Can You Trust Your Pose? Confidence Estimation in Visual Localization

Luca Ferranti[1,2], Xiaotian Li[2], Jani Boutellier[1], Juho Kannala[2]

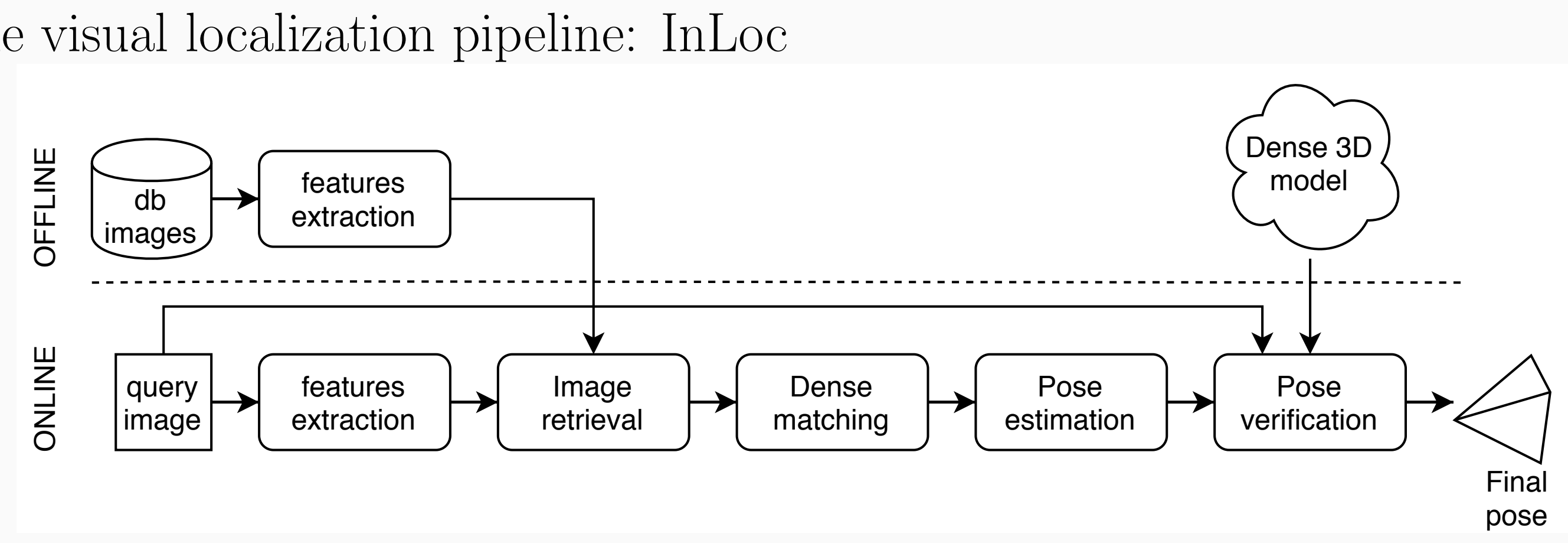
[1] University of Vaasa, Vaasa Finland
[2] Aalto University, Espoo, Finland

Introduction

Camera pose estimation in large-scale environments is still an open question and, despite recent promising results, it may still fail in some situations. The research so far has focused on improving subcomponents of estimation pipelines, to achieve more accurate poses. However, there is no guarantee for the result to be correct, even though the correctness of pose estimation is critically important in several visual localization applications, such as in autonomous navigation. In this paper we bring to attention a novel research question, pose confidence estimation, where we aim at quantifying how reliable the visually estimated pose is. We develop a novel confidence measure to fulfill this task and show that it can be flexibly applied to different datasets, indoor or outdoor, and for various visual localization pipelines. We also show that the proposed techniques can be used to accomplish a secondary goal: improving the accuracy of existing pose estimation pipelines. Finally, the proposed approach is computationally light-weight and adds only a negligible increase to the computational effort of pose estimation.

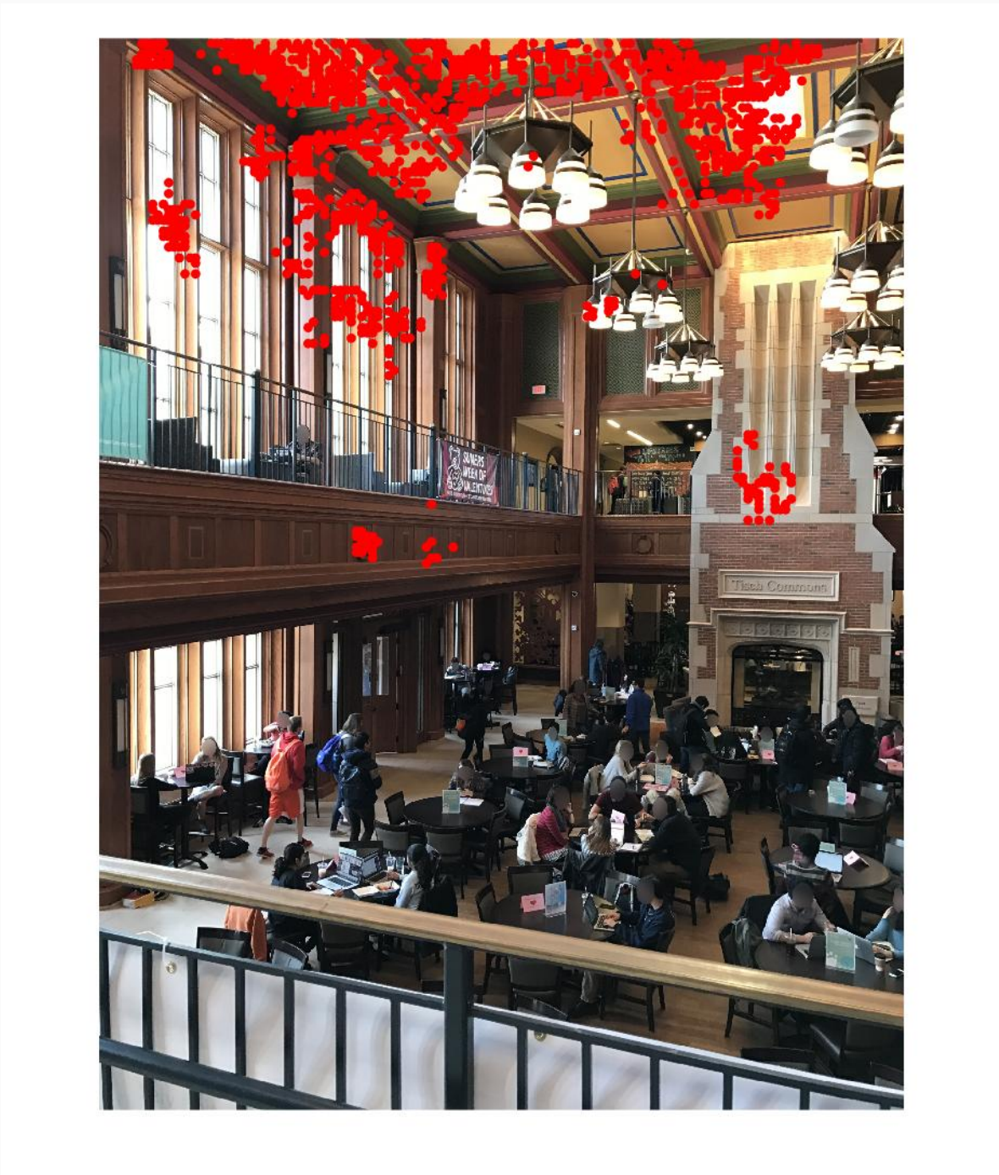
Visual Localization

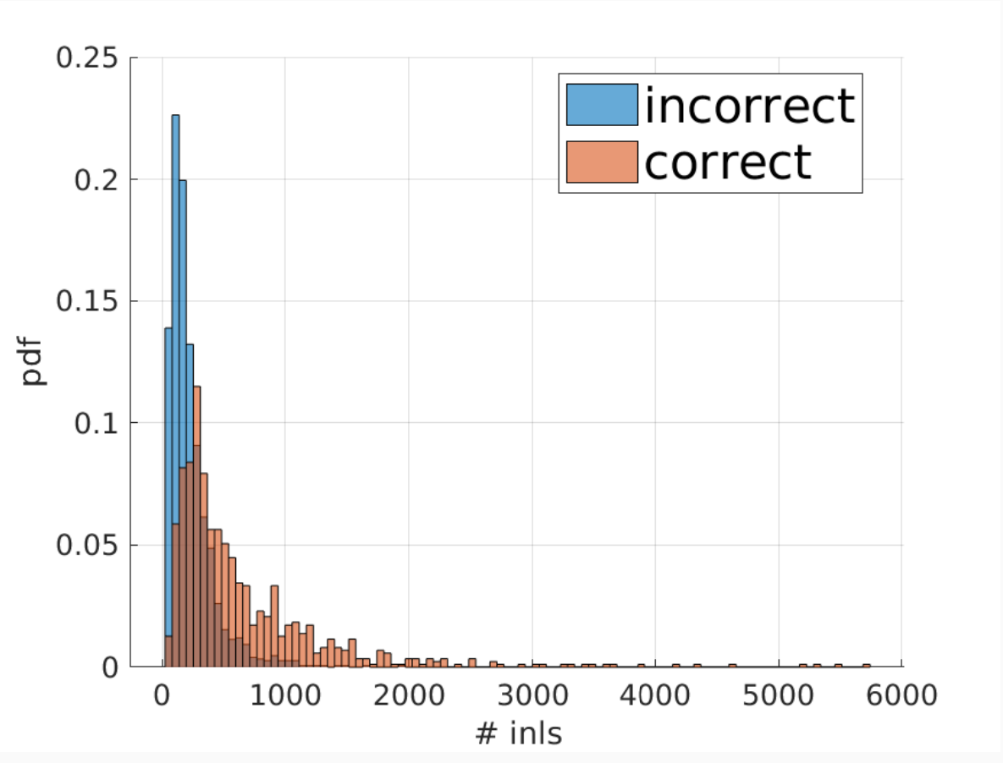
- Visual Localization aims at computing camera position and orientation (shortly referred as camera pose) from a 2D picture.
- example visual localization pipeline: InLoc



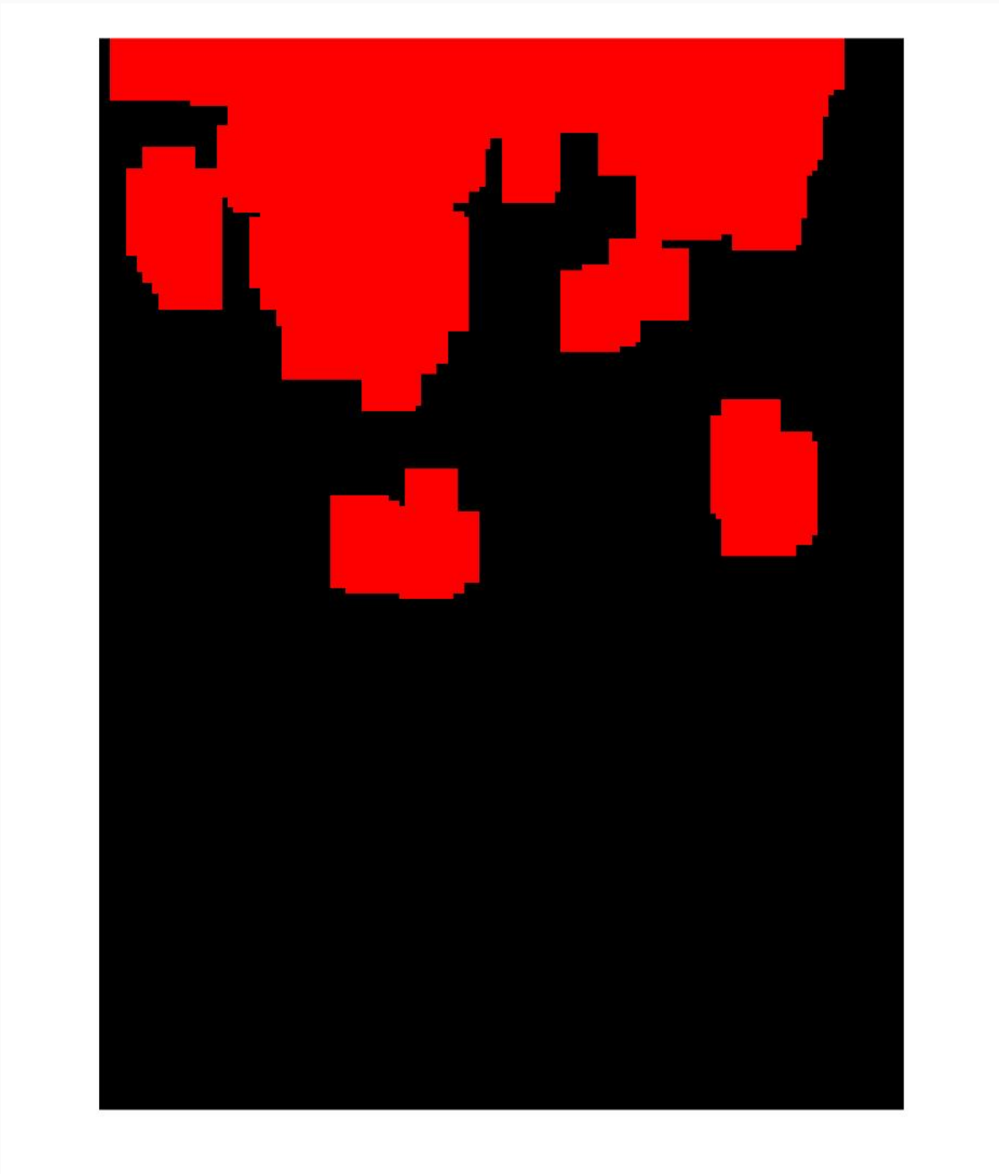
- **Image Retrieval:** 100 most similar database images are retrieved
- **Dense Matching:** for each query-database image pair, 2D-2D points correspondences are formed. These induce 2D-3D correspondences. Only 10 best pairs are kept.
- **Pose Estimation:** for each pair, camera pose is computed suing P3P-RANSAC, obtaining 10 candidate poses.
- **Pose Verification:** 10 candidate poses are reranked to choose the best candidate.

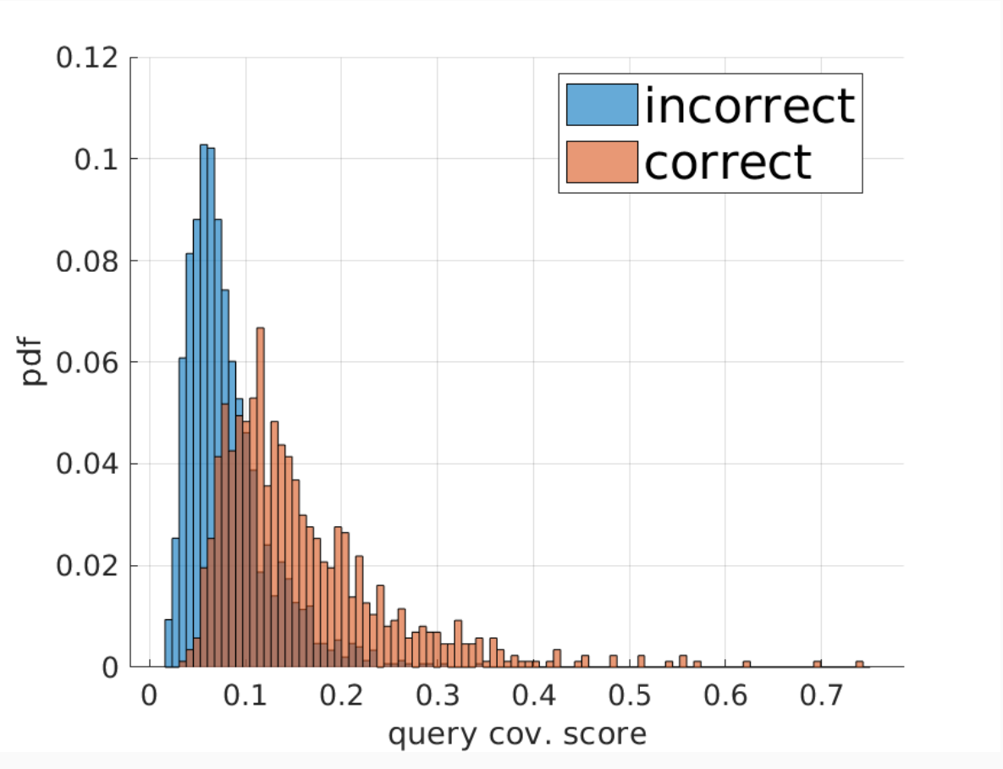
Proposed Confidence Measure





- The inliers count alone is not very robust and several factors, such as repetitive patterns, occlusions and moving objects may lead to incorrect poses with a high inliers count
- Instead, we also compute a coverage map of the inliers, i.e. the portion of the image covered by the inliers and from it a coverage score.

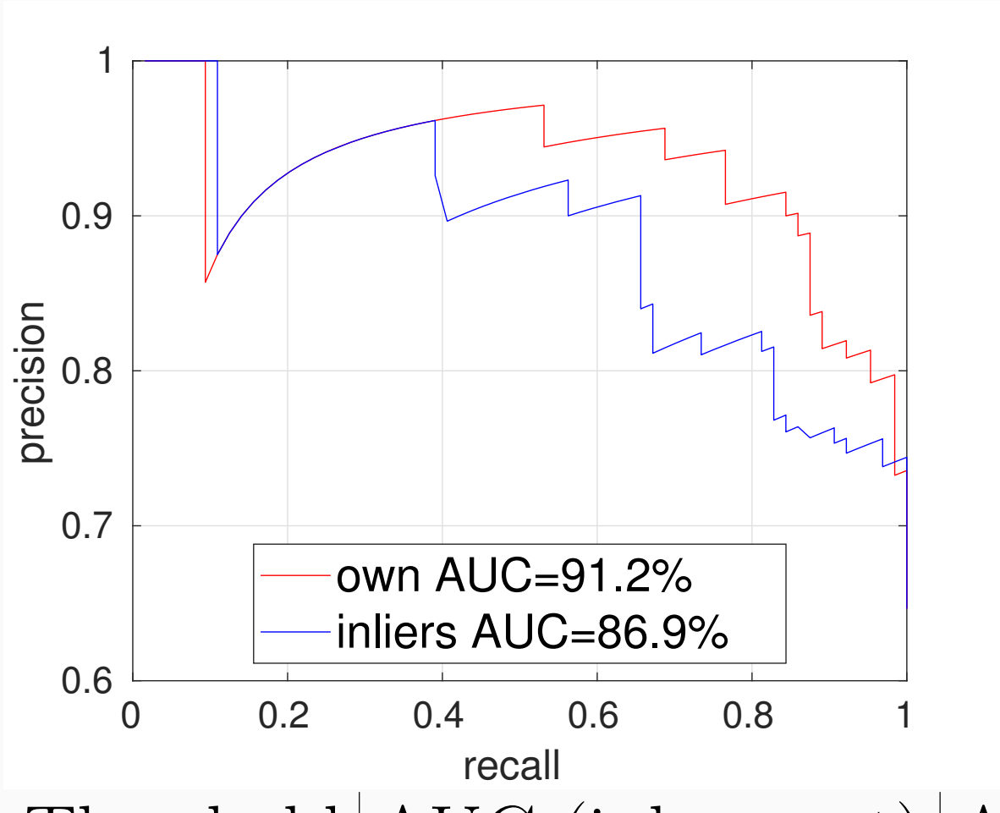


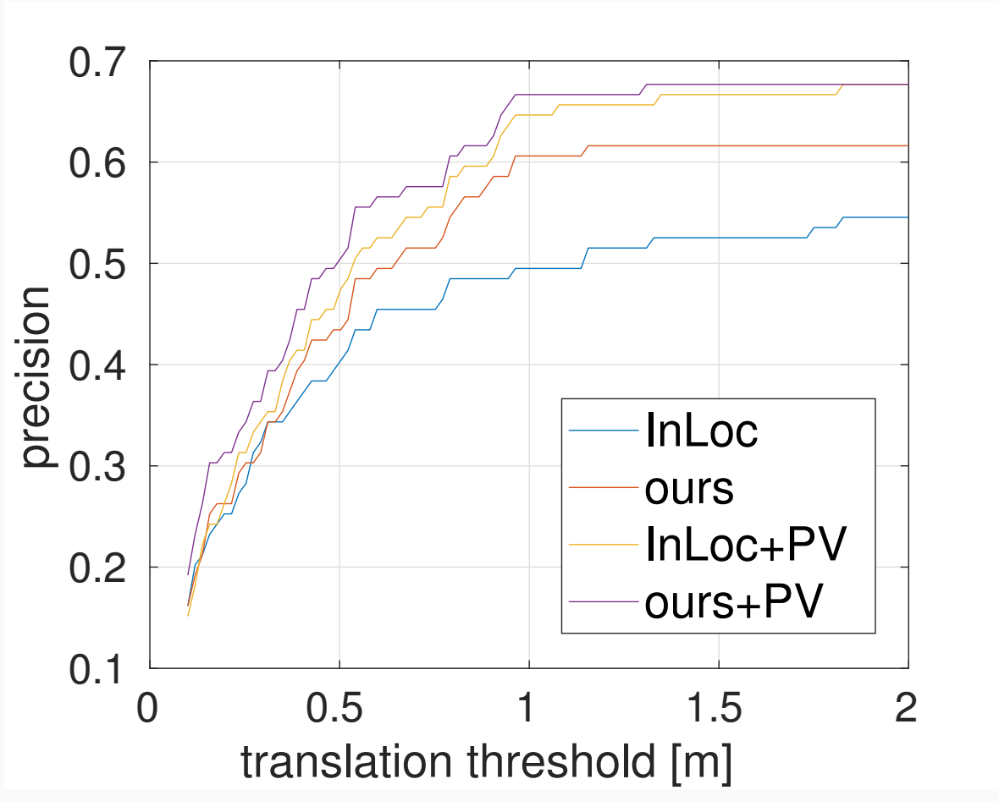


- For a query-database image pair we have three scalar parameters: inliers count x_1, query image coverage score x_2 and database image coverage score x_3.
- To compute our final confidence measure, we use logistic regression, i.e. $\gamma = \text{logsig}\left(b + \sum_{i=1}^{3} w_i x_i\right)$

Results on InLoc

- Our metric gave a more robust precision-recall curve compared to the inliers count
- The metric outperformed the inliers count also when used at different error thresholds than the training error threshold
- Our model was also brought inside InLoc pipeline to choose the best candidate pose of a single query image. Choosing the best candidate pose with our method gave an improvement to the baseline accuracy.

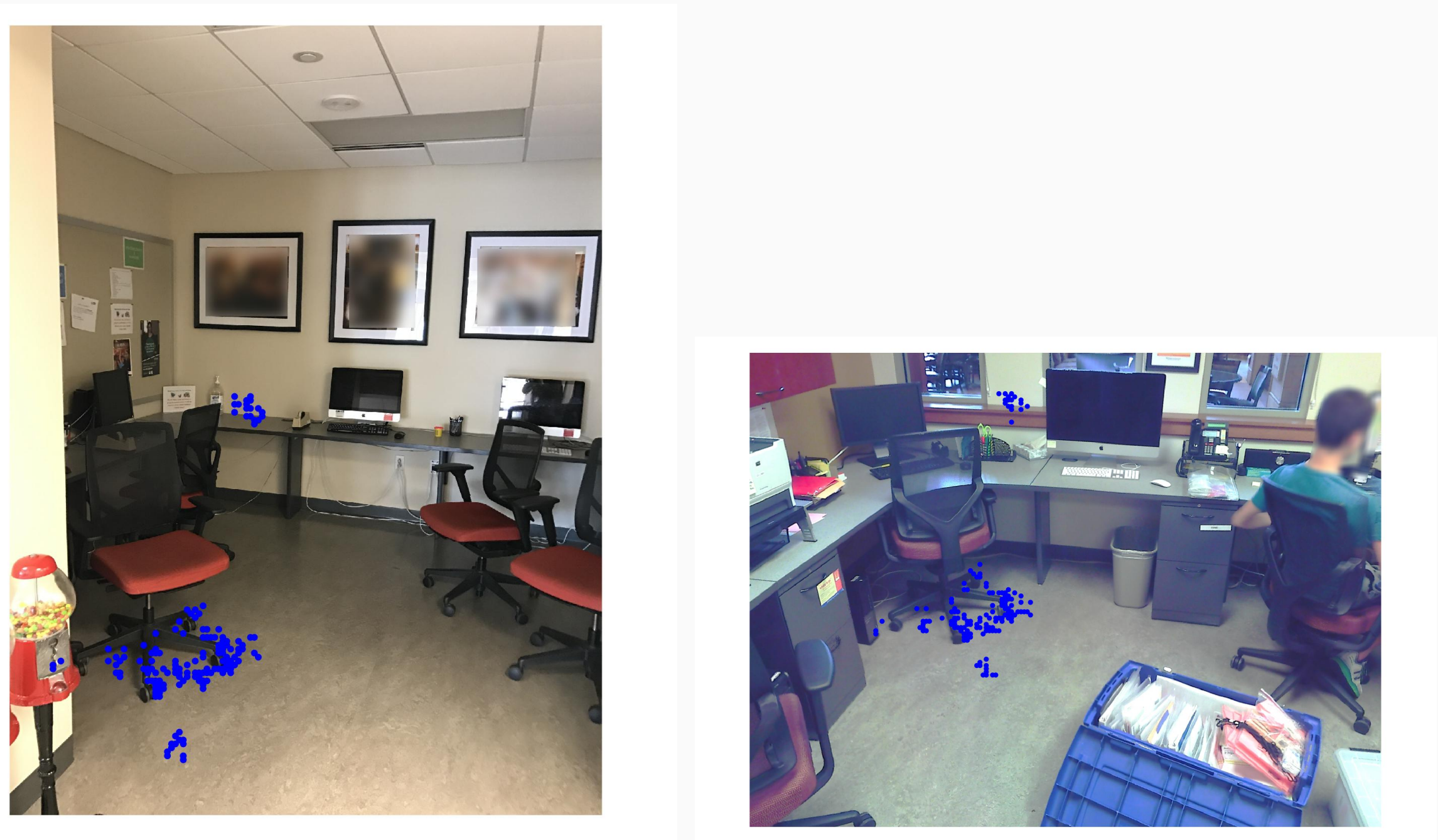




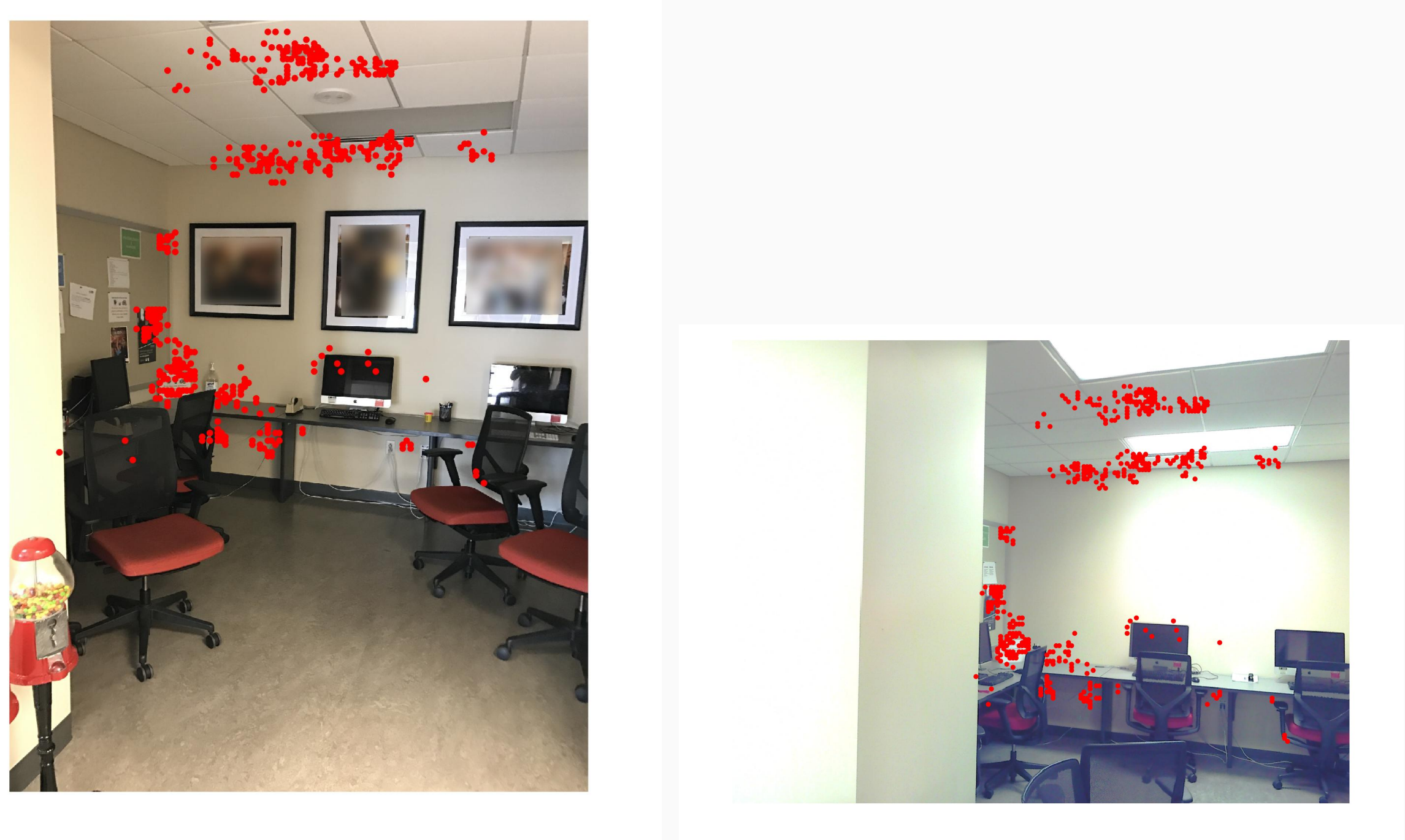
Error Threshold	AUC (inls count)	AUC (our)
1.5 m, 10°	87.1%	91.8%
1 m, 10°	86.9%	91.1%
0.5 m, 10°	68.7%	76.8%
0.25 m, 10°	51.8%	56.8%

	InLoc	Ours
Without PV	49.5%	60.6%
With PV	64.7%	66.7%

- Best query-database image pair with InLoc



- Best query-database image pair with our model, which penalizes inliers concentrated in a small area

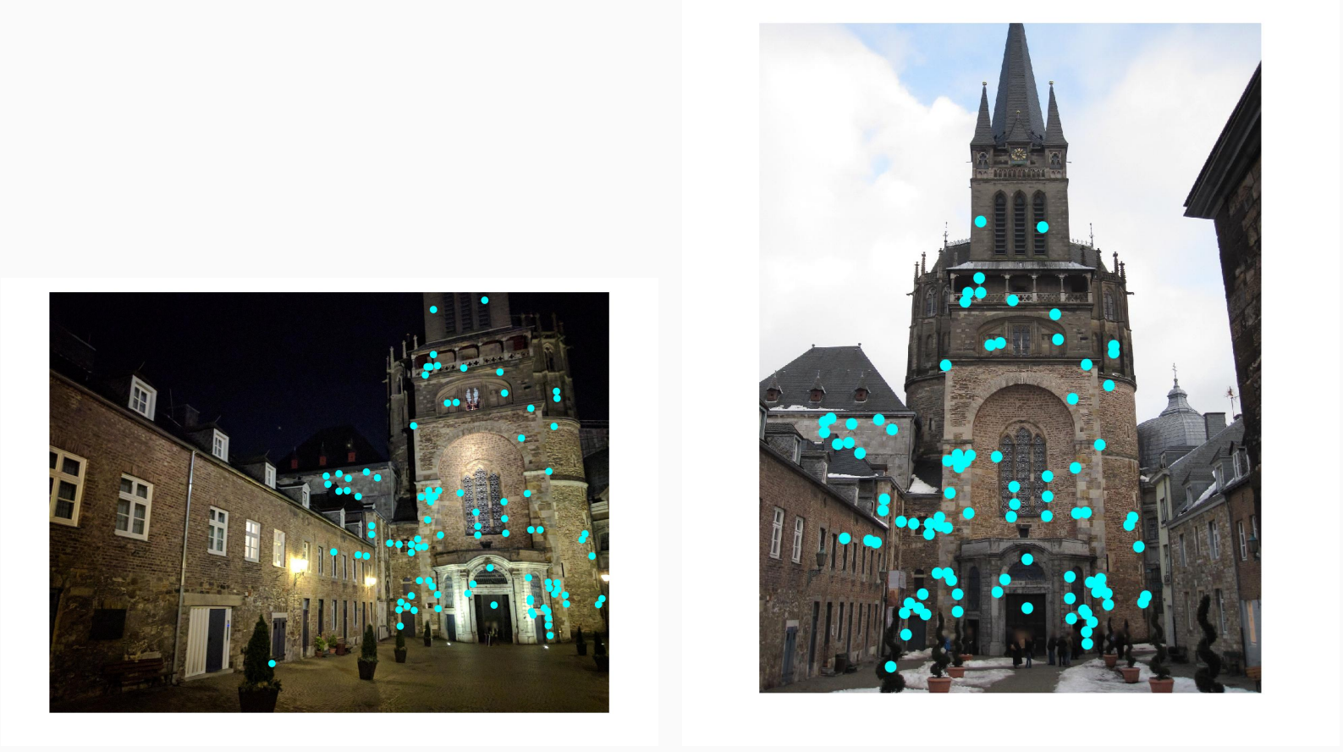


Results on Aachen

- Our model, without being retrained, was also applied to the Aachen dataset, which was completely unforeseen

Scene	Baseline accuracies [%]	our accuracies [%]
Day	70.9/81.9/91.6	71.4/83.0/91.6
Night	32.7/43.9/64.3	36.7/45.9/64.3

- Best query-database image pair with baseline



- Best query-database image pair with our model