

FaVoR: Features via Voxel Rendering for Camera Relocalization

Vincenzo Polizzi, Marco Cannici, Davide Scaramuzza, Jonathan Kelly



Overview

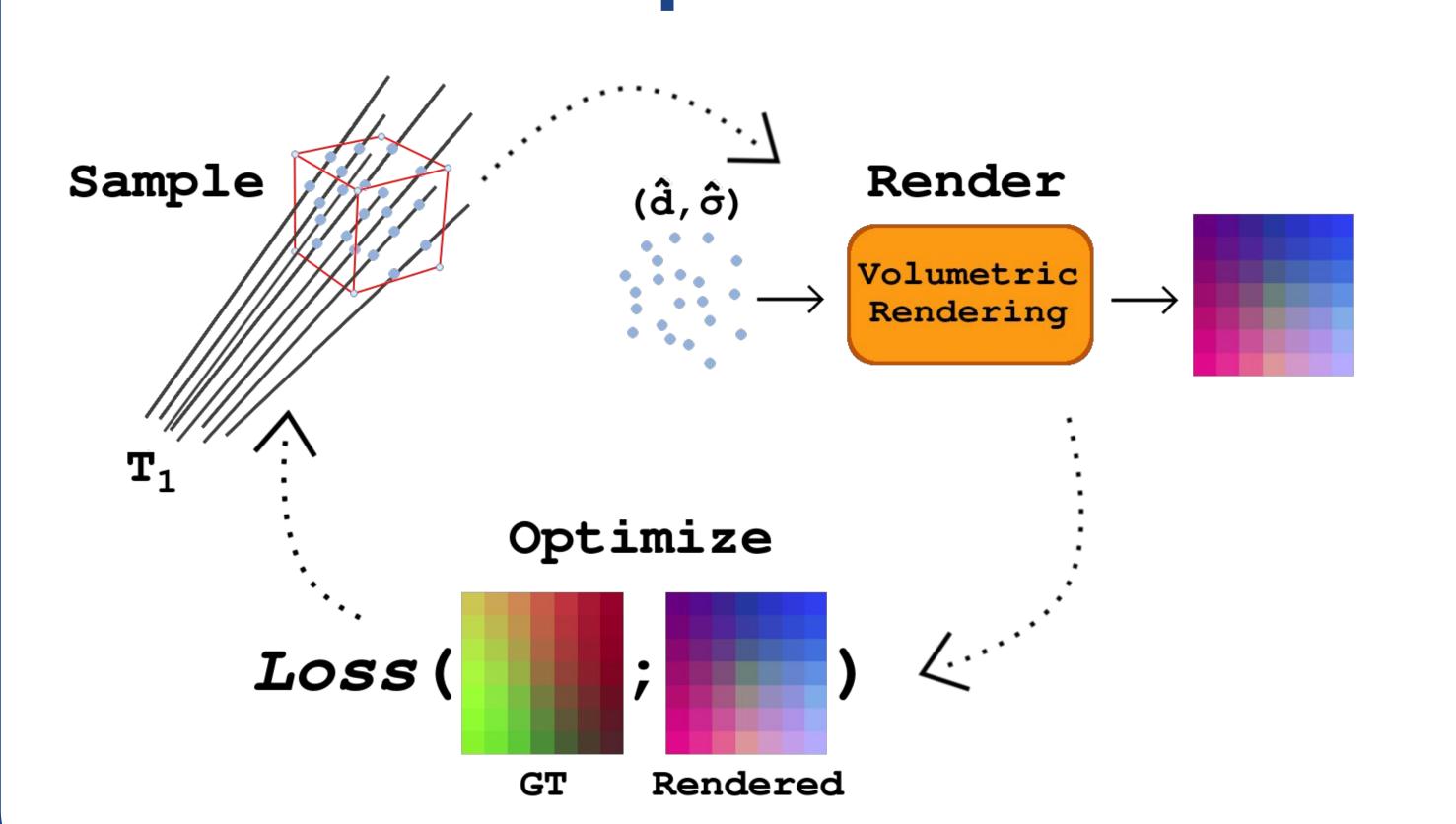
Goal: Overcoming feature descriptor variance by explicitly modeling their view-dependent appearance variations

Method	Sparse	Dense	Low Mem. Req.	High Dim. Req.
Previous				
Our		X		

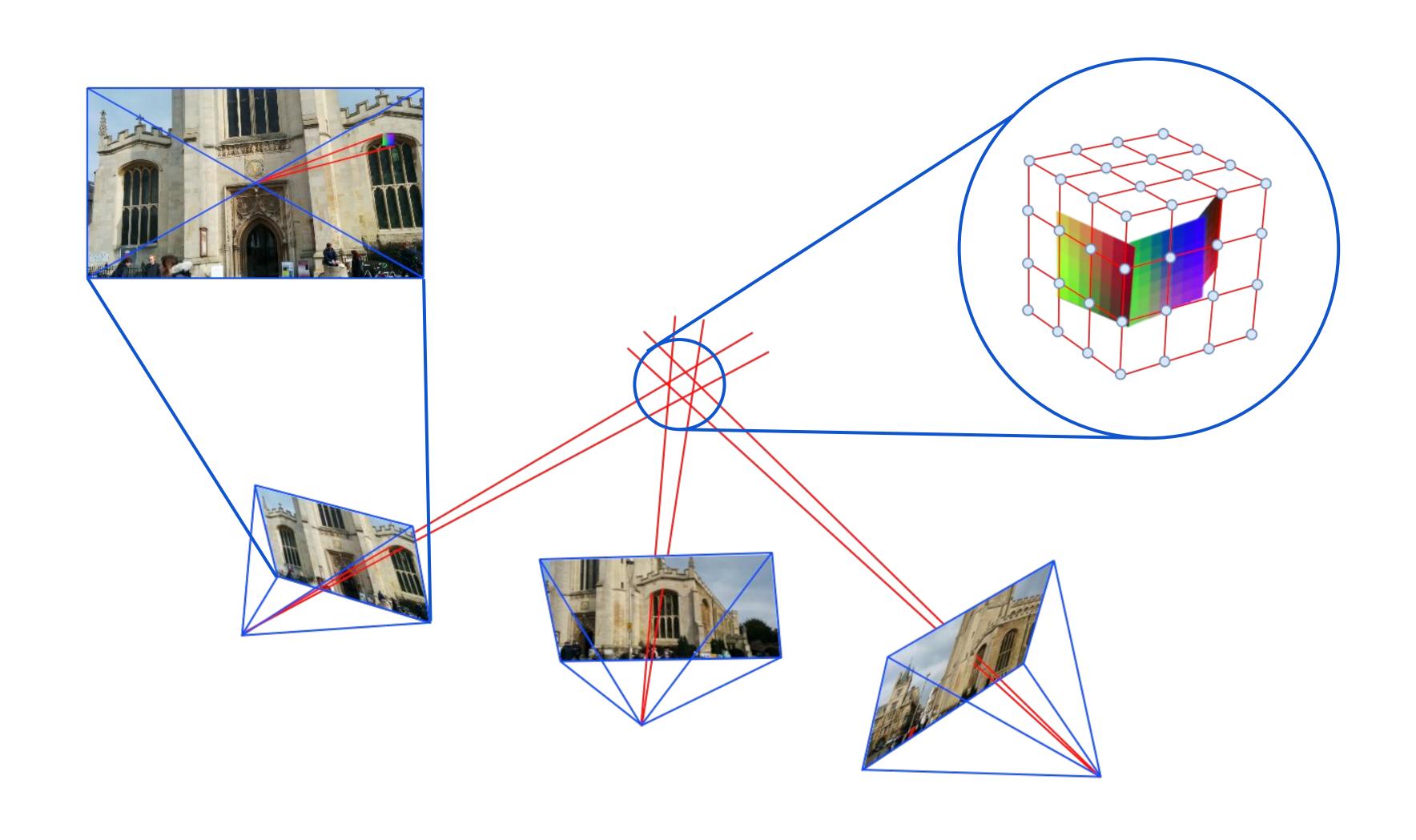
Triangulation



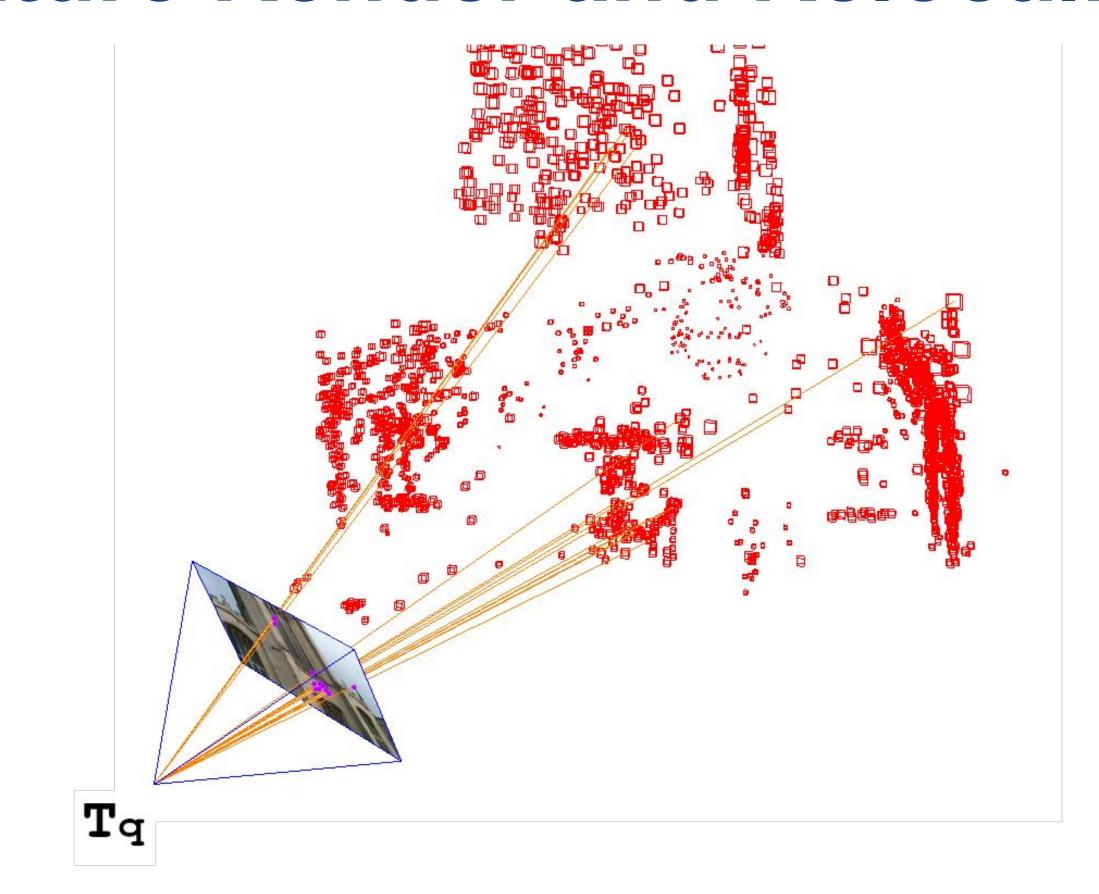
Voxel Optimization



Patch Projection and Voxel Creation



Feature Render and Relocalization

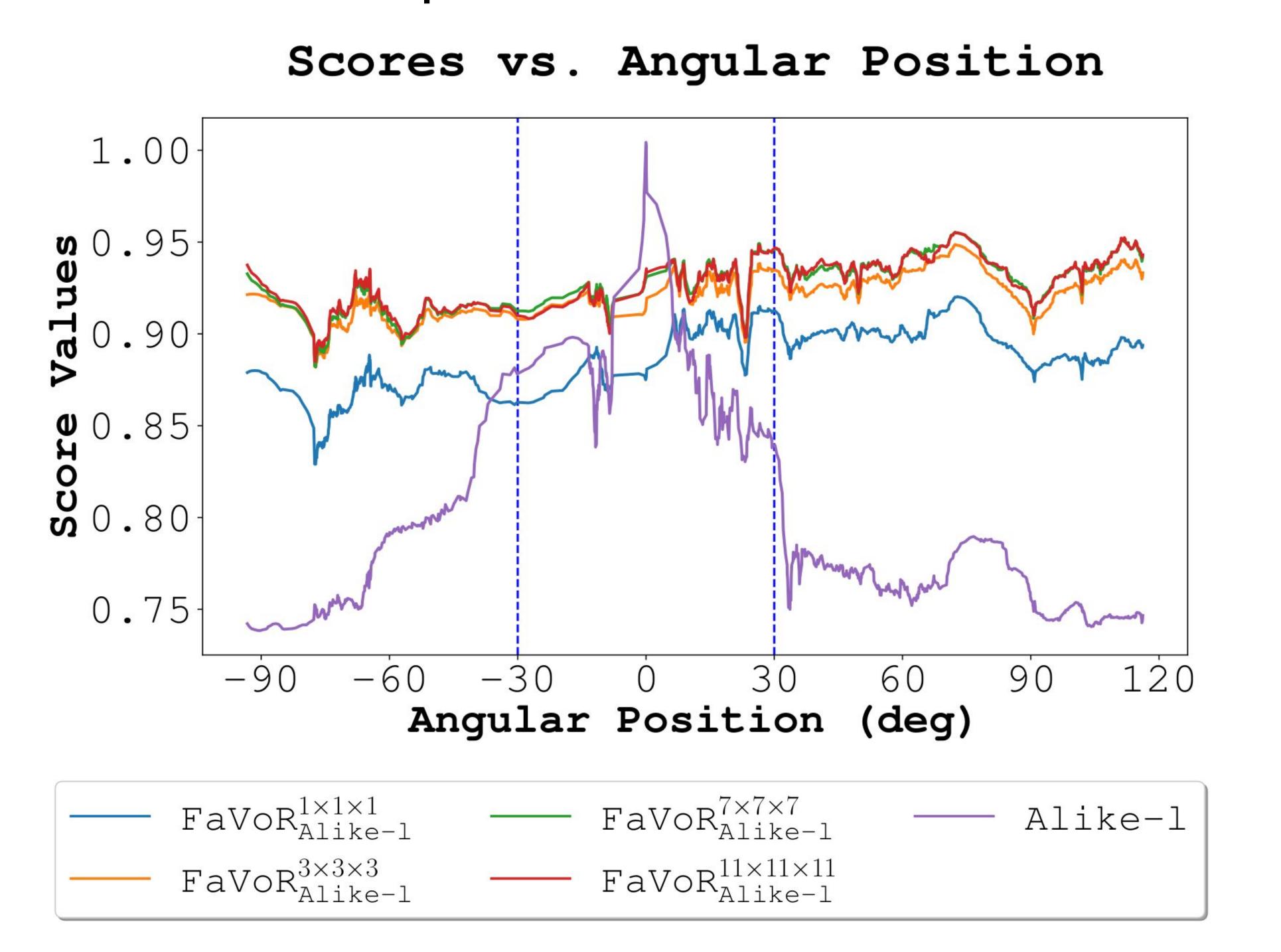


Results

Median pose error (cm, deg) at different iterations of the Render+PnP-RANSAC scheme

Dataset	Method	Median Pose Error	
7-Scenes	FQN-MN	27.4, 7.4	
	CROSSFIRE	4.4, 1.4	
	NeRF-Loc	2.3, 1.3	
	Our _{Alike-I}	1.3, 0.4	
		w Court	w/o Court
	FQN-MN	881, 8.6	38, 1.0
Cambridge	CROSSFIRE	_	37, 1.0
Landmarks	NeRF-Loc	13, 0.2	10, 0.3
	Our _{Alike-I}	15, 0.3	12, 0.3

FaVoR's view invariance across different voxel resolutions, compared to standard methods

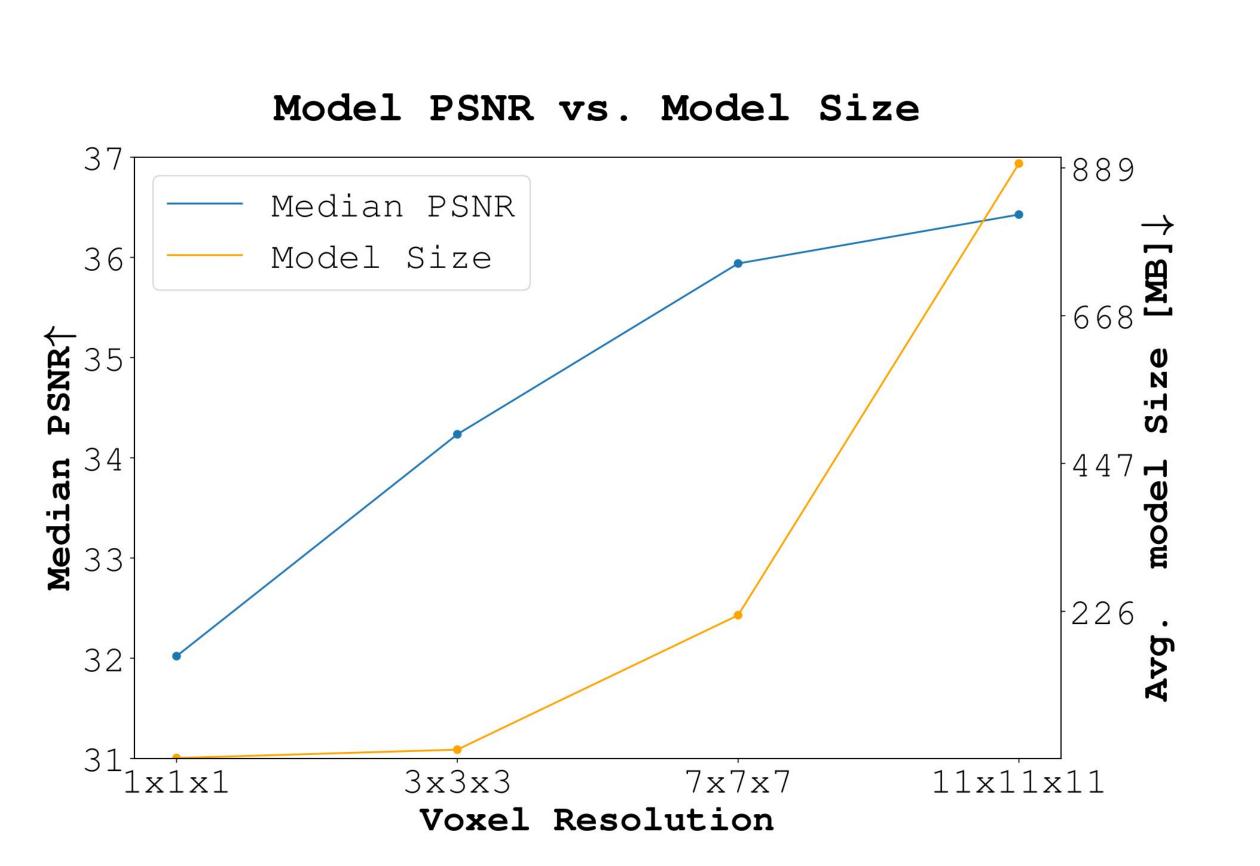


Pose Initialization

Localization errors (cm, deg) for different initial poses: DenseVLAD for *Retrieval* and the first training set pose for *Constant*

Method	Prior Err.	Iter 1	Iter 2	Iter 3
Retrieval	21.9, 12.13	0.8, 0.21	0.7, 0.19	0.7, 0.18
Constant	147.6, 29.94	1.0, 0.28	0.7, 0.19	0.7, 0.18

Qualitative Evaluation



Descriptor patch optimized with L2, cosine, and TV losses vs. ground truth

Ground
Truth
Rendered L2 Norm

