GS-CPR: Efficient Camera Pose Refinement via 3D Gaussian Splatting

Changkun Liu[1], Shuai Chen[2], Yash Bhalgat[2], Siyan Hu[1], Ming Cheng[3], Zirui Wang[2], Victor Prisacariu[2] and Tristan Braud[1]

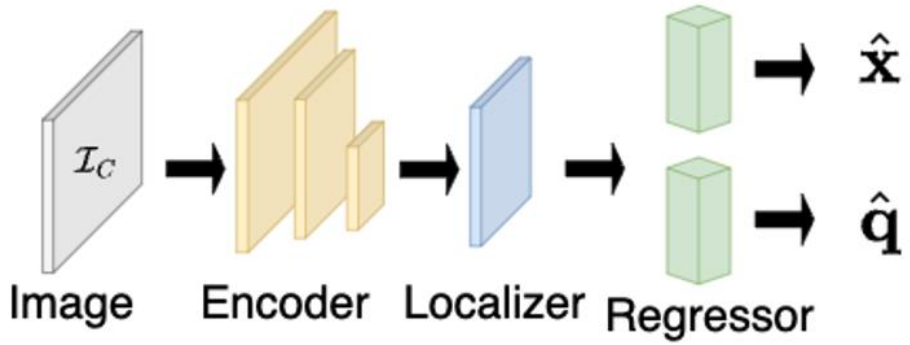
ICLR 2025

Scan & Try !!!

1 香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
2 UNIVERSITY OF OXFORD
3 DARTMOUTH

Absolute pose regression (APR) methods

End-to-end camera pose regression



- Fast inference and low memory footprint.
- Low accuracy and prone to overfit training set [1][2].

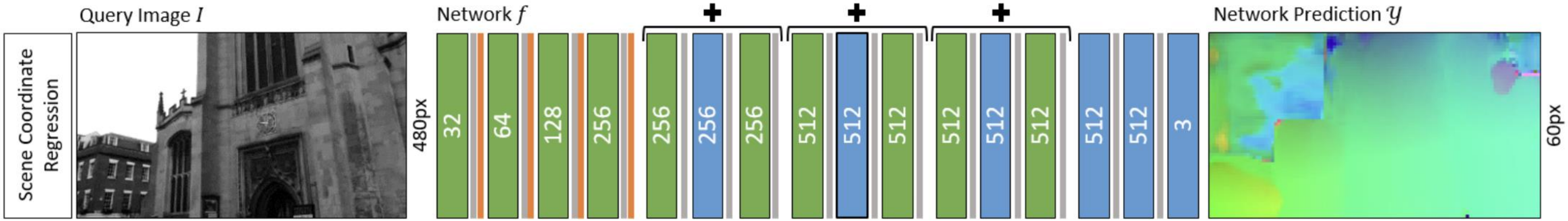
[1] Sattler, Torsten, et al. "Understanding the limitations of cnn-based absolute camera pose regression." CVPR, 2019

[2] Liu, Changkun, et al. "Hr-apr: Apr-agnostic framework with uncertainty estimation and hierarchical refinement for camera relocalisation." IEEE ICRA, 2024

Scene coordinate regression (SCR) methods

End-to-end 2D-3D regression

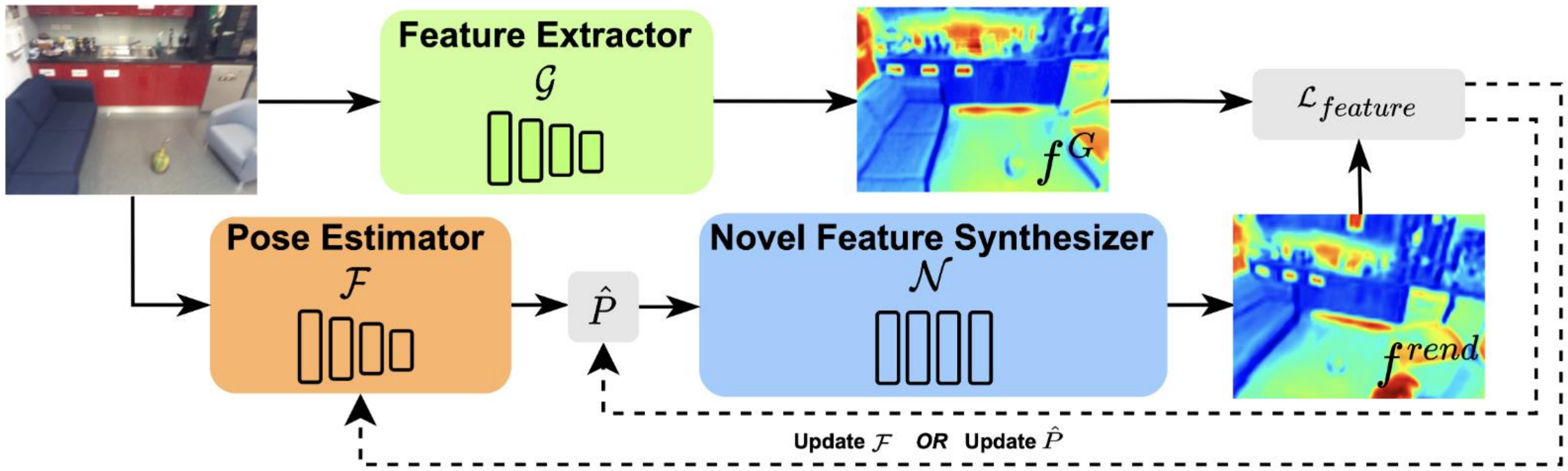
- Accurate in small-scale indoor scenes & relatively fast inference
- Low accuracy in large scenes



Brachmann, Eric, and Carsten Rother. "Visual camera re-localization from RGB and RGB-D images using DSAC." IEEE TPAMI, 2021.

Neural Rendering-based Pose (NRP) Estimation/Refinement

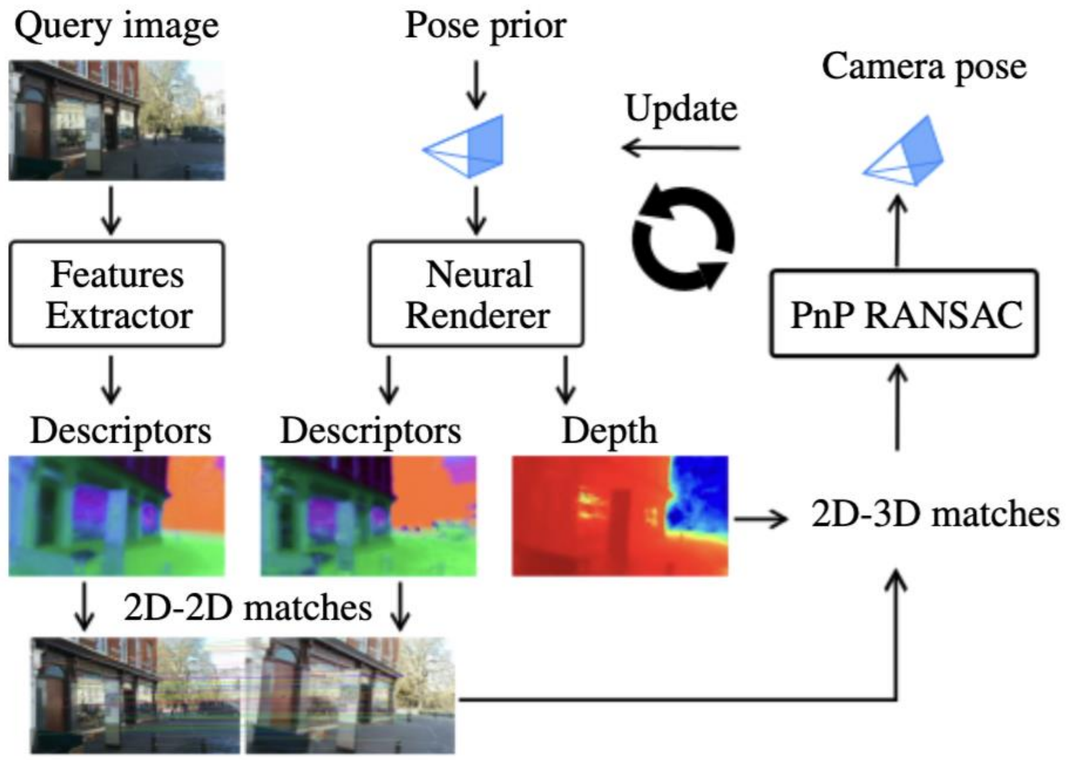
- Can improve APR methods
- Low-efficiency iterative refinement
- Cannot improve SCR methods
- Lower accuracy VS. classical structure-based methods



Chen, Shuai, et al. "Neural refinement for absolute pose regression with feature synthesis." CVPR, 2024.

Neural Rendering-based Pose (NRP) Estimation/Refinement

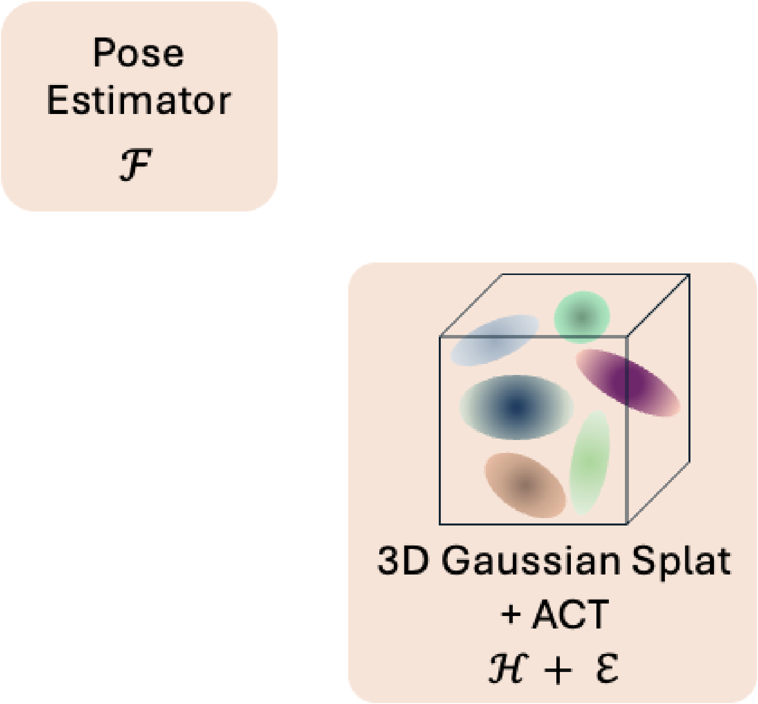
- Utilize geometry information
- Lower accuracy VS. classical structure-based methods
- Need to train scene-specific descriptors



Moreau, Arthur, et al. "Crossfire: Camera relocalization on self-supervised features from an implicit representation." ICCV, 2023.

Our GS-CPR

- Efficient one-shot refinement achieves SOTA accuracy
- Direct RGB matching utilizing MASt3R [1]

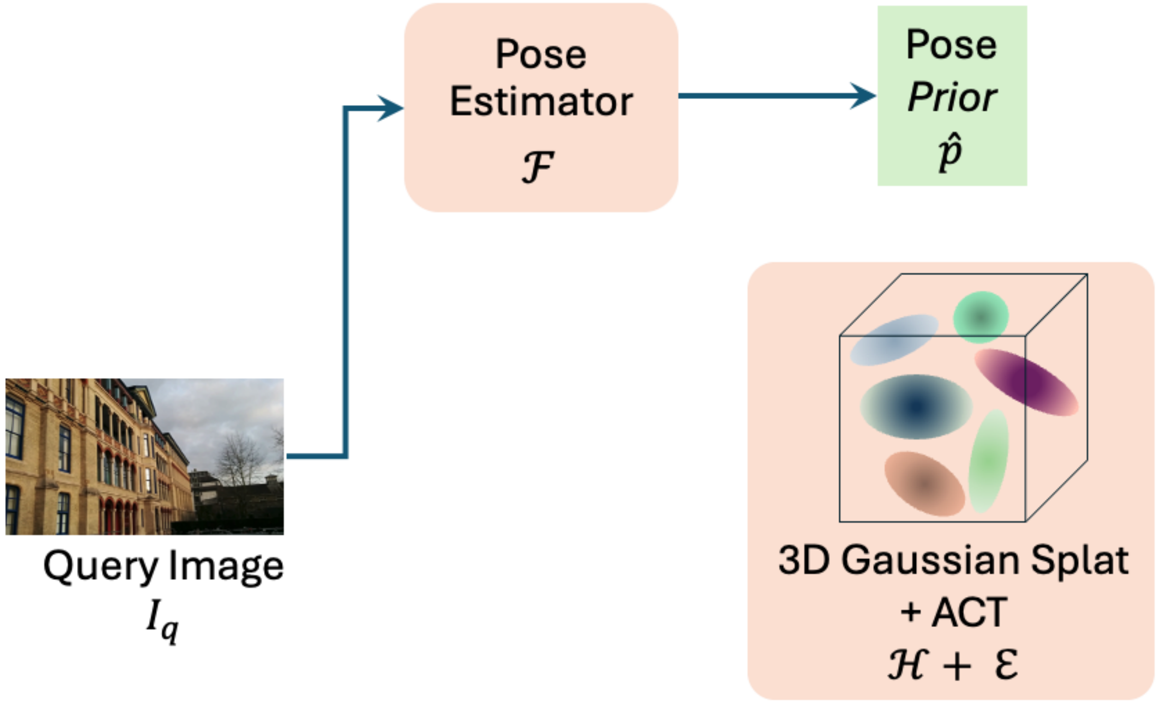


We assume the availability of a pretrained pose estimator (APR/SCR) and a 3DGS model of the scene.

Leroy, Vincent, Yohann Cabon, and Jérôme Revaud. "Grounding image matching in 3d with mast3r." ECCV, 2024.

Our GS-CPR

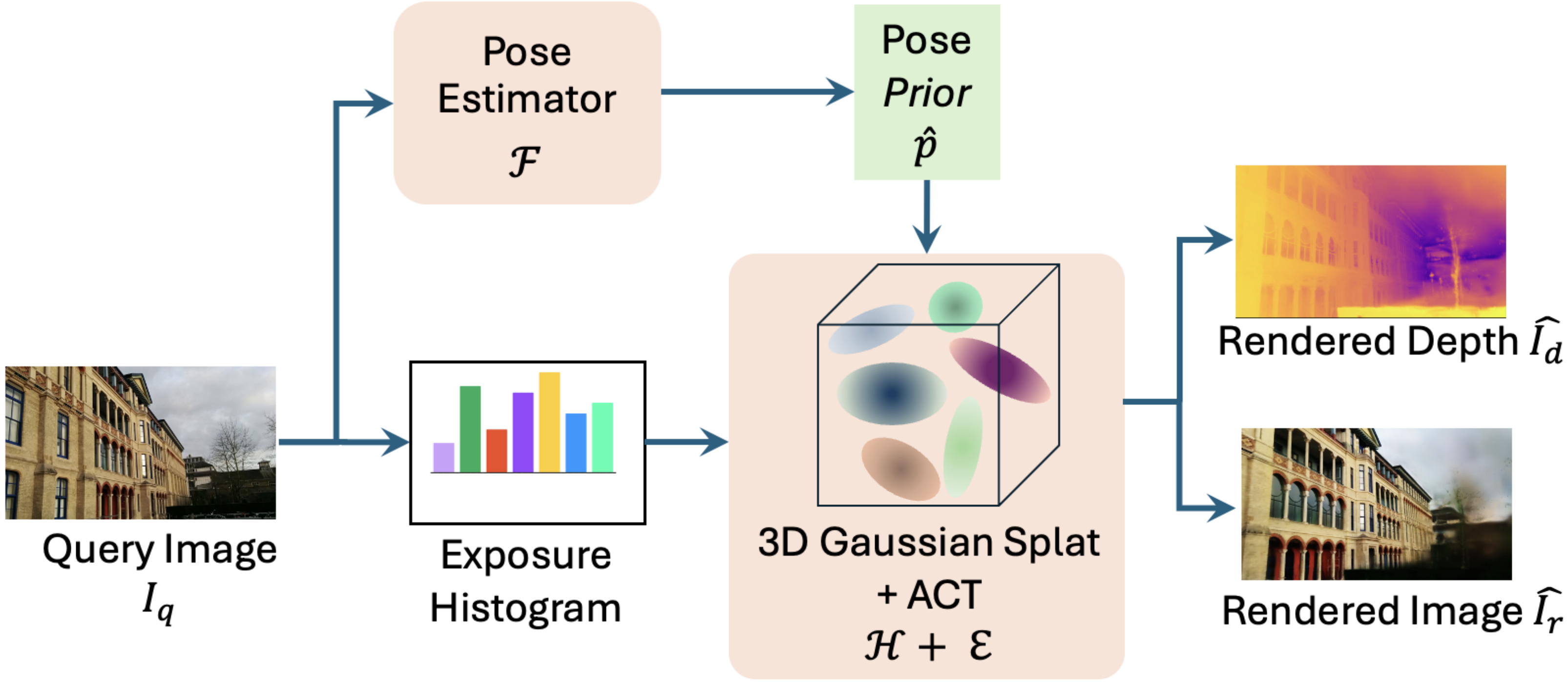
- Efficient one-shot refinement achieves SOTA accuracy
- Direct RGB matching utilizing MASt3R



For a query image, we first obtain an initial estimated pose from the pose estimator (APR/SCR). Our goal is to output a refined pose.

Our GS-CPR

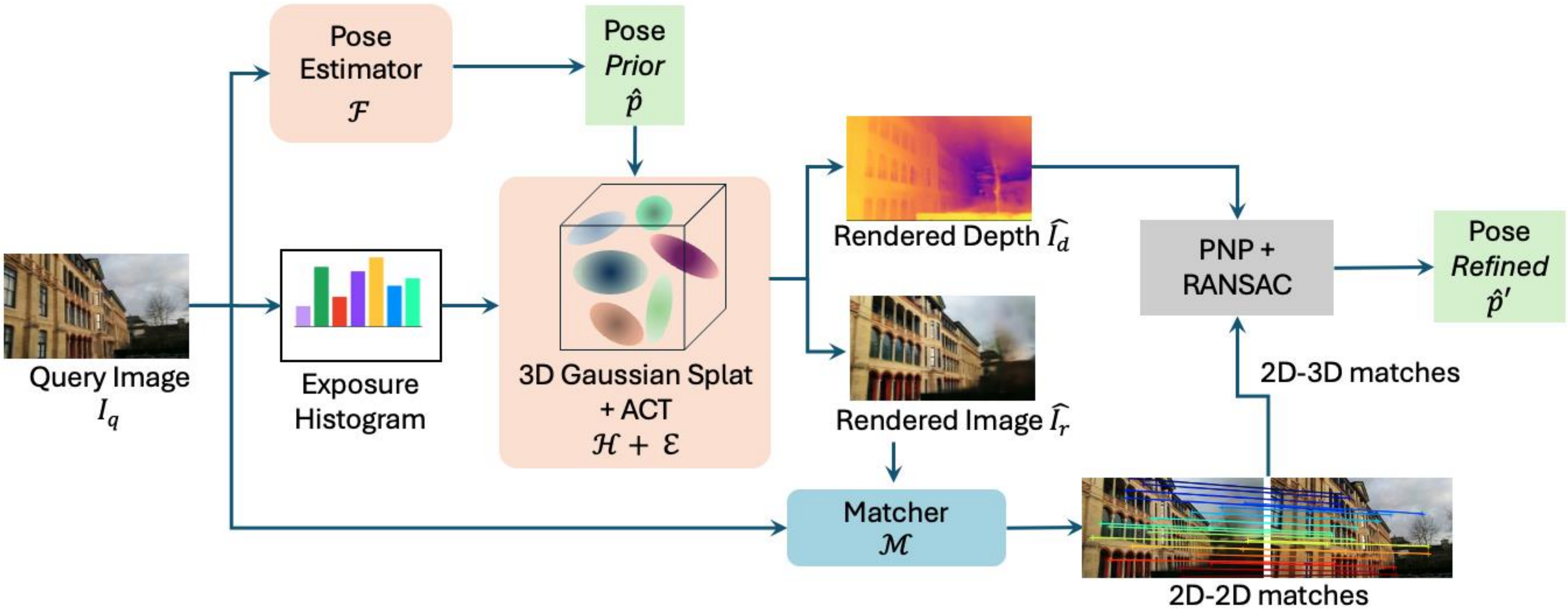
- Efficient one-shot refinement achieves SOTA accuracy
- Direct RGB matching utilizing MASt3R



Given the initial estimated pose, the 3DGS model renders an RGB image and a depth map.

Our GS-CPR

- Efficient one-shot refinement achieves SOTA accuracy
- Direct RGB matching utilizing MASt3R

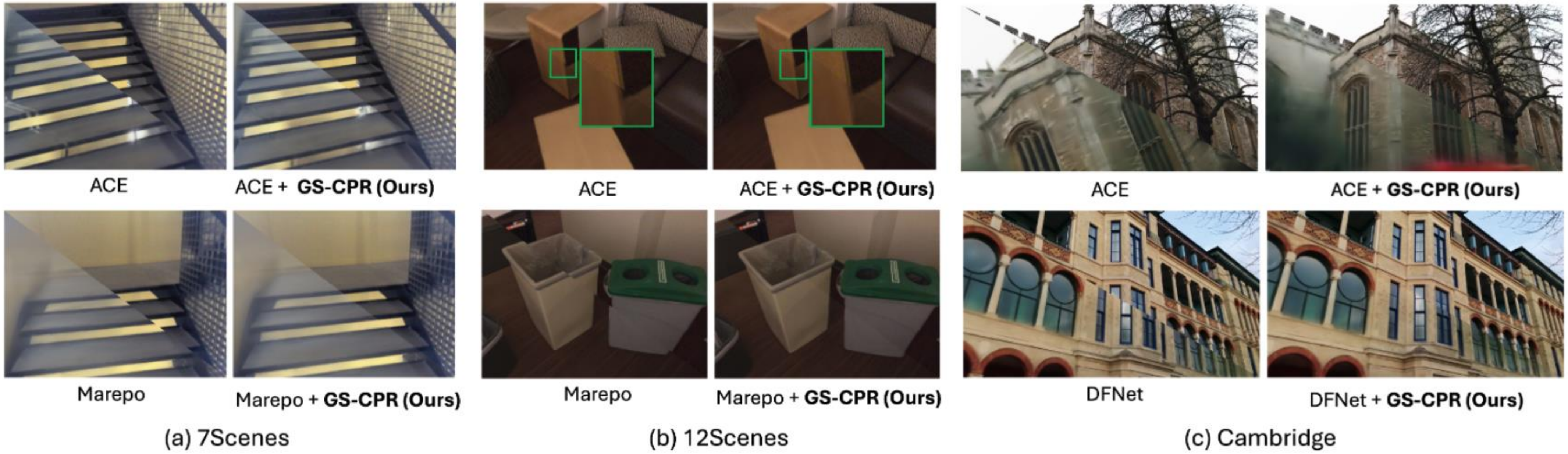


Finally, we obtain the refined pose by feeding these 2D-3D matches into PnP + RANSAC

Visualization

Each figure is divided by a diagonal line, with the bottom left part rendered using the estimated/refined pose and the top right part displaying the ground truth image.

t refers to the refinement step



Visualization on Pose Refinement Before VS. After

ACE (CVPR 2023)

ACE + **GS-CPR (ours)**

Each figure is divided by a diagonal line, with the bottom left part rendered using the estimated/refined pose and the top right part displaying the ground truth image.

Exposure Adaptation and Moving Objects Filtering

(a) Ground Truth (b) Scaffold-GS/PSNR:16.5 dB (c) Ours/PSNR:18.4 dB

GT image

(c) w. Seg Mask + w. ACT (d) w/o. Seg Mask + w/o. ACT

Quantitative Results (Indoor)

7Scenes

	Methods	Avg. ↑ [5cm, 5°]	Avg. ↑ [2cm, 2°]
APR	DFNet	43.1	8.4
	Marepo	84.0	33.7
IR+SfM points	HLoc (SP + SG) (Sarlin et al., 2020; 2019)	95.7	84.5
	DVLAD+R2D2 (Torii et al., 2015; Revaud et al., 2019)	95.7	87.2
SCR	DSAC*	97.8	80.7
	ACE	97.1	83.3
	GLACE	95.6	82.2
NRP	DFNet + NeFeS_{50}	78.3	45.9
	HR-APR	76.4	40.2
	NeRFMatch	78.4	-
	NeRFLoc (Liu et al., 2023)	89.5	-
	DFNet + **GS-CPR (ours)**	94.2	76.5
	Marepo + **GS-CPR (ours)**	99.4	89.6
	ACE + **GS-CPR (ours)**	**100**	**93.1**

12Scenes

Methods	Avg. Err ↓ [cm/°]	Avg. ↑ [5cm, 5°]	Avg. ↑ [2cm, 2°]
Marepo	2.1/1.04	95	50.4
DSAC*	**0.5**/0.25	99.8	96.7
ACE	0.7/0.26	**100**	97.2
GLACE	0.7/0.25	**100**	97.5
Marepo + **GS-CPR (ours)**	0.7/0.28	98.9	90.9
ACE + **GS-CPR (ours)**	**0.5/0.21**	**100**	**98.7**

Quantitative Results (Outdoor)

Cambridge Landmarks

	Methods	Kings	Hospital	Shop	Church	Avg. ↓ [cm/°]
IR + SfM points	HLoc (SP+SG) (k=1)	13/0.22	18/0.38	6/0.25	9/0.28	12/0.28
	HLoc (SP+SG) (k=10)	11/0.2	15/0.31	4/0.21	7/0.22	9/0.24
APR	PoseNet	93/2.73	224/7.88	147/6.62	237/5.94	175/5.79
	MS-Transformer	85/1.45	175/2.43	88/3.20	166/4.12	129/2.80
	LENS (Moreau et al., 2022)	33/0.5	44/0.9	27/1.6	53/1.6	39/1.15
	DFNet	73/2.37	200/2.98	67/2.21	137/4.02	119/2.90
	PMNet (Lin et al., 2024)	68/1.97	103/1.31	58/2.10	133/3.73	90/2.27
SCR	ACE	29/0.38	31/0.61	5/0.3	19/0.6	21/0.47
	GLACE[1]	20/0.32	20/0.41	5/0.22	9/0.3	14/0.32
NRP	FQN-MN	28/0.4	54/0.8	13/0.6	58/2	38/1
	CrossFire	47/0.7	43/0.7	20/1.2	39/1.4	37/1
	DFNet + NeFeS_{30}[2]	37/0.64	98/1.61	17/0.60	42/1.38	49/1.06
	DFNet + NeFeS_{50}	37/0.54	52/0.88	15/0.53	37/1.14	35/0.77
	HR-APR	36/0.58	53/0.89	13/0.51	38/1.16	35/0.78
	MCLoc	31/0.42	39/0.73	12/0.45	26/0.8	27/0.6
	DFNet + **GS-CPR (ours)**	23/0.32	42/0.74	10/0.36	27/0.62	26/0.51
	ACE + **GS-CPR (ours)**	**20/0.29**	**21/0.40**	**5/0.24**	**13/0.40**	**15/0.33**

[1] We report the accuracy based on official open-source models (Wang et al., 2024a).
[2] Results of DFNet + NeFeS_{30} taken from Liu et al. (2024a).

Thanks for watching

Project Page: https://xrim-lab.github.io/GS-CPR/

ICLR 2025

Scan & Try !!!