Rethinking Pose Refinement in 3D Gaussian Splatting under Pose Prior and Geometric Uncertainty

Supplementary Material

Supplementary

In the supplementary material, we show

- integrating retrieval-based pose initialization into our method;
- a comparison using various matching methods;
- an ablation study analyzing each module in our pipeline;
- a runtime discussion;
- an additional implementation detail.

A. Leveraging Image Retrieval Prior

Pose refinement in 3D Gaussian Splatting relies heavily on the quality of the initial pose prior, since the refinement module only adjusts the pose locally around this estimate. Consequently, the choice of pose estimator providing the prior, typically an image retrieval system, APR, or SCR, plays a crucial role in determining the overall localization performance. Among these options, image retrieval offers a unique advantage: unlike APR or SCR, it does not require any per-scene training. When combined with a 3DGS scene representation, this enables a purely geometry-driven localization pipeline that performs relocalization without additional learning.

As illustrated in Fig. A, conventional 3DGS-based refinement pipelines use the pose of the Top-1 retrieved database (DB) image as the pose prior. While this approach is simple and efficient, it inherently assumes that the retrieved Top-1 image is spatially close to the query. In practice, this assumption frequently breaks down. Even if the retrieved DB image is visually similar to the query, it may still be captured from a different viewpoint, for example, the opposite side of the same building. Fig. B demonstrates such a failure case in the Cambridge Landmarks *Church* scene. The Top-1 retrieved image exhibits high visual similarity, yet its pose lies on the opposite side of the church, leading to a large spatial discrepancy. Refining such an incorrect prior remains challenging, regardless of the rendering quality or scene detail available in the 3DGS model.

Our proposed UGS-Loc addresses this limitation by integrating retrieval results into the Monte Carlo refinement framework. Instead of relying on a single deterministic prior, we use the poses of the Top-K retrieved DB images as initial hypotheses (particles). Each hypothesis is assigned an importance weight derived from matching confidence and geometric uncertainty, allowing the system to down-weight misleading candidates and preserve only the reliable ones through resampling. As shown in Fig. B, although most initial particles lie on the incorrect side of the building, UGS-Loc samples the poses on the correct region by leveraging importance weighting, which leverages the confidence from the matching module and uncertainty of matching points. This demonstrates the robustness of our retrieval-integrated sampling scheme and highlights its advantage over traditional deterministic pose refinement.

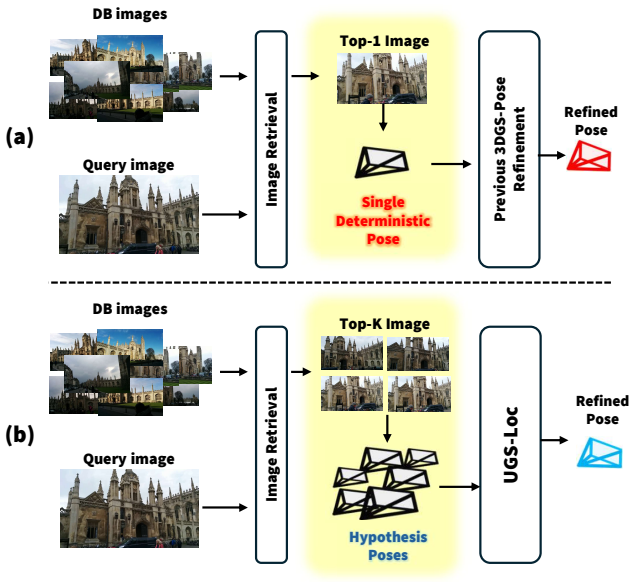

Figure A. **Leveraging Image Retrieval into UGS-Loc.** Illustration comparing deterministic 3DGS-based pose refinement and our UGS-Loc when using image retrieval as a pose prior. (a) Prior methods rely solely on the Top-1 retrieved image to obtain a single deterministic pose hypothesis, which is then refined through a 3DGS-based pipeline, making them vulnerable to incorrect or ambiguous retrieval results. (b) Our UGS-Loc instead leverages Top-K retrieved images to generate multiple pose hypotheses, followed by importance weighting and refinement. This enables the localization to suppress misleading hypotheses and converge reliably toward the true camera pose.

We evaluate the effectiveness of integrating image retrieval priors into our UGS-Loc framework, as shown in Tabs. A and C. We utilize NetVlad [1] as an image retrieval model. For a fair comparison, we also include an iterative extension of the standard 3DGS-based pose refinement pipeline, in which the refined pose is repeatedly fed back as the new initialization. This allows us to directly compare UGS-Loc against both the conventional single-step refinement and its iterative variant.

Tab. C reports the localization results on the large-scale

Table A. **Localization Result on 7scenes with Image Retrieval.** We report the median translation error (cm) and rotation errors (°). Pose prior is initialized with the poses of the database images retrieved by NetVlad [1].

Iteration	Method	Top-K	Chess	Fire	Heads	Office	Pumpkin	Redkitchen	Stairs	Avg. ↓ [cm/°]
1	GS-CPR	1	1.66/0.43	2.38/0.61	1.34/0.67	3.33/0.65	2.16/0.45	2.56/0.54	4.51/1.10	2.56/0.64
2		1	0.56/0.16	0.80/0.28	0.48/0.32	1.18/0.31	1.13/0.25	0.77/0.19	2.24/0.66	1.03/0.31
2	**UGS-Loc (Ours)**	5	**0.40/0.13**	**0.54/0.21**	**0.40/0.27**	**0.88/0.23**	**0.85/0.19**	**0.64/0.16**	**1.80/0.50**	**0.79/0.24**

Table B. **Average Recall on 7-Scenes with Image Retrieval.** We report the recall rate within the $[2\text{cm}, 2^\circ]$ and $[5\text{cm}, 5^\circ]$ error thresholds across the 7Scenes.

Iteration	Method	Top-K	Acc ↑ $[2\text{cm}, 2^\circ]$	Acc ↑ $[5\text{cm}, 5^\circ]$
1	GS-CPR	1	43.0	76.6
2		1	79.9	94.4
2	**UGS-Loc**	5	**88.9**	**97.2**
	ACE		83.3	97.1

Cambridge Landmarks dataset. Remarkably, when combined with image retrieval, our UGS-Loc achieves performance comparable to the performance of ACE [2], a Scene Coordinate Regression method that relies on per-scene training. Furthermore, compared to using only the Top-1 retrieved image as a deterministic pose prior, UGS-Loc reduces the median translation error by 17%, despite operating under the same retrieval assumptions and without any additional learning.

Although the retrieval-based experiment involves one additional refinement iteration compared to the standard setting, the improvement remains noteworthy. This demonstrates that UGS-Loc effectively mitigates the uncertainty introduced by imperfect retrieval priors and consistently converges toward accurate poses. The evaluation on the 7Scenes dataset, presented in Tab. A, further confirms this trend. UGS-Loc robustly handles pose priors obtained from retrieval, outperforming both the original 3DGS-based refinement and its iterative variant across all scenes.

B. Various Matching Modules

Our UGS-Loc framework refines poses using 2D–3D correspondences obtained by lifting 2D–2D matches through rendered depth. Since correspondence quality directly depends on the underlying 2D matcher, we additionally evaluate UGS-Loc with a broader set of matching modules beyond the main paper, ELoFTR [17], XFeat [14], and Match-Anything [9], and report the median translation and rotation errors on the Cambridge Landmarks dataset in Tab. D.

Across all matchers, UGS-Loc consistently achieves strong pose refinement performance, demonstrating that our method does not rely on any specific matching backbone. While transformer-based dense matchers generally provide higher correspondence coverage, even lightweight or classical matchers yield comparable localization accuracy when combined with UGS-Loc. This robustness highlights that the gains from our framework primarily arise from uncertainty-guided sampling and multi-hypothesis pose refinement, rather than from matcher capacity itself.

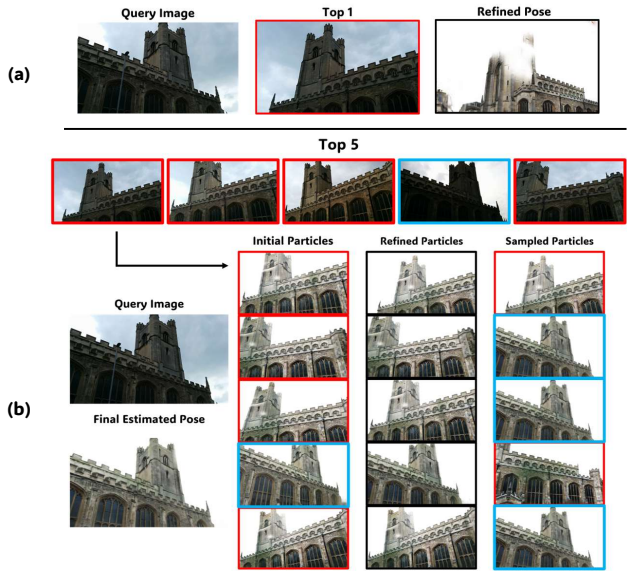

Figure B. **Visualization of UGS-Loc with Image Retrieval.** We illustrate how integrating image retrieval with our pipeline enables robust pose refinement. In the Cambridge *Church* scene, the deterministic Top-1 baseline (a) suffers from a failure case where the retrieved image (red box) views the church from the opposite side of the scene, yielding an incorrect pose prior. In contrast, using Top-K retrieved poses (b), UGS-Loc effectively suppresses such erroneous hypotheses and converges to the correct pose, highlighted by blue boxes.

C. Module Ablation

Tabs. E and F report how each component of our pipeline, Monte Carlo Refinement (MCR) and Uncertainty-based PnP (UPnP), contributes to the final localization accuracy. We evaluate both DFNet [3] and ACE [2] as pose priors to analyze the generality of each module across different initialization qualities. For both priors, enabling only UPnP yields modest improvements by suppressing unreliable correspondences through depth-uncertainty guidance. Incorpo-

Table C. **Localization Result on 7scenes with Image Retrieval.** We report the median translation error (cm) and rotation errors (°). Pose prior is initialized with the poses of the database images retrieved by NetVlad [1].

Iteration	Method	Top-K	Kings	Hospital	Shop	Church	Avg. ↓ [cm/°]
1	GS-CPR	1	26/0.33	35/0.57	13/0.43	23/0.61	24/0.49
2		1	24/0.26	27/0.46	6.2/0.25	8.8/0.27	17/0.31
3		1	19/0.23	20/0.41	5.9/0.21	6.6/0.22	13/0.27
3	**UGS-Loc (Ours)**	5	**18**/0.20	**15**/0.32	**4.3/0.17**	5.4/0.18	**11**/0.22
3		10	**18/0.19**	**15/0.31**	4.5/**0.17**	**5.1/0.17**	**11/0.21**
	ACE + UGS-Loc		18/0.18	14/0.30	4.2/0.16	6.3/0.20	11/0.21

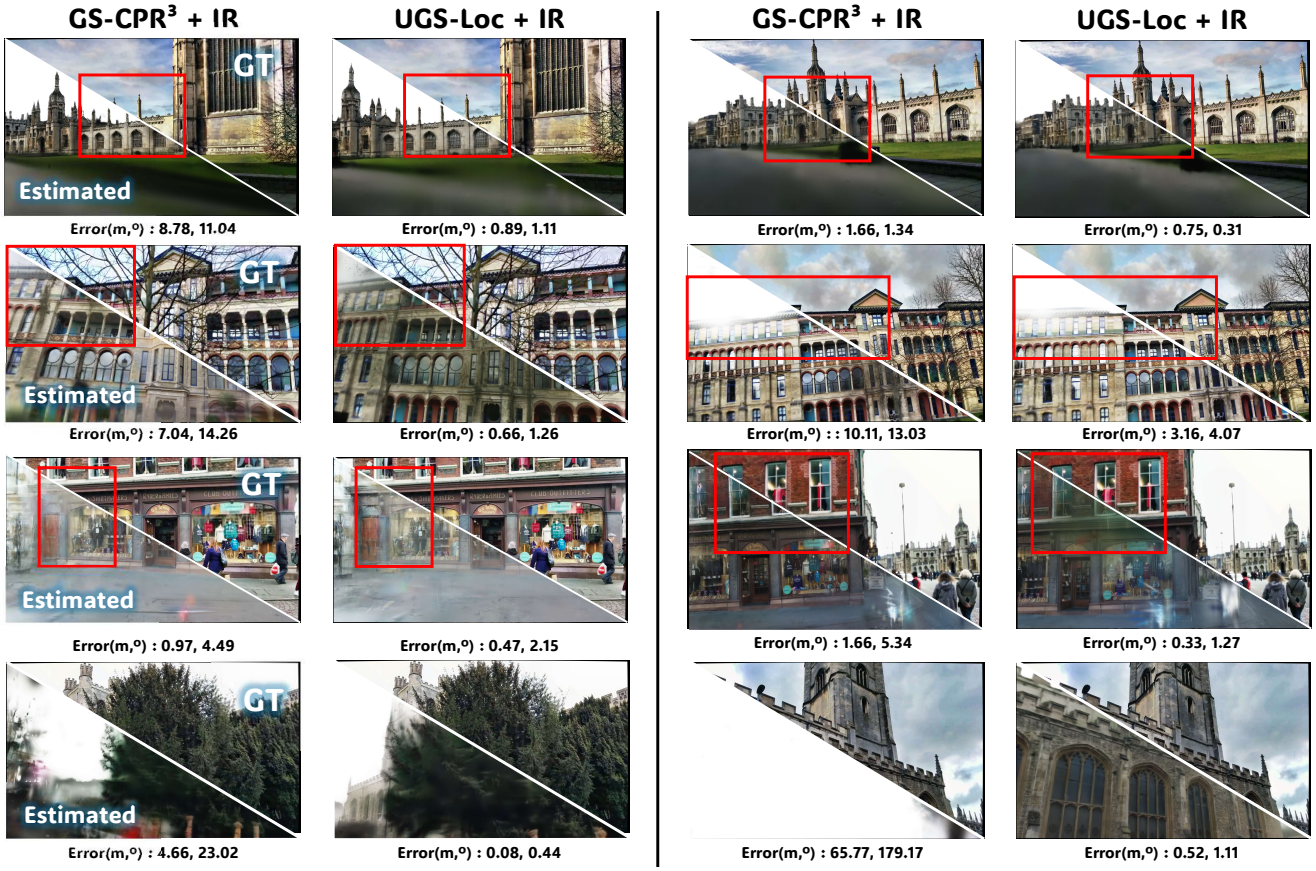

Figure C. **Visualization of Localization Quality on the Cambridge Landmark Dataset.** Each pair of columns compares the ground-truth view (top-right) against the view rendered from (i) the iteratively modified GS-CPR refinement with 3 iterations (denoted as GS-CPR^3) [12] and (ii) our UGS-Loc refinement (bottom-left). IR denotes that the pose prior is initialized with Image Retrieval (IR) [1]. Tighter visual alignment along the diagonal boundary indicates a more accurate pose estimate. Red bounding boxes highlight regions where misalignment is most apparent, emphasizing how UGS-Loc corrects errors that remain unresolved by deterministic 3DGS-based refinement.

rating only MCR produces a larger gain, reflecting the benefit of exploring multiple pose hypotheses rather than relying on a single deterministic prior.

As shown in Tab. E, applying our uncertainty-based PnP alone already improves the median translation error by roughly 11% compared to the standard 3DGS-based pose refinement. When combined with the modified Monte Carlo refinement, the performance further improves with two iterations; the MCL structure achieves a substantial reduction in both translation and rotation errors. When the two modules are combined, we observe the best performance across all scenes. Under DFNet pose prior, the combined system reduces the average error to 11cm and 0.22°, which is the same performace utilizing a better pose prior (ACE). This

Table D. **Evaluation with Different 2D Matching Module.** We report the median translation error (cm) and rotation error (°) on the Cambridge Landmark dataset [7]. We utilize various matching modules to establish 2D-3D correspondences.

Prior	Matchers	Kings	Hospital	Shop	Church	Avg. ↓ [cm/°]
DFNet	SP [15] + LG [11]	20/0.22	17/0.31	5.5/0.26	7.7/0.25	13/0.26
	Xfeat [14]	**19/0.18**	15/0.28	5.5/0.26	6.6/0.20	12/0.23
	ELoFTR [17]	20/0.21	**13/0.25**	4.1/0.19	6.5/0.20	**11/0.20**
	MatchAny [9]	20/0.20	14/**0.25**	4.0/0.16	6.3/0.20	**11/0.20**
	MASt3R [8]	**19**/0.19	15/0.29	**3.9/0.15**	**5.5/0.17**	**11/0.20**
ACE	SP [5] + LG [11]	19/0.23	15/0.27	4.0/0.19	7.8/0.26	**11**/0.26
	Xfeat [14]	19/**0.19**	**12/0.25**	3.9/0.17	7.2/0.24	**11/0.21**
	ELoFTR [17]	19/0.20	13/**0.25**	3.7/**0.15**	7.0/0.23	**11/0.21**
	MatchAny [9]	19/0.20	13/0.26	**3.6**/0.17	6.6/**0.20**	**11/0.21**
	MASt3R [8]	**18/0.18**	14/0.30	4.2/0.16	**6.3/0.20**	**11/0.21**

Table E. **Module Ablation on Cambridge Landmark.** We report the median translation error (cm) and rotation error (°) on Cambridge Landmark dataset [7]. We utilize DFNet [3] and ACE [2] as a pose prior for module ablation. MCR and UPnP denote Monte Carlo refinement and Uncertainty sampling-based PnP optimization.

Prior	MCR	UPnP	Kings	Hospital	Shop	Church	Avg. ↓ [cm/°]
DFNet			21/0.25	29/0.54	9.4/0.35	14/0.38	19/0.38
		✓	20/0.25	25/0.58	9.2/0.34	14/0.39	17/0.39
	✓		**19**/0.20	17/0.30	4.4/0.18	5.8/0.18	12/**0.22**
	✓	✓	**19/0.19**	**15/0.29**	**3.9/0.15**	**5.5/0.17**	**11/0.20**
ACE			21/0.25	19 /0.36	4.7/0.20	9.5/0.29	14/0.28
		✓	19/0.21	18 /0.33	4.6/0.20	8.3/0.26	12/0.25
	✓		**18**/0.20	15 /0.31	4.3/0.18	6.8/0.21	**11**/0.23
	✓	✓	**18/0.18**	**14 /0.30**	**4.2/0.16**	**6.3/0.20**	**11/0.21**

Table F. **Module Ablation on 7scenes [16].** We report the median translation error (cm) and rotation error (°) on 7-scenes dataset. We utilize DFNet [3] as a pose prior for module ablation. MCR and UPnP denote Monte Carlo refinement and Uncertainty sampling-based PnP optimization.

Prior	MCR	UPnP	Chess	Fire	Heads	Office	Pumpkin	Redkitchen	Stairs	Avg. ↓ [cm/°]
DFNet			0.63/0.18	0.93/0.33	0.56/0.36	1.27/0.33	1.18/0.28	0.93/0.24	2.18/0.59	1.10/0.33
		✓	0.53/0.16	0.81/0.31	0.58/0.38	1.10/0.29	0.99/0.23	0.85/0.23	2.40/0.63	1.04/0.32
	✓		0.40/0.13	0.55/0.21	0.38/**0.25**	0.83/0.23	0.89/0.21	0.63/**0.15**	**1.18/0.35**	0.69/0.22
	✓	✓	**0.36/0.12**	**0.49/0.20**	**0.36/0.25**	**0.78/0.22**	**0.81/0.18**	**0.59/0.15**	1.19/**0.35**	**0.65/0.21**

consistent convergence across different priors demonstrates that each module addresses a distinct source of uncertainty. MCR mitigates pose prior bias, whereas UPnP handles geometric unreliability during 2D–3D lifting.

Overall, the ablation confirms that both components contribute complementary strengths, and their combination forms a robust and reliable pose refinement pipeline independent of the choice of pose estimator.

D. Inference Cost

We do not claim that UGS-Loc offers faster inference than deterministic 3DGS-based pose refinement. Because our framework integrates the conventional 3DGS refinement pipeline with an MCL-inspired multi-hypothesis strategy, additional computation is naturally introduced. However, rather than executing iterative refinement serially as in traditional 3DGS pipelines, UGS-Loc leverages *multiprocessing* to evaluate multiple pose hypotheses in parallel. This design helps mitigate much of the computational overhead and yields practical efficiency in multi-particle settings. Another important factor influencing inference speed is the choice of the 2D matching module. As discussed in Sec. B, UGS-Loc maintains strong localization performance even when paired with lightweight matchers such as XFeat or SP+LG,

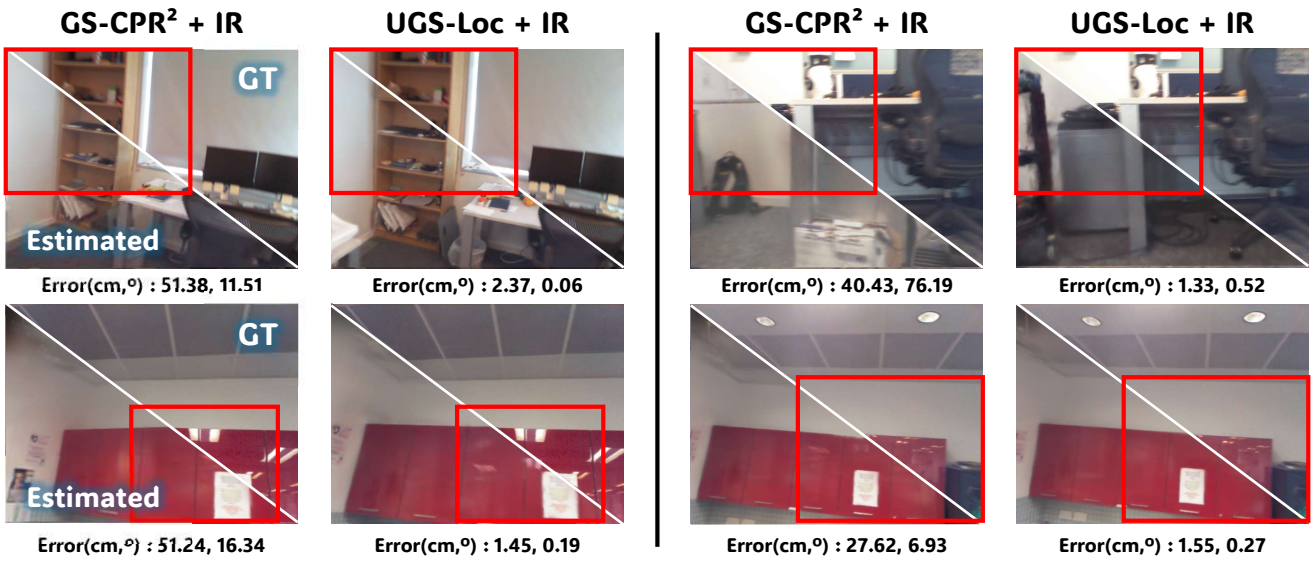

Figure D. **Visualization of Localization Quality on the 7-Scenes Dataset.** Each pair of columns compares the ground-truth view (top-right) against the view rendered from (i) the iteratively modified GS-CPR refinement with 2 iterations (denoted as GS-CPR^2) [12] and (ii) our UGS-Loc refinement (bottom-left). IR denotes that the pose prior is initialized with Image Retrieval (IR) [1].

providing flexibility for balancing speed and accuracy beyond the MASt3R-based configuration.

Comparison with GS-CPR. Due to differences in hardware, a direct inference time comparison with GS-CPR [12] requires normalization. According to the original GS-CPR paper, the runtime of each component is as follows: 3.7ms for a single rendering, 71ms for a MASt3R [8] forward pass, and 94 ms for the PnP optimization with MASt3R matching. However, under our environment, using the same codebase, rendering, MASt3R inference, and PnP optimization take 12.4ms, 189ms, and 182ms, respectively, showing a consistent slowdown across all processes. This discrepancy arises from hardware differences rather than algorithmic overhead.

E. Additional Implementation Detail

This section addresses the hyperparameters and implementation details used in our pose refinement pipeline. UGS-Loc performs PnP-based refinement over two iterations, maintaining eight particles per iteration. For efficiency, we use a 256 resolution in the first iteration and a 512 resolution in the second. For PnP optimization, the reprojection error threshold is set to 1.0 px for 7Scenes and 2.5 px or Cambridge. The maximum number of iterations for PnP-RANSAC is fixed to 1000, consistent with GS-CPR [12]. Our uncertainty-guided sampling–based PnP terminates early once the confidence threshold reaches 0.99. To prevent extreme values in correspondence uncertainty values u_i, we apply percentile clipping using the 5%–95% range. During the final iteration of Monte Carlo refinement, we re-render all particles from their refined poses and compute SSIM scores [18] to obtain the particle weights. For 7Scenes, the final pose is obtained via a weighted average, whereas for Cambridge, we select the best single particle to avoid performance degradation caused by noisy samples.

For scene representation, we adopt Scaffold-GS [13] and apply object and sky masks obtained from an off-the-shelf model [4] to prevent overly noisy Gaussian reconstructions. Unlike several prior localization works [6, 10], our 7Scenes experiments train the Gaussian Splatting model using RGB images only, without requiring depth supervision or per-scene fine-tuning with deep features beyond GS training.

References

[1] Relja Arandjelovic, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic. Netvlad: Cnn architecture for weakly supervised place recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5297–5307, 2016. 1, 2, 3, 5, 6

[2] Eric Brachmann, Tommaso Cavallari, and Victor Adrian Prisacariu. Accelerated coordinate encoding: Learning to relocalize in minutes using rgb and poses. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5044–5053, 2023. 2, 4

[3] Shuai Chen, Xinghui Li, Zirui Wang, and Victor A Prisacariu. Dfnet: Enhance absolute pose regression with direct feature matching. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part X*, pages 1–17. Springer, 2022. 2, 4

[4] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexan-

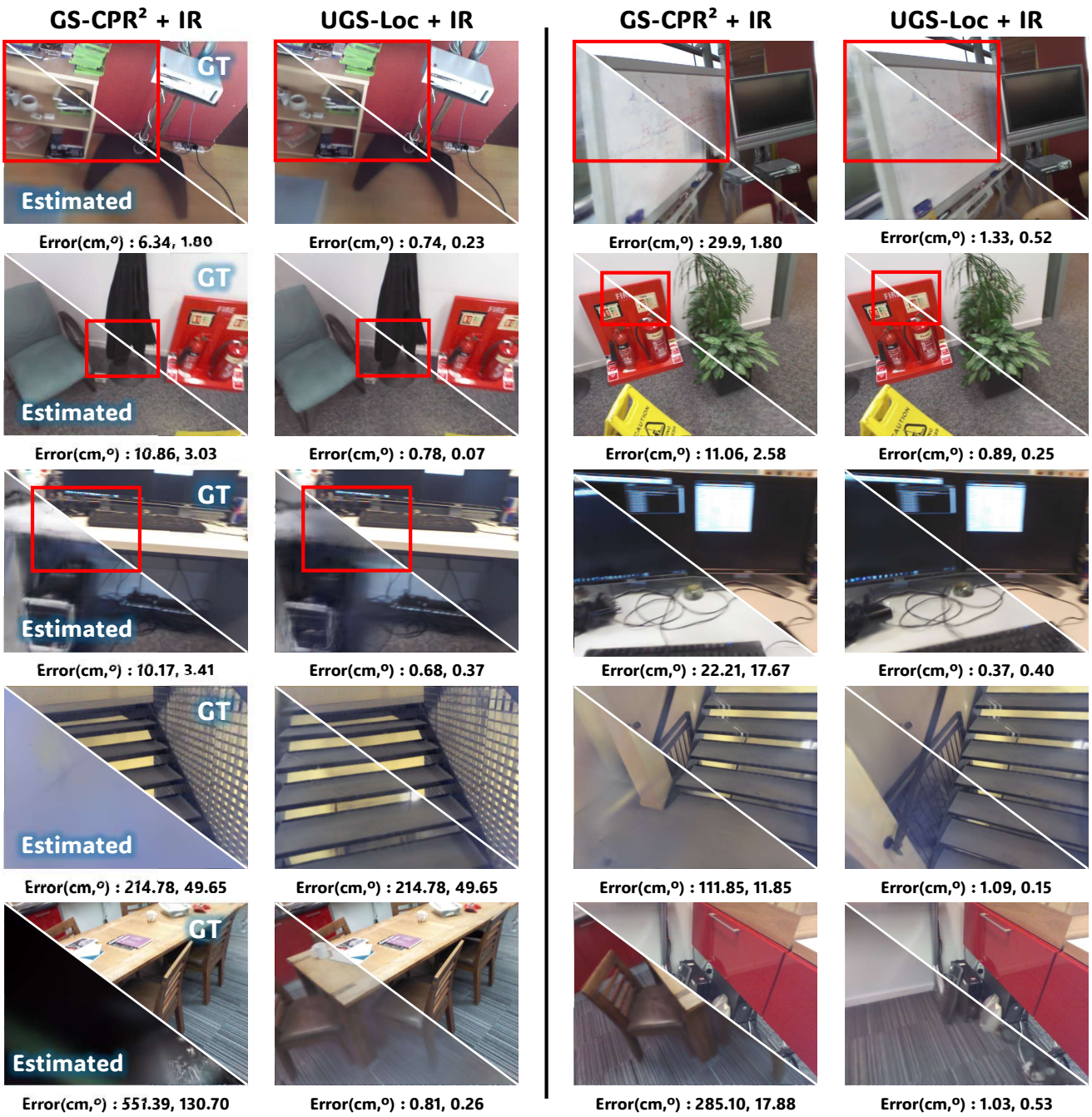

Figure E. **Visualization of Localization Quality on the 7-Scenes Dataset.** Each pair of columns compares the ground-truth view (top-right) against the view rendered from (i) the iteratively modified GS-CPR refinement with 2 iterations (denoted as GS-CPR^2) [12] and (ii) our UGS-Loc refinement (bottom-left). IR denotes that the pose prior is initialized with Image Retrieval (IR) [1].

der Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1290–1299, 2022. 5

[5] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 224–236, 2018. 4

[6] Zhiwei Huang, Hailin Yu, Yichun Shentu, Jin Yuan, and Guofeng Zhang. From sparse to dense: Camera relocalization with scene-specific detector from feature gaussian splat-

ting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 27059–27069, 2025. 5

[7] Alex Kendall, Matthew Grimes, and Roberto Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *Proceedings of the IEEE international conference on computer vision*, pages 2938–2946, 2015. 4

[8] Vincent Leroy, Yohann Cabon, and Jérôme Revaud. Grounding image matching in 3d with mast3r. In *European Conference on Computer Vision*, pages 71–91. Springer, 2024. 4, 5

[9] Siyuan Li, Lei Ke, Martin Danelljan, Luigi Piccinelli, Mattia Segu, Luc Van Gool, and Fisher Yu. Matching anything by segmenting anything. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18963–18973, 2024. 2, 4

[10] Sihang Li, Siqi Tan, Bowen Chang, Jing Zhang, Chen Feng, and Yiming Li. Unleashing the power of data synthesis in visual localization. *arXiv preprint arXiv:2412.00138*, 2024. 5

[11] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. Lightglue: Local feature matching at light speed. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17627–17638, 2023. 4

[12] Changkun Liu, Shuai Chen, Yash Bhalgat, Siyan Hu, Ming Cheng, Zirui Wang, Victor Adrian Prisacariu, and Tristan Braud. Gs-cpr: Efficient camera pose refinement via 3d gaussian splatting. *arXiv preprint arXiv:2408.11085*, 2024. 3, 5, 6

[13] Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai. Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20654–20664, 2024. 5

[14] Guilherme Potje, Felipe Cadar, André Araujo, Renato Martins, and Erickson R Nascimento. Xfeat: Accelerated features for lightweight image matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2682–2691, 2024. 2, 4

[15] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4938–4947, 2020. 4

[16] Jamie Shotton, Ben Glocker, Christopher Zach, Shahram Izadi, Antonio Criminisi, and Andrew Fitzgibbon. Scene coordinate regression forests for camera relocalization in rgb-d images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2930–2937, 2013. 4

[17] Yifan Wang, Xingyi He, Sida Peng, Dongli Tan, and Xiaowei Zhou. Efficient loftr: Semi-dense local feature matching with sparse-like speed. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 21666–21675, 2024. 2, 4

[18] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 5