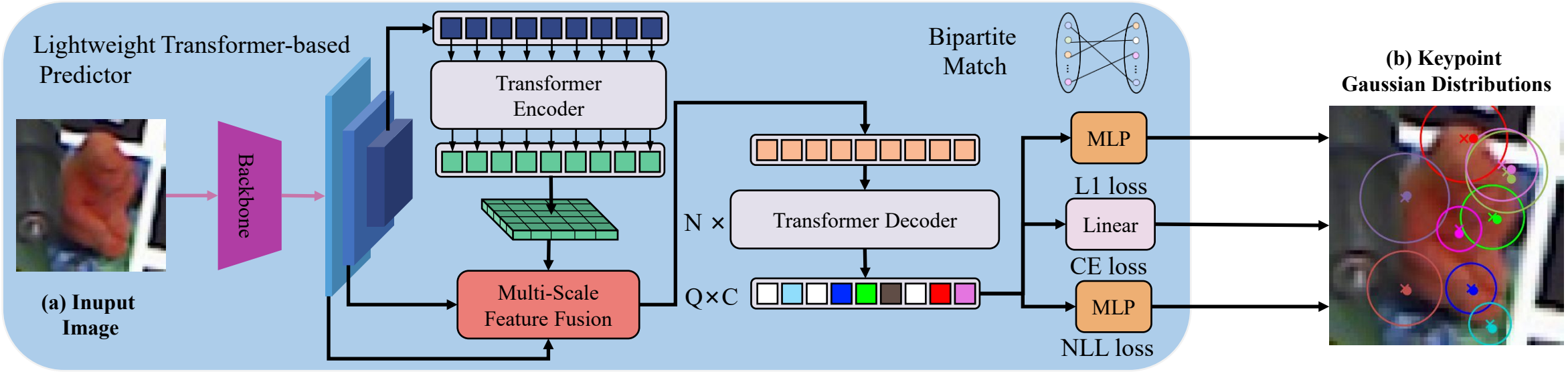
[25] Yang, H., & Pavone, M. (2023). Object pose estimation with statistical guarantees: Conformal keypoint detection and geometric uncertainty propagation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8947-8958).

3 Overview of the proposed method



Our method produces a guaranteed 6D pose confidence region by propagating uncertainty from 2D keypoint sets established via inductive conformal prediction.

3.1 Keypoint Detection



Keypoint detection and uncertainty quantification using a direct regression-based approach, replacing computationally intensive heatmaps.

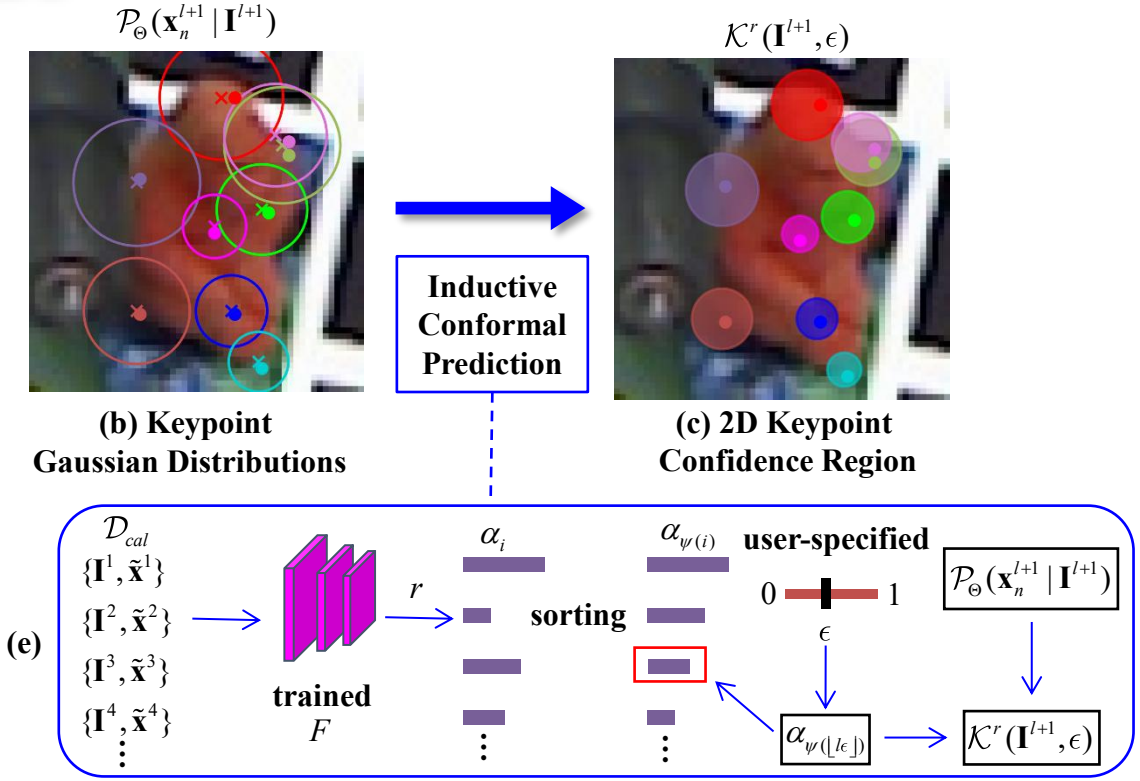
Training a model using negative log-likelihood (NLL) loss.

$$\mathcal{L}_{NLL} = -\sum_{n=1}^{N} \log \mathcal{P}_\Theta(\mathbf{x}_n|\mathbf{I})\Big|_{\mathbf{x}_n=\tilde{\mathbf{x}}_n}$$

A deep neural network maps an image to N Gaussian distributions.

$$\mathcal{P}_\Theta(\mathbf{x}_n|\mathbf{I}) \sim \mathcal{N}(\mathbf{x}_n, \boldsymbol{\sigma}_n^2)$$

3.2 Inductive Conformal Prediction



Non-conformity function for measuring the difference between a new sample and the calibration set.

$$\phi(\tilde{\mathbf{x}}_n, F(\mathbf{I})_n) = \frac{\|\tilde{\mathbf{x}}_n - \mathbf{x}_n\|}{\det(\boldsymbol{\sigma}_n)}$$

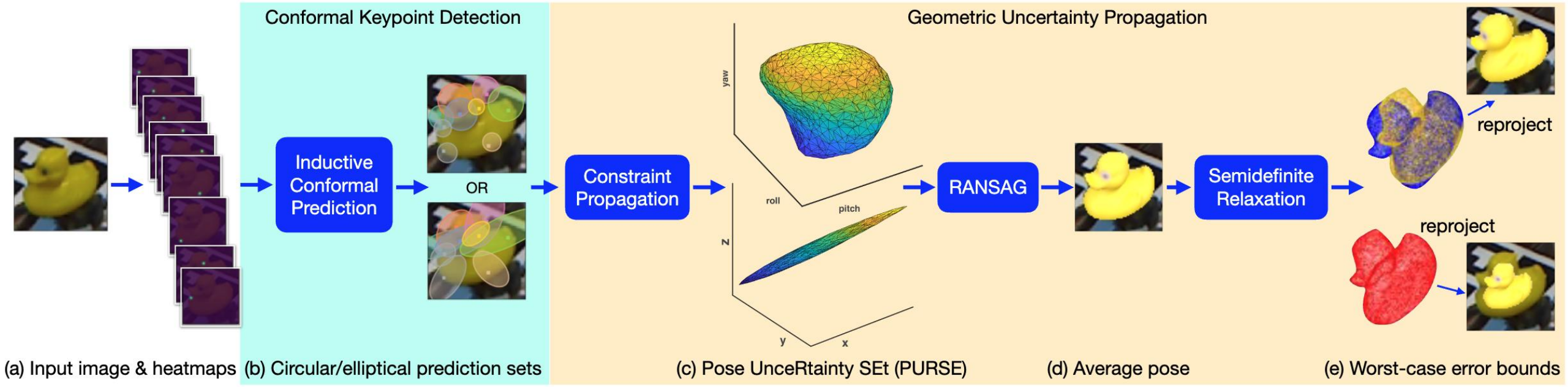
Conformal prediction of 2D confidence regions for a new sample.

$$\mathcal{K}^r = \{\mathbf{x}' \mid \|\mathbf{x}'_n - \mathbf{x}_n\| \leq \det(\sigma_n)\alpha, \forall n\}$$

ICP produces statistically guaranteed keypoint confidence regions for finite samples, ensuring the regions contain the true value at a user-specified coverage rate.

$$\mathbb{P}\left[\tilde{\mathbf{x}}^{l+1} \in \mathcal{K}^r(\mathbf{I}^{l+1}, \epsilon)\right] \geq 1 - \epsilon$$

3.3 Sampling-based Uncertainty Propagation



Sampling keypoints from 2D confidence regions and repeatedly solving PnP has the following issues:

Slow speed: The time increases with the number of samples.

Inaccuracy: P3P uses only three keypoints, resulting in low-quality pose samples.

Large intervals: The convex hull of noisy pose samples unnecessarily expands the final confidence interval.

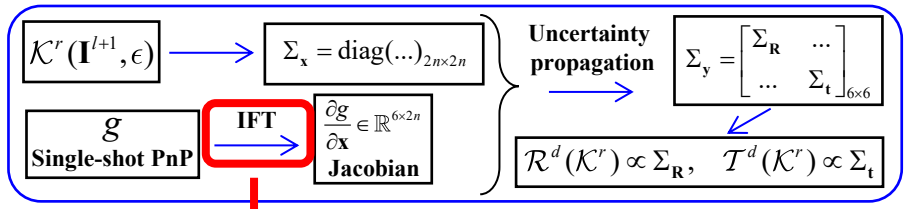
3.4 Implicit Function Theorem

$\mathcal{K}^r(\mathbf{I}^{l+1}, \epsilon)$

(c) 2D Keypoint Confidence Region

Deterministic Propagation

(f)



Implicit Function Theorem

$$\mathbf{y} = g(\mathbf{x}, \mathbf{z}, \mathbf{K})$$

PnP function

$$O(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{K}) = \sum_{n=1}^{N} \|\mathbf{r}_n\|_2^2$$

Objective function

$$\frac{\partial O(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{K})}{\partial \mathbf{y}} = 0$$

Stationary Condition

$$f(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{K}) = [f_1, ..., f_m]^T$$

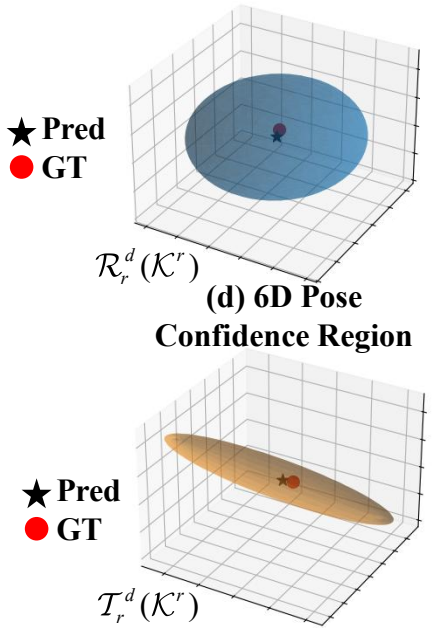
$$f_j = \frac{\partial O(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{K})}{\partial y_j} = 2\sum_{n=1}^{N} \langle \mathbf{r}_n, -2\frac{\partial \boldsymbol{\pi}_n}{\partial y_j} \rangle$$

IFT Constraint Function

$$\frac{\partial g}{\partial \mathbf{x}} = -\left[\frac{\partial f}{\partial \mathbf{y}}\right]^{-1}\left[\frac{\partial f}{\partial \mathbf{x}}\right]$$

Jacobian

3.5 6D Pose Confidence Region



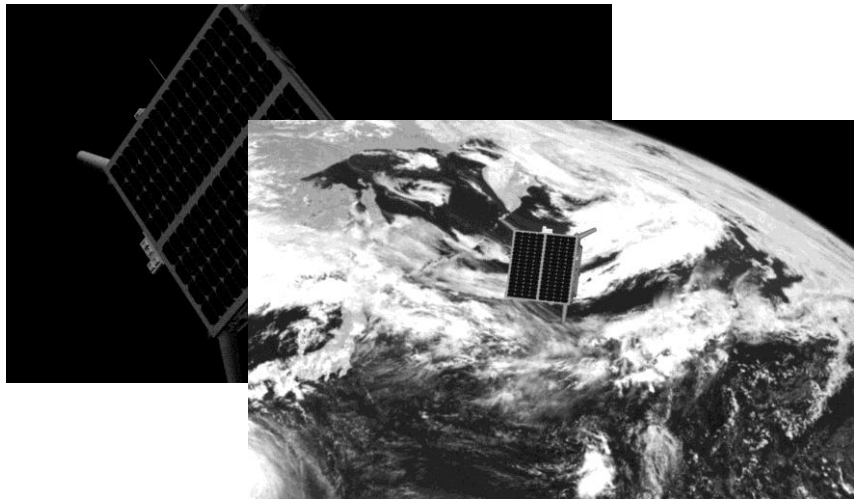
(d) 6D Pose Confidence Region

Based on the Jacobian matrix, we apply uncertainty propagation from the 2D keypoint to 6D pose.

$$\mathbf{\Sigma_y} = \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \mathbf{\Sigma_x} \left(\frac{\partial \mathbf{g}}{\partial \mathbf{x}} \right)^T$$

This covariance matrix defines an ellipsoid in the 6D pose space, which is the final, compact confidence region output by our method.

Dataset



SPEED (Spacecraft Pose Estimation Dataset): A satellite dataset for testing applicability in safety-critical scenarios.

LMO (LineMOD Occlusion): Images of common objects with significant object clutter and occlusion.

4 Dataset & Metric

Keypoint Coverage Rate: **The probability that the confidence region contains the ground truth keypoint.**

$$\eta^{kpt} = \frac{1}{K}\sum_{k=1}^{K}\mathbb{I}\left(\tilde{\mathbf{x}}^k \in \mathcal{K}^h(\mathbf{I}^k, \epsilon)\right)$$

6D Pose Coverage Rate: **The probability that the confidence region contains the ground truth pose.**

$$\eta^{\mathbf{R}} = \frac{1}{K}\sum_{k=1}^{K}\mathbb{I}\left(\mathbf{R} \in \mathcal{R}_h^s(\mathbf{I}^k, \epsilon)\right), \eta^{\mathbf{t}} = \frac{1}{K}\sum_{k=1}^{K}\mathbb{I}\left(\mathbf{t} \in \mathcal{R}_h^s(\mathbf{I}^k, \epsilon)\right)$$

Confidence Region Volume: **A measure of the compactness (or tightness) of the confidence region.**

$$V_{\mathbf{R}} = \frac{4}{3}\pi\sqrt{\det(\mathbf{\Sigma}_{\mathbf{R}})}, V_{\mathbf{t}} = \frac{4}{3}\pi\sqrt{\det(\mathbf{\Sigma}_{\mathbf{t}})}$$

5.1 Runtime and Accuracy

	Ours		[9]	
	$\mathcal{K}^r$	$\mathcal{C}_r^d$	$\mathcal{K}^h$	$\mathcal{C}_h^s$
LMO [23]	**0.0038**	**0.0361**	0.0076	0.0550
SPEED [22]	**0.0032**	**0.0358**	0.0064	0.0521

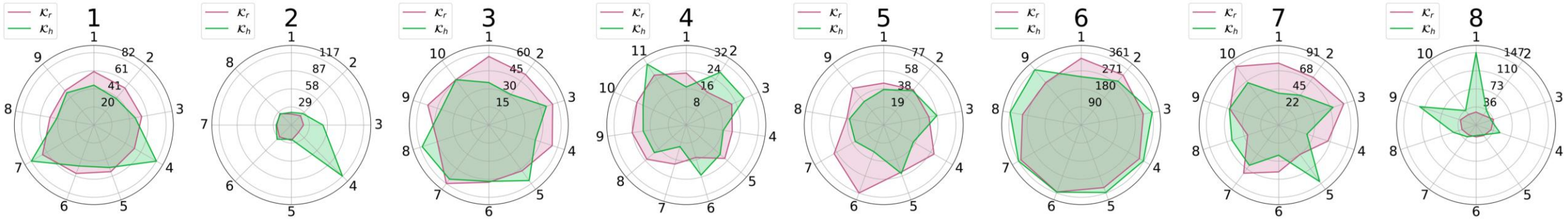
Inference and Propagation Time

LMO [23] Objects	Ours	[9]		[11]
		$\epsilon = 0.1$	$\epsilon = 0.4$	
ape (1)	76.78	77.70	**79.52**	69.14
can (2)	**91.96**	73.41	75.97	86.09
cat (3)	90.11	87.36	**90.59**	65.12
driller (4)	**89.29**	79.32	83.08	61.44
duck (5)	**84.39**	82.71	82.54	73.06
eggbox (6)	6.85	0	0	**8.43**
glue (7)	69.83	56.49	**71.08**	55.37
holepuncher (8)	**86.44**	81.65	82.89	69.84
mean	**74.45**	67.33	70.71	61.06
SPEED [22]	**97.09**	57.80	57.40	57.46

2D Keypoint Detection Accuracy

5.2 2D Keypoint Confidence Region

Radar Chart of 2D Keypoint Confidence Region Radii



Regression-based

Heatmap-based

2D Keypoint Confidence Region Coverage Rate

LMO [23] Objects	Ours $\mathcal{K}^r$		[9] $\mathcal{K}^h$	
	$\epsilon = 0.1$	$\epsilon = 0.4$	$\epsilon = 0.1$	$\epsilon = 0.4$
1	**90.37**	61.51	88.35	64.87
2	89.97	59.90	**93.54**	61.81
3	91.63	64.63	**92.30**	63.69
4	**91.84**	59.30	90.86	64.74
5	**90.31**	64.09	90.13	64.94
6	88.40	59.90	**88.86**	60.37
7	**92.02**	54.34	91.97	59.83
8	90.66	59.42	**93.31**	66.86
mean	90.65	60.38	**91.16**	63.38
SPEED [22]	**89.66**	61.25	88.88	62.64

5.3 6D Pose Confidence Region

Coverage Rate

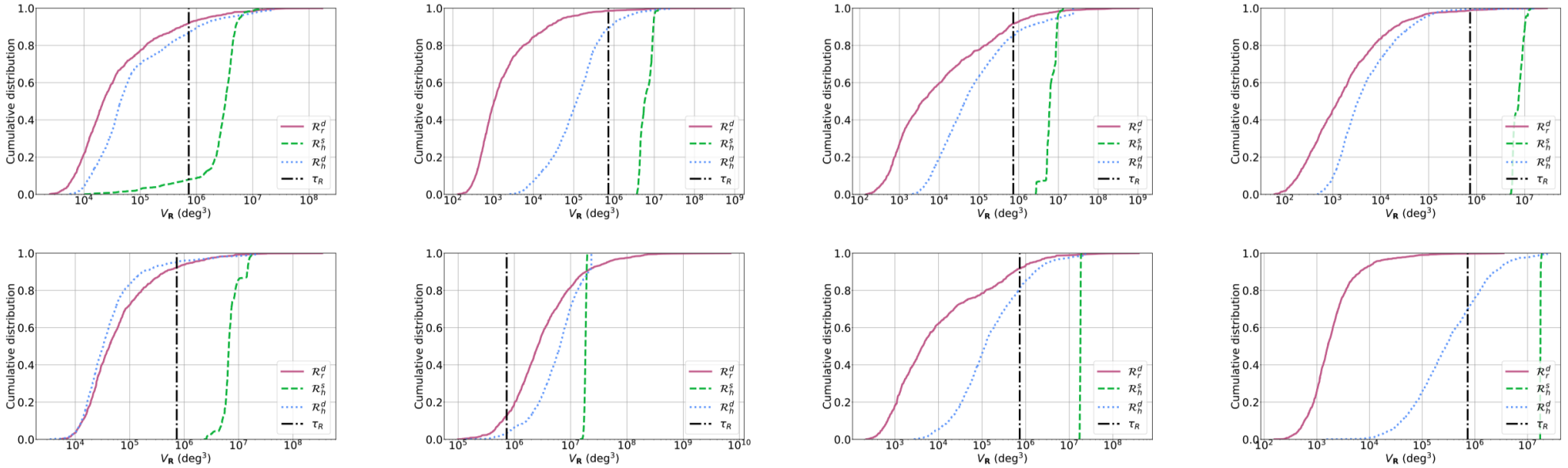
LMO Objects	Ours		[9] + Samp.		[9] + Det.	
	$\mathcal{T}_r^d$	$\mathcal{R}_r^d$	$\mathcal{T}_h^s$	$\mathcal{R}_h^s$	$\mathcal{T}_h^d$	$\mathcal{R}_h^d$
1	70.52	91.26	**97.26**	N/A	76.88	**93.38**
2	88.73	89.98	99.25	N/A	**99.59**	**98.18**
3	77.28	87.55	**98.29**	N/A	88.02	**90.59**
4	85.09	96.62	**97.12**	N/A	90.28	**98.85**
5	97.18	**79.70**	**99.81**	N/A	90.13	76.41
6	98.36	**1.46**	77.81	N/A	**98.63**	1.37
7	79.73	87.76	**98.64**	N/A	89.99	**91.84**
8	69.01	86.94	**99.92**	N/A	98.51	**98.02**
mean	83.24	77.66	**96.61**	N/A	91.50	**81.08**
SPEED	86.69	88.81	6.40	N/A	**87.10**	**90.92**

N/A: Indicates that the confidence regions of all images exceed the threshold.

5.3 6D Pose Confidence Region

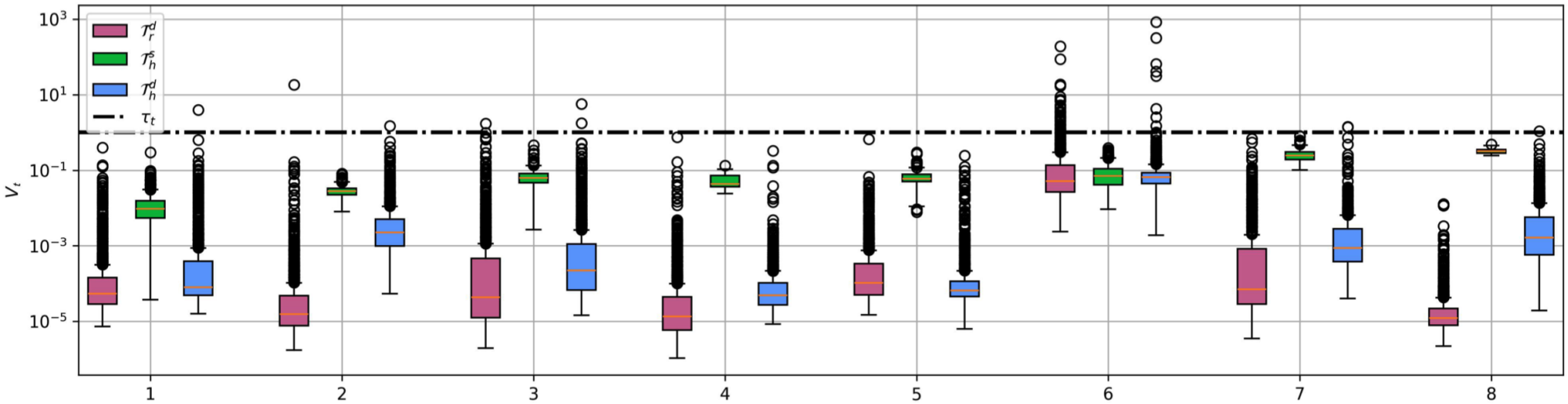
Rotation

Confidence Region volume Distribution



Translation

Confidence Region volume Distribution



5.3 6D Pose Confidence Region

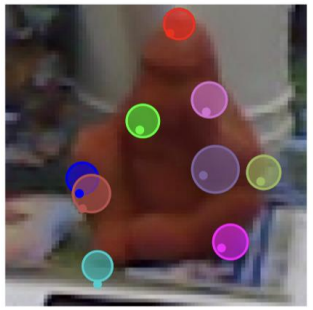
Confidence Region Volume

LMO Objects	Ours				[9] + Samp.				[9] + Det.			
	$\mathcal{T}_r^d$	Out	$\mathcal{R}_r^d$	Out	$\mathcal{T}_h^s$	Out	$\mathcal{R}_h^s$	Out	$\mathcal{T}_h^d$	Out	$\mathcal{R}_h^d$	Out
1	0.9	0	**28.96**	69	12.6	0	252.2	1041	**0.5**	0	50.3	64
2	**0.7**	0	**9.9**	9	28.3	0	N/A	1207	0.9	0	45.0	12
3	2.8	3	50.7	113	66.8	0	N/A	1052	**1.4**	0	**42.9**	73
4	1.5	0	6.5	12	52.0	0	N/A	1214	**0.1**	0	**3.5**	2
5	0.4	0	52.8	43	63.1	0	N/A	1064	**0.2**	0	**24.7**	24
6	**57.6**	11	**514.4**	1055	82.2	1	N/A	1095	59.9	10	515.7	1057
7	4.8	0	**62.9**	98	252.4	0	N/A	809	**0.8**	0	67.5	56
8	**0.03**	0	**4.6**	3	322.2	0	N/A	1210	0.3	0	51.6	9
mean	8.6	2	**91.3**	175	109.9	0	252.2	1070	**8.0**	1	100.1	162
SPEED	**0.7**	0	**0.2**	0	287.8	920	210.6	898	73.4	50	4.4	18

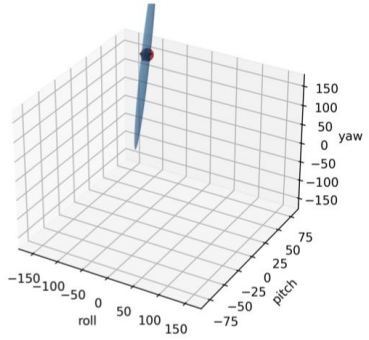
Out: The number of images whose confidence region volume exceeds the threshold.

N/A: Indicates that the confidence regions of all images exceed the threshold.

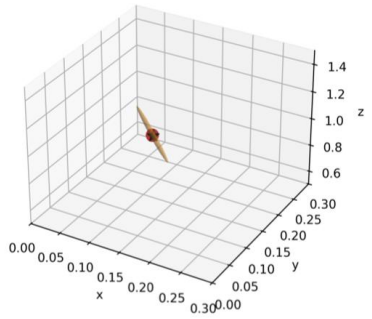
5.4 Confidence Region Visualization



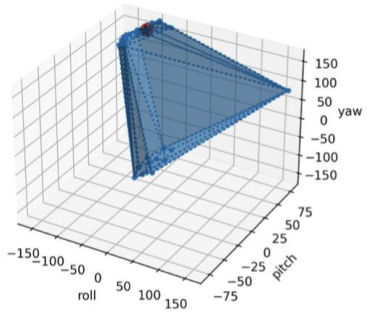
(a) $\mathcal{K}^r$



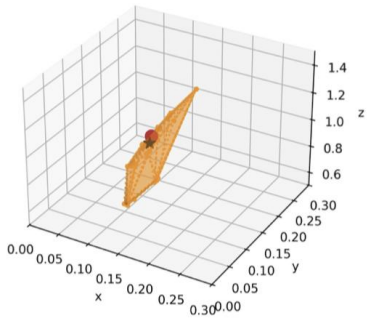
(b) $\mathcal{R}_r^d : V_{\mathbf{R}} = 19.6$



(c) $\mathcal{T}_r^d : V_{\mathbf{t}} = 0.05$

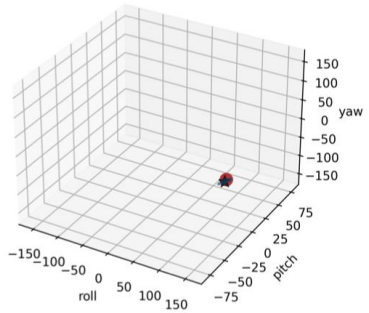


(d) $\mathcal{R}_h^s : V_{\mathbf{R}} = 2479.6$

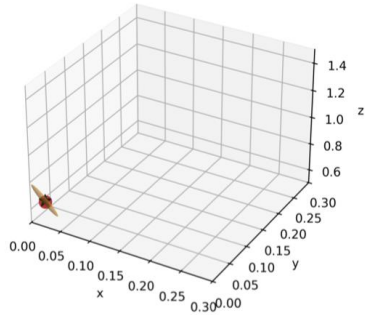


(e) $\mathcal{T}_h^s : V_{\mathbf{t}} = 0.5$

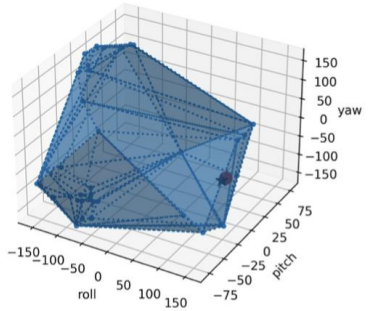
(f) $\mathcal{K}^r$



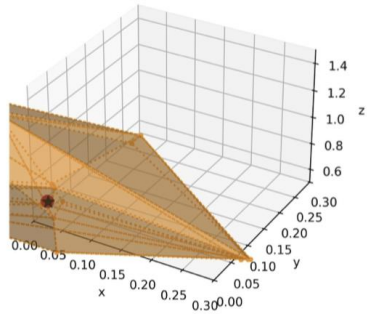
(g) $\mathcal{R}_r^d : V_{\mathbf{R}} = 1.0$



(h) $\mathcal{T}_r^d : V_{\mathbf{t}} = 0.01$

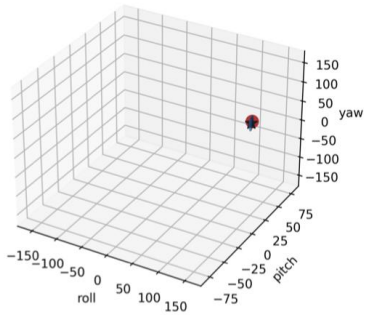


(i) $\mathcal{R}_h^s : V_{\mathbf{R}} = 8472.9$

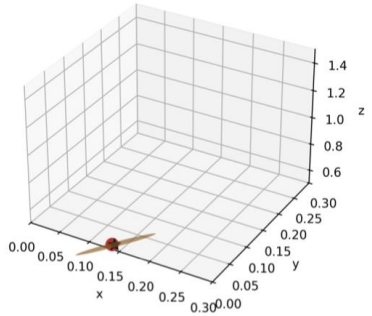


(j) $\mathcal{T}_h^s : V_{\mathbf{t}} = 53.8$

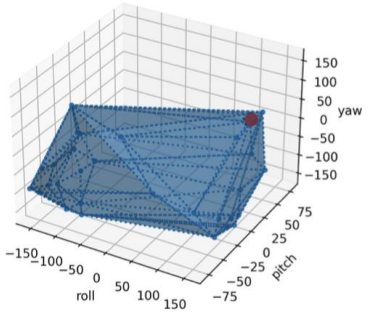
(k) $\mathcal{K}^r$



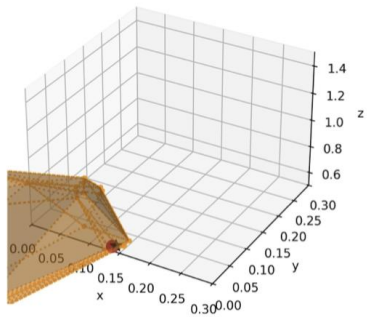
(l) $\mathcal{R}_r^d : V_{\mathbf{R}} = 1.9$



(m) $\mathcal{T}_r^d : V_{\mathbf{t}} = 0.03$

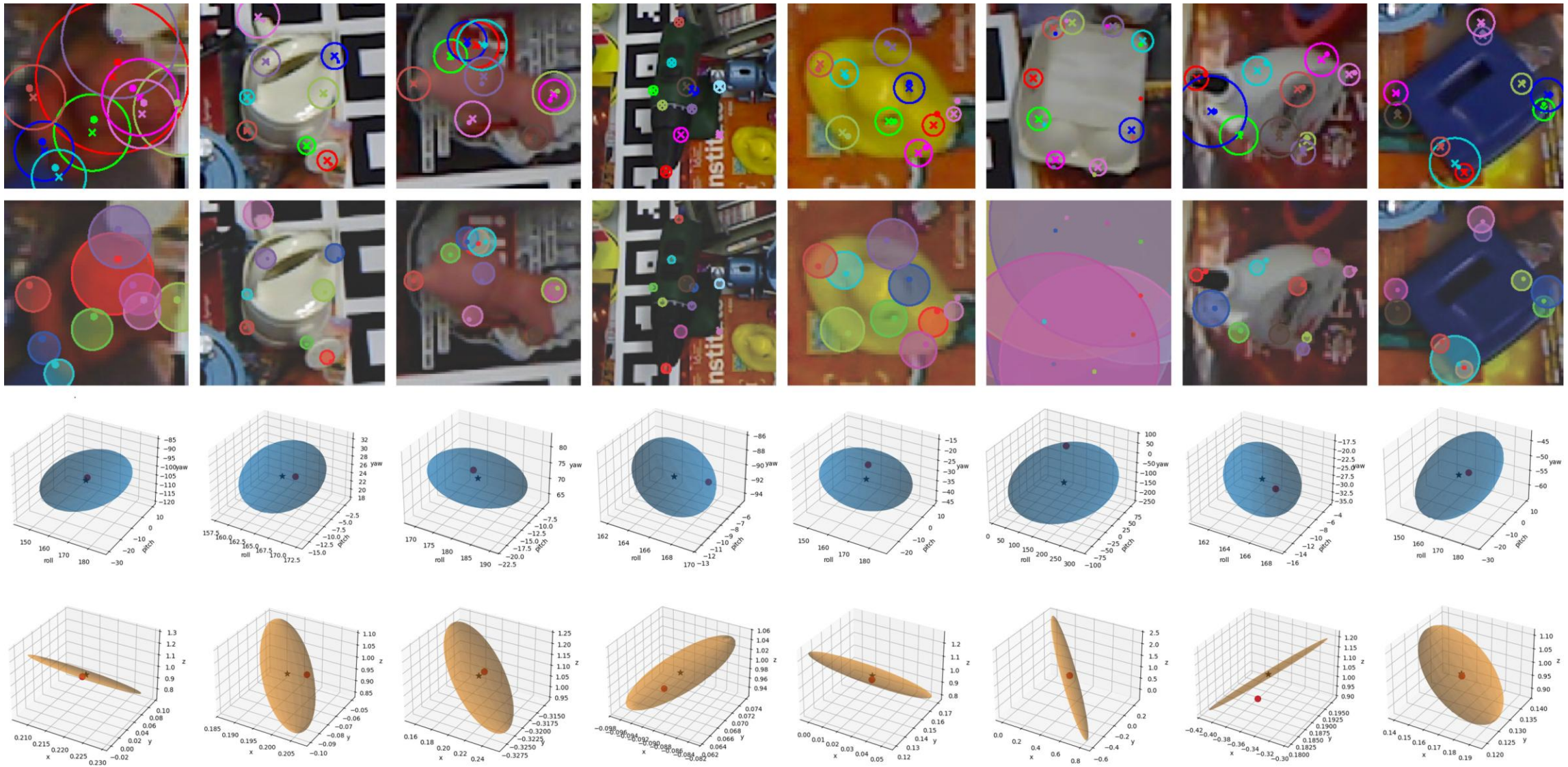


(n) $\mathcal{R}_h^s : V_{\mathbf{R}} = 5779.0$

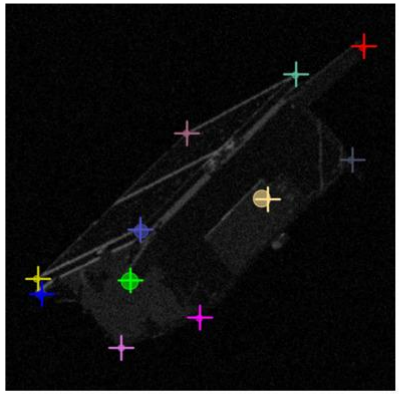
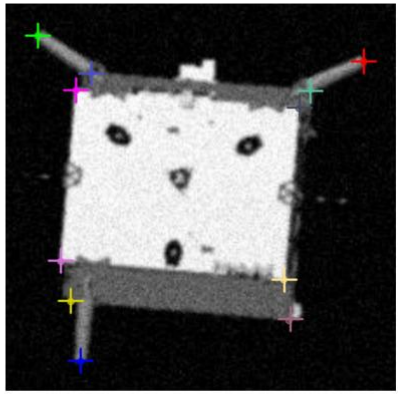
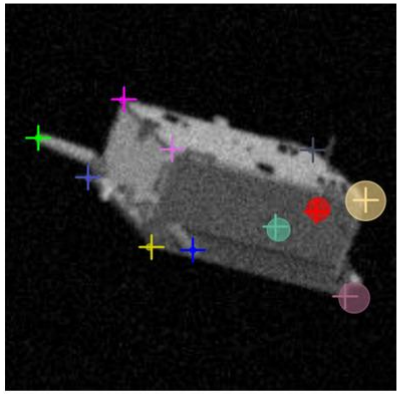
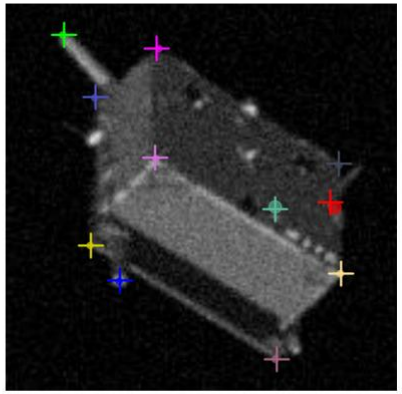
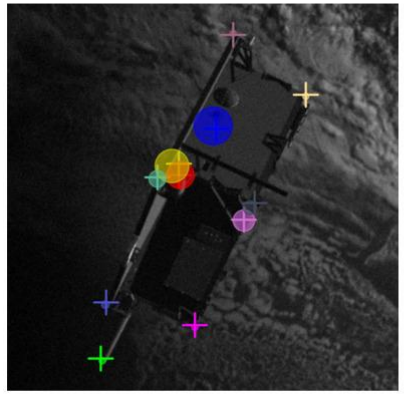
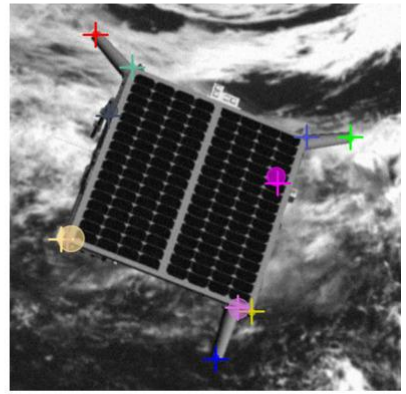


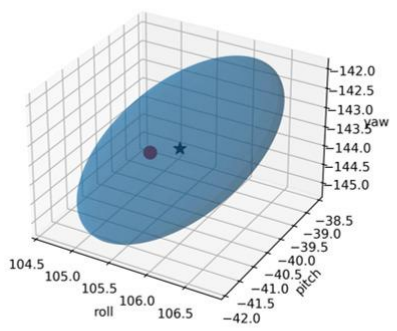
(o) $\mathcal{T}_h^s : V_{\mathbf{t}} = 14.8$

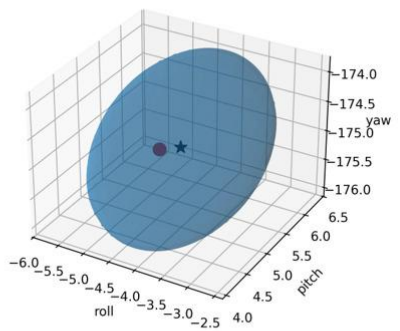
5.4 Confidence Region Visualization

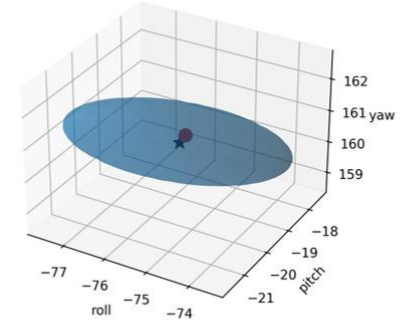


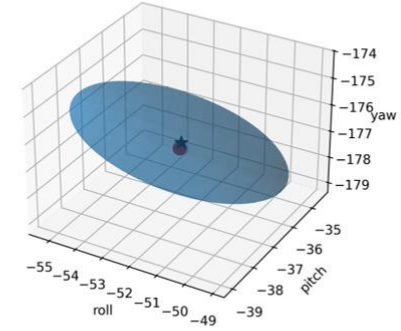
5.4 Confidence Region Visualization

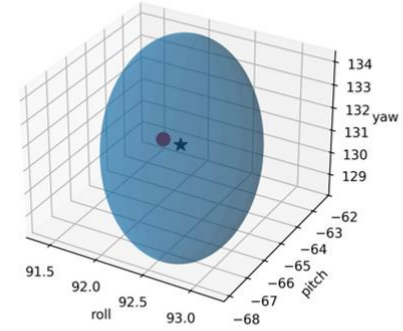







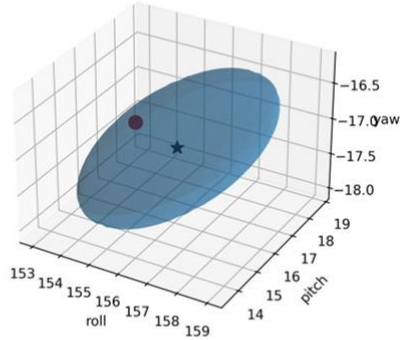


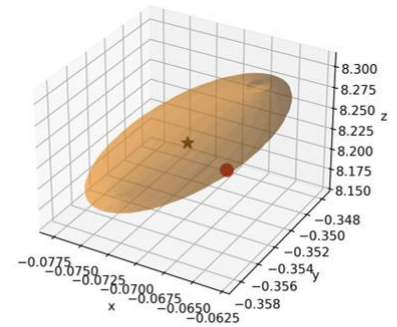


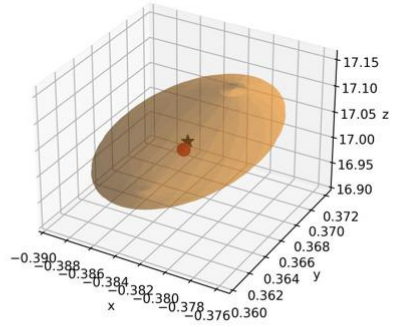


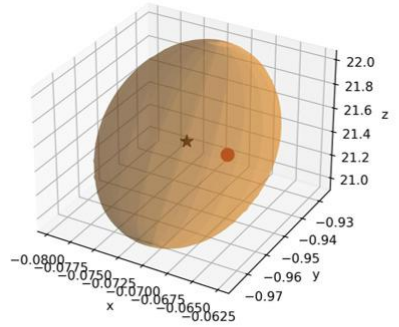


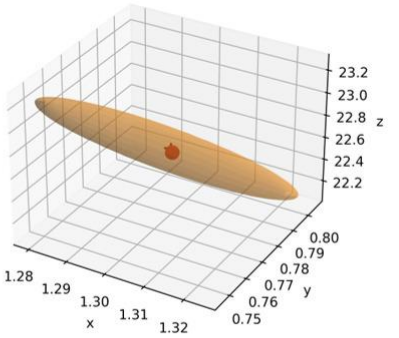


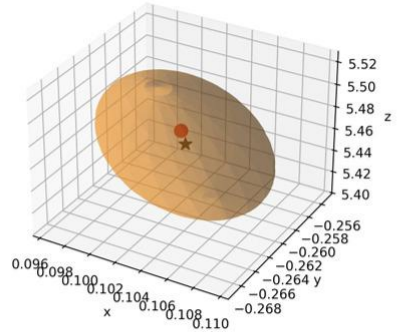


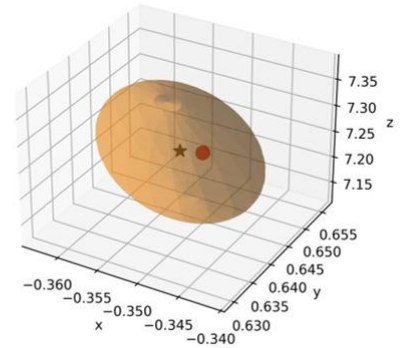












6 Conclusion

We propose an efficient and direct framework that employs an analytical method based on the Implicit Function Theorem to estimate compact 6D pose confidence regions, superseding slow and inaccurate sampling-based approaches.

Our method provides statistically guaranteed data to support subsequent safety-critical applications.