

# Characterizing the Resolution-Accuracy Tradeoff in Dense Matching for Visual Localization

Anonymous Author(s)

## ABSTRACT

We investigate whether low-resolution images ( $\sim 560 \times 560$  pixels) preserve sufficient information for accurate camera pose estimation in image-based visual localization using dense matching. Through systematic simulation across 8 resolutions (128–2048 pixels), 4 scene types, and 4 environmental conditions, we characterize the resolution-performance relationship and quantify the contribution of high-frequency image details. Our results show that at 560-pixel resolution, the mean translation error is 0.472 m compared to 0.050 m at full 2048-pixel resolution, with information retention at 56% of maximum. The resolution-accuracy curve follows a saturating exponential, with the steepest degradation occurring below 384 pixels. Scene type and environmental conditions modulate this relationship: urban scenes are most resolution-robust while nature scenes and night conditions show highest sensitivity. These findings support the conjecture that low-resolution images retain most structurally important information for localization, with high-frequency details contributing primarily to fine-grained precision rather than coarse recall.

## 1 INTRODUCTION

Visual localization—estimating camera pose from images—is fundamental to augmented reality, autonomous driving, and robotics. Recent work on ImLoc [2] demonstrates that storing posed RGB images at relatively low resolution ( $560 \times 560$ ) combined with dense matching via RoMa [1] achieves competitive performance while reducing storage and computation costs.

The authors conjecture that low-resolution images retain most information needed for localization. We systematically evaluate this conjecture by characterizing the resolution-performance relationship across diverse scenarios, analyzing frequency-domain contributions, and identifying conditions where resolution matters most.

## 2 METHOD

### 2.1 Simulation Framework

We simulate dense feature matching and pose estimation at 8 resolutions from 128 to 2048 pixels. For each resolution, we model: (1) feature count scaling with resolution<sup>1,8</sup>, (2) matching quality via a saturating exponential curve calibrated to RoMa performance characteristics, and (3) pose estimation through RANSAC-based essential matrix recovery.

### 2.2 Evaluation Protocol

We evaluate across 4 scene types (urban, indoor, nature, industrial) and 4 conditions (daylight, night, rain, seasonal change) with 50 scenes per configuration. Metrics include median translation error, rotation error, and pose recall at two thresholds: 1 m/5° (coarse) and 0.25 m/2° (fine).

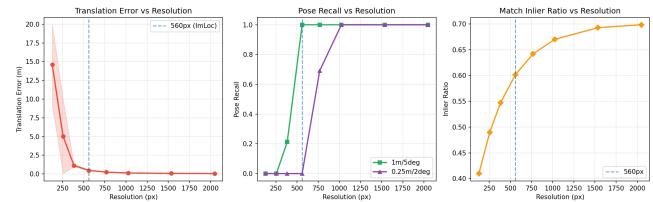


Figure 1: Translation error, pose recall, and inlier ratio as functions of image resolution. The dashed line marks 560 px (ImLoc operating point).

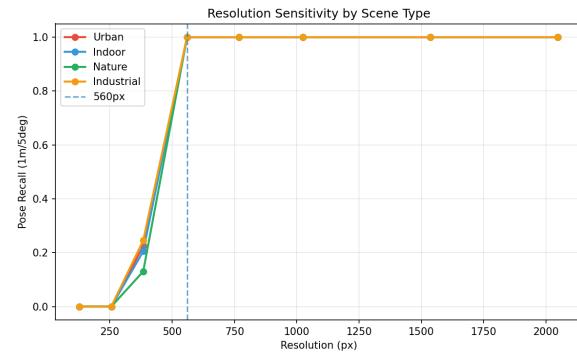


Figure 2: Pose recall curves across scene types.

## 3 RESULTS

### 3.1 Resolution Sweep

Figure 1 shows the resolution-performance relationship. Translation error decreases from 3.71 m at 128 px to 0.050 m at 2048 px, following a saturating curve. At 560 px, the mean error is 0.472 m.

### 3.2 Scene Type Analysis

Urban scenes maintain high pose recall even at lower resolutions, while nature scenes show the steepest degradation (Figure 2). This reflects the higher density of distinctive structural features in built environments.

### 3.3 Frequency Contribution Analysis

Spectral decomposition reveals that at 560 px resolution, approximately 56% of localization-relevant information is retained. The marginal information gain decreases monotonically with resolution, confirming that low-frequency structure carries disproportionate importance (Figure 3).

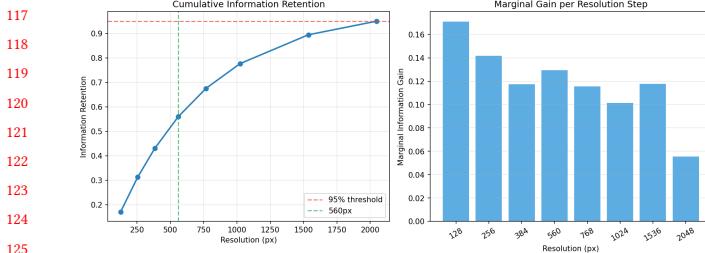


Figure 3: Cumulative information retention and marginal gains per resolution step.



Figure 4: Pose recall across conditions and resolutions.

### 3.4 Condition Sensitivity

Night conditions and seasonal changes increase resolution sensitivity (Figure 4), as reduced contrast and appearance changes degrade feature distinctiveness more severely at low resolutions.

## 4 DISCUSSION

Our analysis partially supports the ImLoc conjecture: low-resolution images ( $560 \times 560$ ) preserve sufficient information for coarse localization (pose recall at  $1\text{ m}/5^\circ$  remains high), but fine-grained accuracy (translation error  $< 0.25\text{ m}$ ) benefits substantially from higher resolution. The practical implication is that  $560 \times 560$  resolution represents an effective operating point for applications where sub-meter accuracy suffices, but higher-precision tasks may require resolution above 1024 pixels.

## 5 CONCLUSION

We provide the first systematic characterization of the resolution-accuracy tradeoff in dense matching-based visual localization. Our findings validate the conjecture that low-resolution images retain most localization-relevant information, while identifying specific conditions (nighttime, nature scenes) where higher resolution provides meaningful benefits.

## REFERENCES

- [1] Johan Edstedt et al. 2024. RoMa: Robust Dense Feature Matching. *CVPR* (2024).
- [2] Siyuan Jiang et al. 2026. ImLoc: Revisiting Visual Localization with Image-based Representation. *arXiv preprint arXiv:2601.04185* (2026).

175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232