

# PERSPECTIVE-N-POINT IN PRACTICE



Performance, Robustness, and Accuracy for Mesh-Based Localisation

F. Vultaggio (1,2), P. Fanta-Jende (1), M. Gerke (2)

1

Austrian Institute of Technology  
- Center for Vision, Automation  
and Control, Unit Assistive and  
Autonomous Systems

2

Technische Universität  
Braunschweig - Institute of  
Geodesy and Photogrammetry



# BACKGROUND

## What is visual localisation?

To find the position and orientation of a query image given a pre-existing *map* of the environment



# BACKGROUND

## What is visual localisation?

“Finding a particular needle in a huge haystack containing many other similarly-looking needles”

Liu, Liu, Hongdong Li, and Yuchao Dai. "Efficient global 2d-3d matching for camera localization in a large-scale 3d map." Proceedings of the IEEE International Conference on Computer Vision. 2017.





Typical query image



Typical reference image

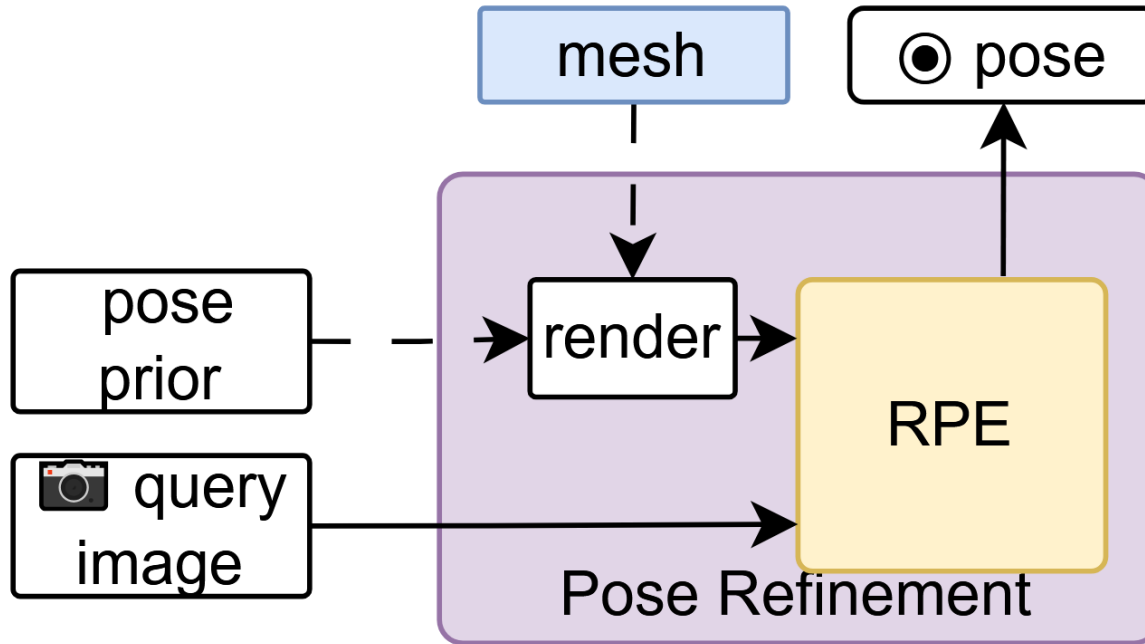
The Solution?

MESHES



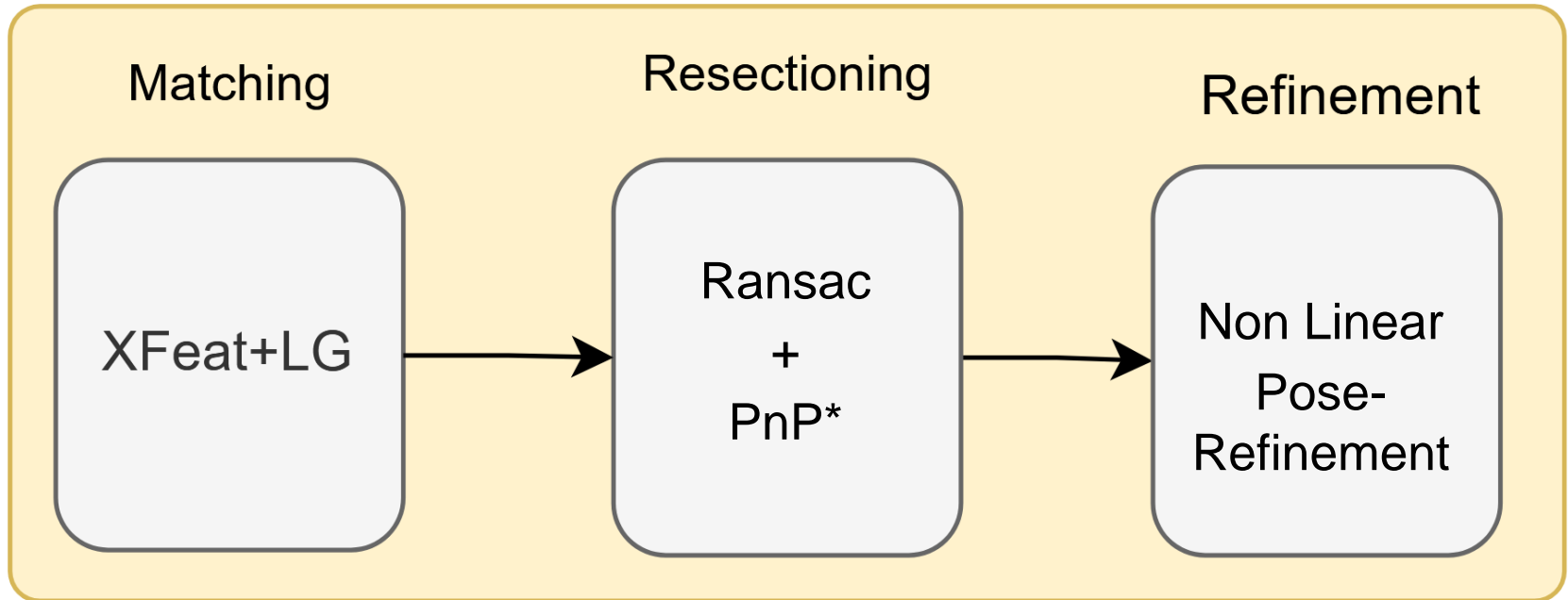
HOW?





\*RPE = Relative pose estimation

Vultaggio, Francesco, et al. "Investigating Visual Localization Using Geospatial Meshes." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 48 (2024): 447-454.



PnP\* = Perspective-n-Point

# TODAY'S QUESTION

What is the best PnP setup to use for visual localization?



# WHY ?

Visual resectioning is an often-overlooked aspect of Visual Localization?



## WHY ?

PnP works focus on synthetic data.  
Often generated inconsistently across works.



# OUR BENCHMARK



# OUR BENCHMARK

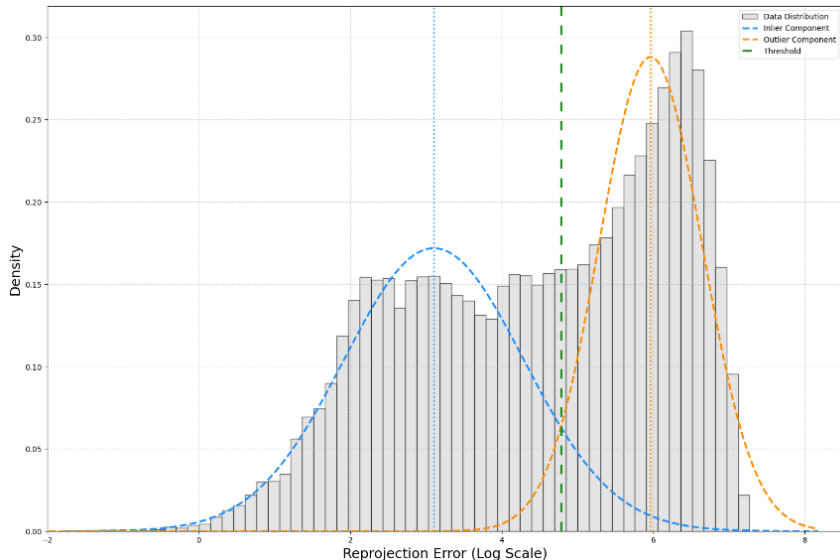
- Render RGB and Depth images
- Match synthetic RGB and real images
- Lift the 2D-2D correspondences to 3D using the Depth image
- Compute the per-correspondence reprojection error
- Camera resection: [experiments]
  - PnP alone
  - RANSAC + PnP
  - RANSAC + PnP + Non Linear Refinement

# SOURCES OF NOISE

- Noise in the 3D reconstruction of the aerial data
- Noise in the mesh generation
  - Noise in the 3D data
- Noise in the camera model estimation
- Noise in the 2D-2D matches
  - Keypoint detection noise
- Gross outliers
  - Mis-association in 2D-2D matches

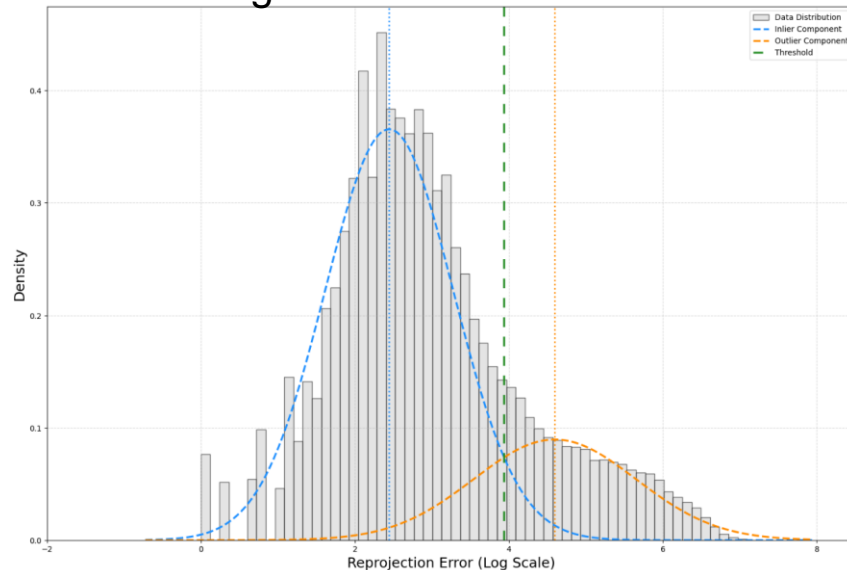
# IMAGE MATCHES ERROR

## SIFT + NN



Bellavia, Fabio, and Carlo Colombo. "Is there anything new to say about SIFT matching?." *International journal of computer vision* 128.7 (2020): 1847-1866.

## Xfeat + LightGlue

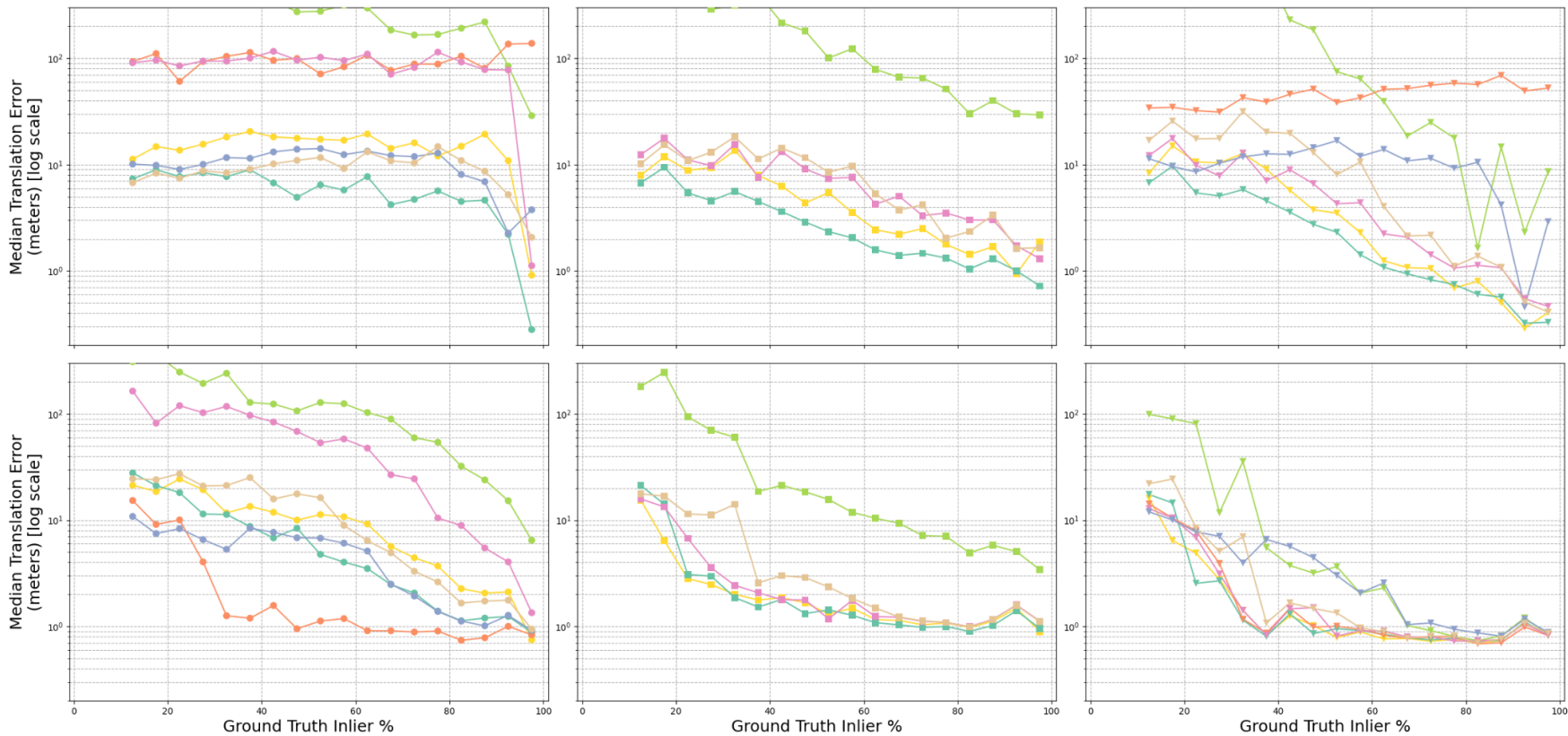


Potje, Guilherme, et al. "Xfeat: Accelerated features for lightweight image matching." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024.

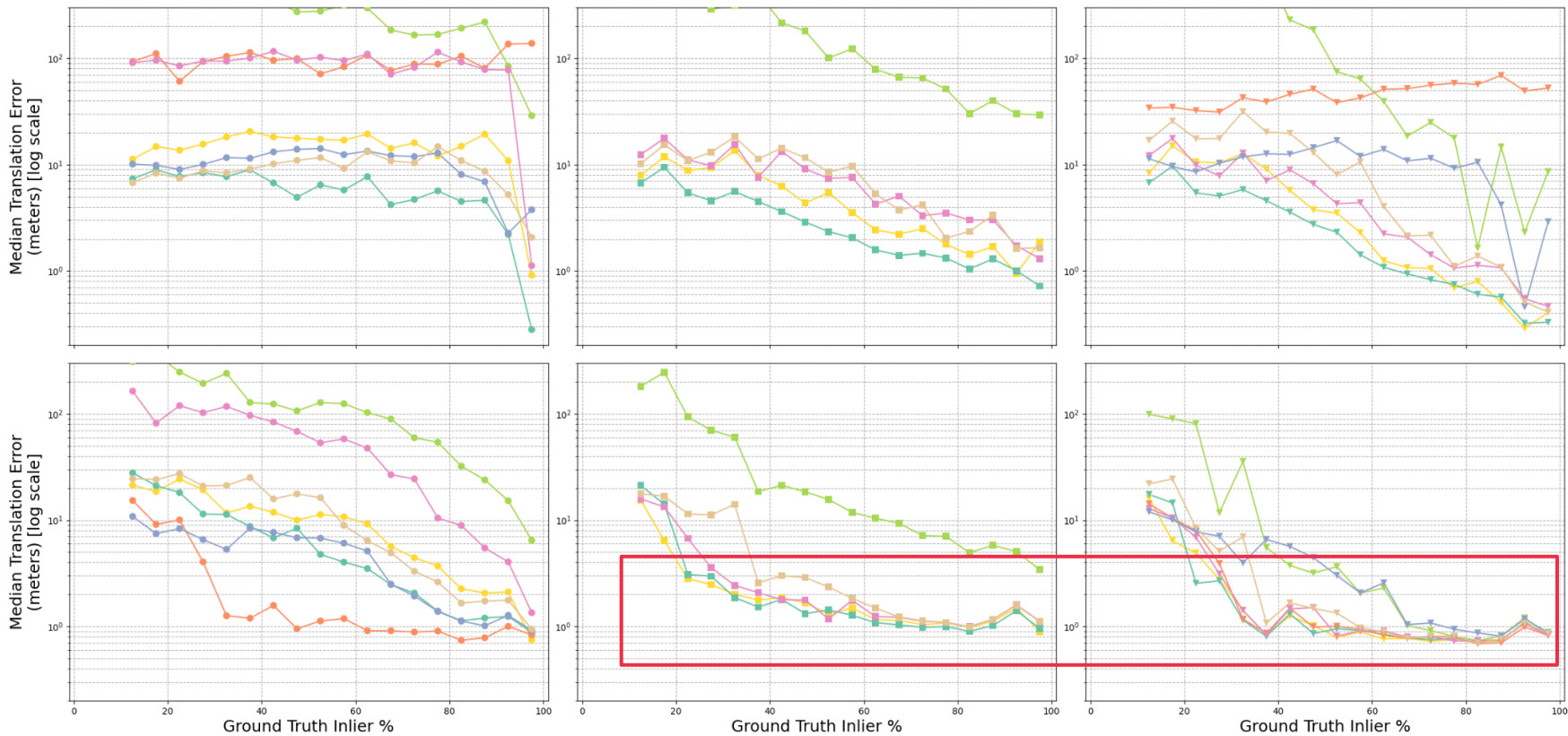
# PNP ALGORITHMS

<b>Algorithm</b>	<b>Reference</b>	<b>Claims</b>
EPnP	(Lepetit et. al. 2009)	Fast
RPnP	(Li et. al 2012)	Degenerate cases
OPnP	(Zheng et. al. 2013)	Accurate
CPnP	(Zeng et. al. 2023)	For modern matchers
oDLT	(Henery et. al. 2024)	Efficient
REPPnP	(Ferraz et. al. 2014)	Robust
R1PPnP	(Zhou et. al. 2019)	Robust

# OUR RESULTS

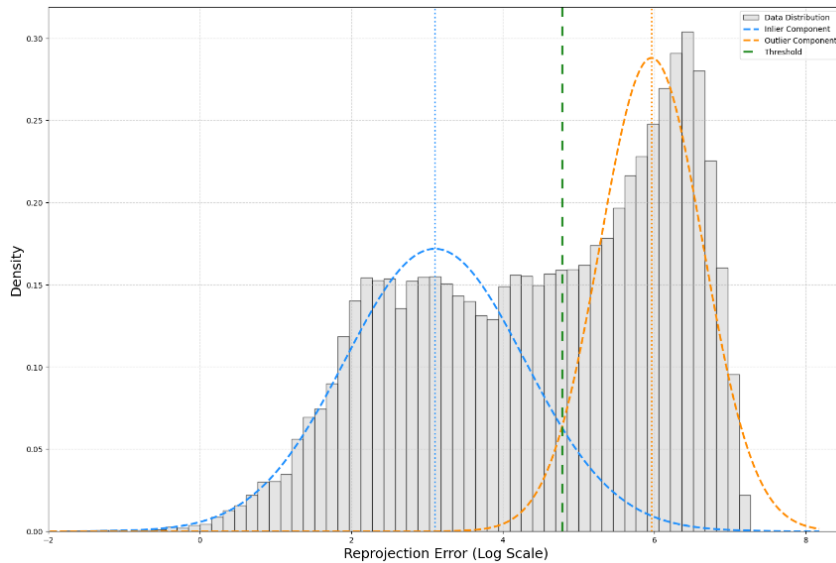


# OUR RESULTS

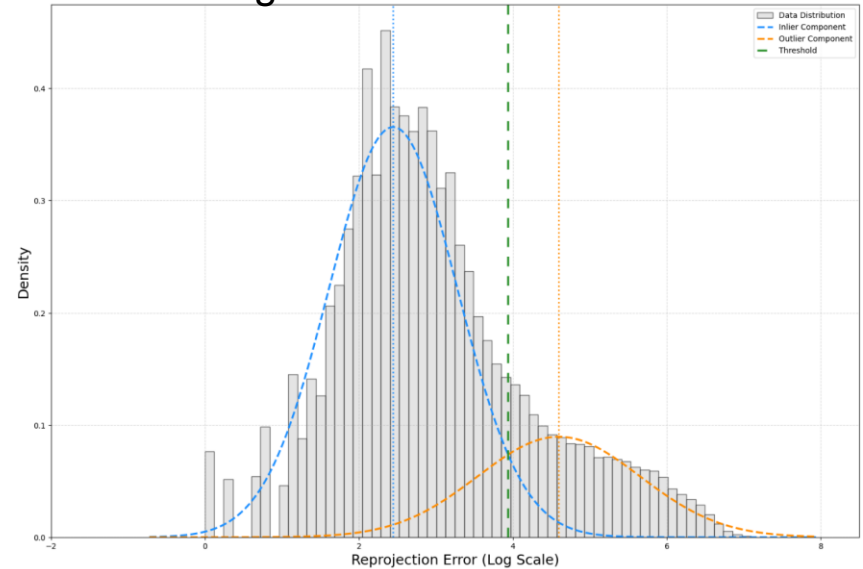


# OUR BENCHMARK

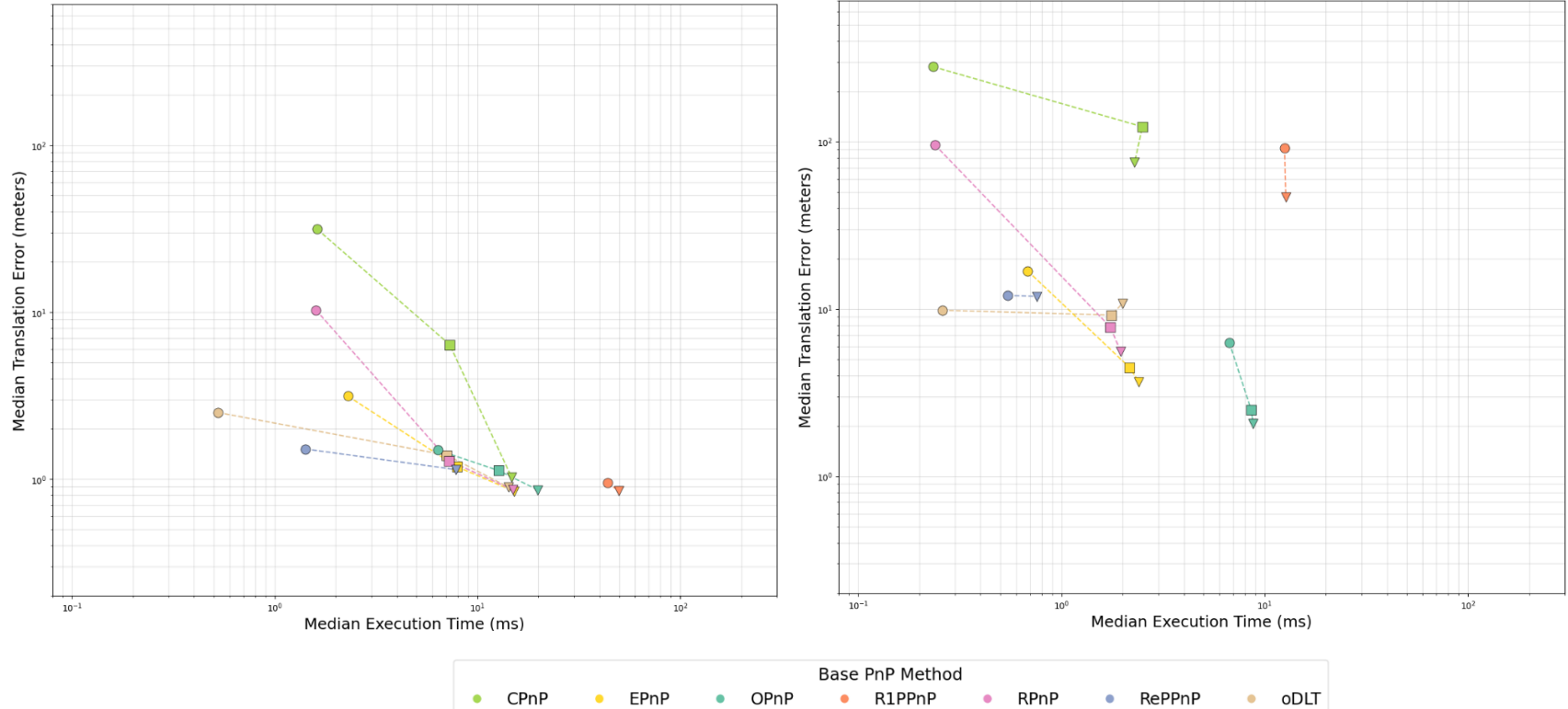
## SIFT



## Xfeat + LightGlue



# OUR RESULTS



# CONCLUSIONS

- **PnP alone is not enough**
  - Pre and Post processing steps are **necessary**
- Real world accuracy is **comparable across PnP variants**
- **Focus on the accuracy of image correspondences**
  - Increase accuracy of keypoint detection
  - Model inaccuracy in the resectioning process

# THANK YOU!



Francesco Vultaggio  
Phillipp Fanta-Jende  
Markus Gerke



Francesco.Vultaggio@ait.ac.at

**Coming soon!**

First geospatial visual localization benchmark dataset