



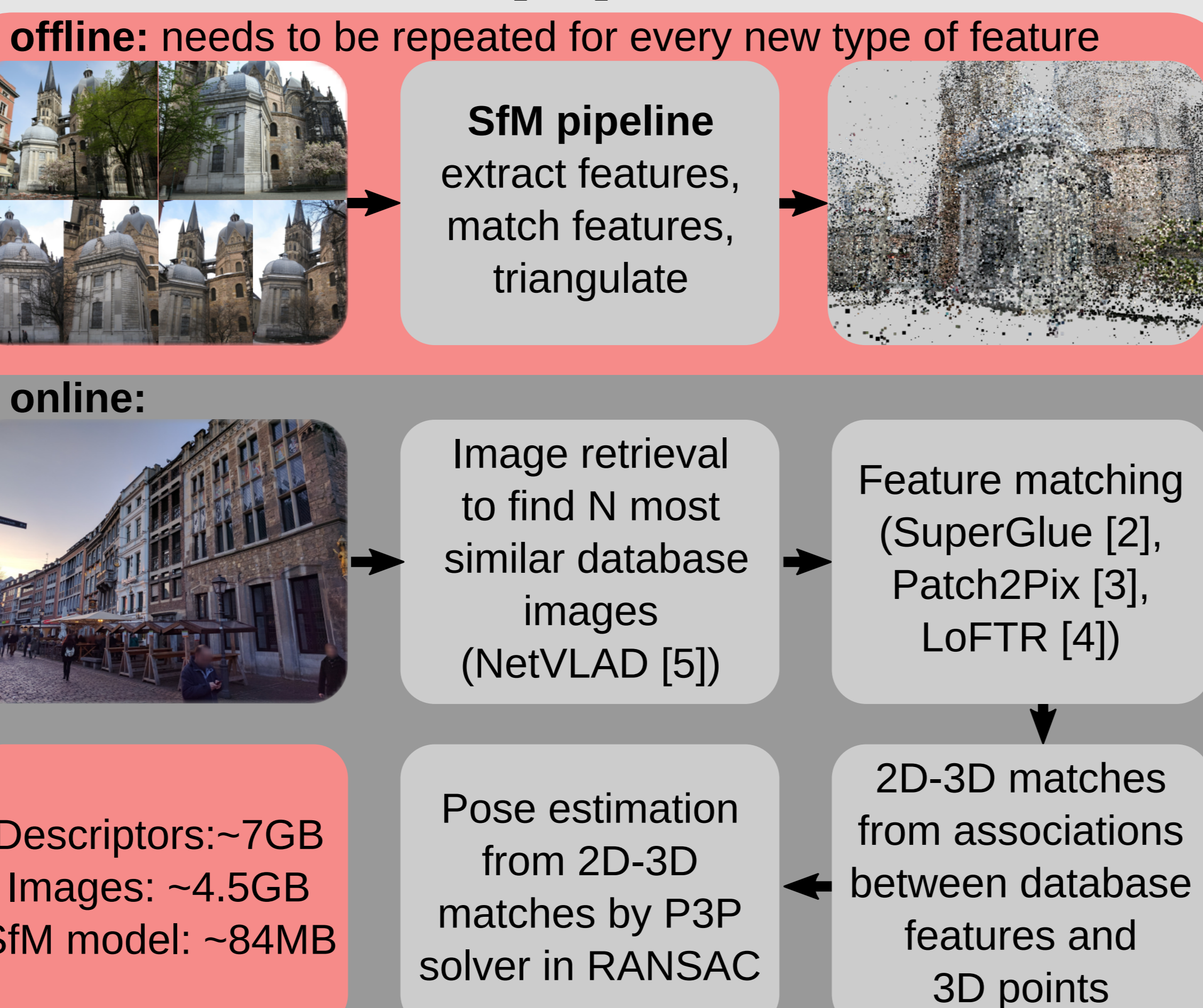
## Motivation

SfM-based representations (commonly used for localization) are specialized  
→ explore mesh-based scene representation that can be used for other tasks as well

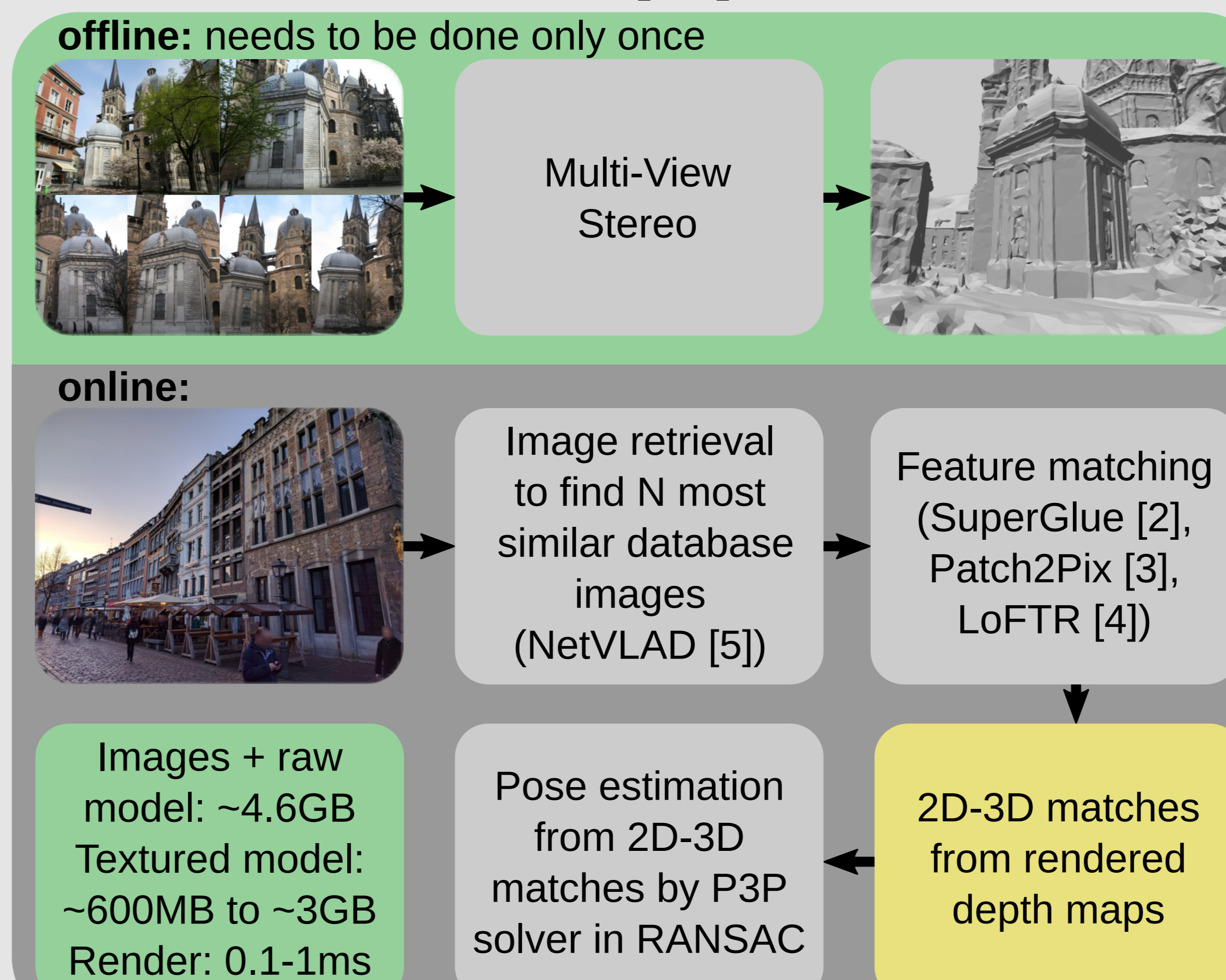
## Contributions

1. Design and evaluate localization pipeline based on meshes
  2. State-of-the-art results for a simple variant
  3. Interesting and promising results for non-photorealist renderings of meshes
  4. Detailed ablation studies with state-of-the-art features
- Code and data available at: [github.com/tsattler/meshloc\\_release](https://github.com/tsattler/meshloc_release)

## SfM-based pipeline



## Mesh-based pipeline

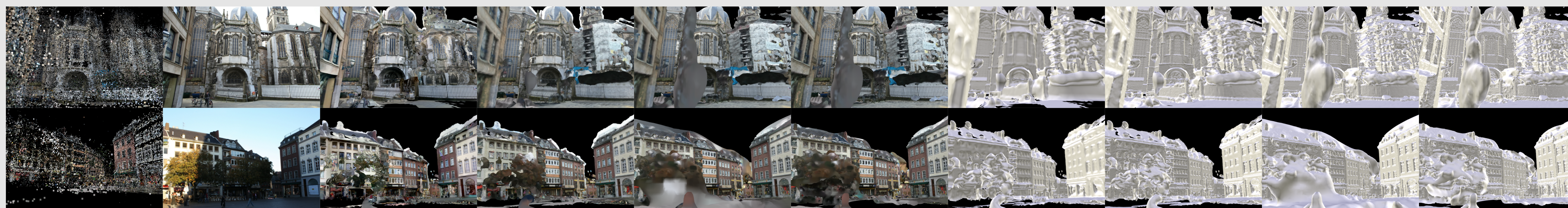


## Main insights

- Simple pipeline, yet **state-of-the-art** results with real images
- **Close-to-state-of-the-art** results with rendered images
- Promising results when localizing against purely geometric representations (untextured meshes), even without re-training features
- **Dense** representations can be **more compact** than the sparse ones

## Results

**Aachen Day-Night v1.1 [1]**  
- nighttime queries only



% queries localized within  
(0.25m, 2°) / (0.5m, 5°) / (5m, 10°)

	SfM point cloud	original images	AC13C textured 645 MB	AC13 colored 617 MB	AC14 colored 1,234 MB	AC15 colored 2,538 MB	AC13C tricolor 47 MB	AC13 tricolor 558 MB	AC14 tricolor 1,116 MB	AC15 tricolor 2,538 MB
800 px										
	-	74.3 / 93.2 / 99.0	72.3 / 91.1 / 99.0	68.1 / 90.1 / 97.4	74.9 / 92.1 / 99.5	75.4 / 89.5 / 98.4	7.3 / 23.0 / 53.9	22.0 / 50.8 / 74.3	33.0 / 65.4 / 79.1	37.2 / 60.7 / 77.5
	-	78.5 / 93.2 / 99.5	74.3 / 91.1 / 99.0	72.3 / 87.4 / 94.2	76.4 / 87.4 / 95.3	78.0 / 89.0 / 95.8	2.6 / 13.6 / 41.9	15.7 / 36.1 / 63.4	27.7 / 62.3 / 78.5	34.0 / 60.7 / 77.0
	-	73.8 / 89.5 / 95.3	66.5 / 87.4 / 94.2	64.4 / 81.7 / 90.1	63.9 / 84.3 / 93.2	65.4 / 84.8 / 94.2	9.4 / 28.3 / 63.4	31.9 / 58.1 / 81.2	38.2 / 67.5 / 84.3	38.2 / 62.8 / 82.2
	-	74.9 / 92.1 / 99.5	70.7 / 90.6 / 99.5	71.7 / 91.1 / 97.9	72.3 / 92.1 / 96.9	72.8 / 92.1 / 98.4	9.4 / 25.1 / 57.6	23.0 / 55.0 / 78.5	39.3 / 68.6 / 80.6	40.3 / 66.0 / 80.1
full res.										
	77.0 / 90.6 / 100.0	77.0 / 92.1 / 99.0	77.0 / 92.1 / 99.5	77.0 / 92.1 / 99.0	75.4 / 91.1 / 99.0	76.4 / 92.1 / 99.0	-	-	-	-
	78.5 / 90.6 / 99.0	-	-	-	-	-	-	-	-	-
	72.3 / 88.5 / 97.9	74.3 / 90.1 / 96.3	74.3 / 90.1 / 96.9	74.3 / 90.1 / 96.3	71.2 / 86.9 / 95.3	72.3 / 88.0 / 96.9	-	-	-	-
	78.0 / 90.6 / 99.0	74.3 / 92.1 / 99.5	73.3 / 92.1 / 99.5	74.3 / 92.1 / 99.5	73.3 / 91.1 / 99.5	74.3 / 92.7 / 99.5	-	-	-	-

Example of LoFTR [4] matches on AC15 mesh:



## 12 Scenes dataset [6]

% queries localized within  
(5cm, 5°) / (7cm, 7°) / (10cm, 10°)

	Scenes with well-aligned RGB images and scene geometry			Scenes with visible misalignment between RGB images and scene geometry		
	apt2/kitchen	office1/gates381	office1/manolis	apt1/living	office1/lounge	office2/5b
MeshLoc (SG [2]) - real	97.6 / 99.0 / 100.0	93.4 / 98.7 / 99.7	95.5 / 99.9 / 100.0	88.2 / 100.0 / 100.0	87.8 / 100.0 / 100.0	88.4 / 99.8 / 100.0
MeshLoc (SG [2]) - color	92.9 / 98.6 / 100.0	78.3 / 97.6 / 99.9	75.7 / 98.0 / 100.0	31.0 / 83.0 / 99.8	60.2 / 95.1 / 99.7	60.0 / 83.2 / 97.8
MeshLoc (SG [2]) - tricolor	21.9 / 56.2 / 76.7	15.3 / 36.2 / 52.2	23.3 / 64.6 / 84.8	15.8 / 40.4 / 60.6	12.5 / 28.1 / 47.1	14.6 / 29.9 / 47.9
Active Search [7]	100.0 / 100.0 / 100.0	98.2 / 99.0 / 99.2	100.0 / 100.0 / 100.0	99.8 / 99.8 / 100.0	100.0 / 100.0 / 100.0	100.0 / 100.0 / 100.0
HLoc [8]	100.0 / 100.0 / 100.0	99.1 / 99.9 / 100.0	100.0 / 100.0 / 100.0	100.0 / 100.0 / 100.0	100.0 / 100.0 / 100.0	100.0 / 100.0 / 100.0

[1] T. Sattler, T. Weyand, B. Leibe, L. Kobbelt: Image Retrieval for Image-Based Localization Revisited. BMVC (2012).  
[2] P.-E. Sarlin, D. DeTone, T. Malisiewicz, A. Rabinovich. "SuperGlue: Learning Feature Matching With Graph Neural Networks." CVPR. 2020.  
[3] Q. Zhou, T. Sattler, L. Leal-Taixé. "Patch2Pix: Epipolar-Guided Pixel-Level Correspondences." CVPR. 2021.  
[4] J. Sun, Z. Shen, Y. Wang, H. Bao, X. Zhou. "LoFTR: Detector-Free Local Feature Matching with Transformers." CVPR. 2021.  
[5] R. Arandjelović, P. Gronat, A. Torii, T. Pajdla, J. Sivic. "NetVLAD: CNN Architecture for Weakly Supervised Place Recognition." PAMI. 2018.  
[6] J. P. C. Valentin, A. Dai, M. Nießner, P. Kohli, P. H. S. Torr, S. Izadi, C. Keskin. "Learning to Navigate the Energy Landscape." 3DV. 2016.  
[7] T. Sattler, B. Leibe, L. P. Kobbelt. "Efficient & Effective Prioritized Matching for Large-Scale Image-Based Localization." PAMI. 2017.  
[8] P.-E. Sarlin, C. Cadena, R. Y. Siegwart, M. Dymczyk. "From Coarse to Fine: Robust Hierarchical Localization at Large Scale." CVPR. 2019.

This work was supported by the EU Horizon 2020 project RICAIP (grant agreement No. 857306), the European Regional Development Fund under project IMPACT (No. CZ.02.1.01/0.0/0.0/15\_003/0000468), a Meta Reality Labs research award under project call 'Benchmarking City-Scale 3D Map Making with Mappillary Metropolis', the Grant Agency of the Czech Technical University in Prague (No. SGS21/119/OHK3/2T/13), the OP VVV funded project CZ.02.1.01/0.0/0.0/16\_019/0000765 "Research Center for Informatics", and the ERC-CZ grant MSMT LL1901.