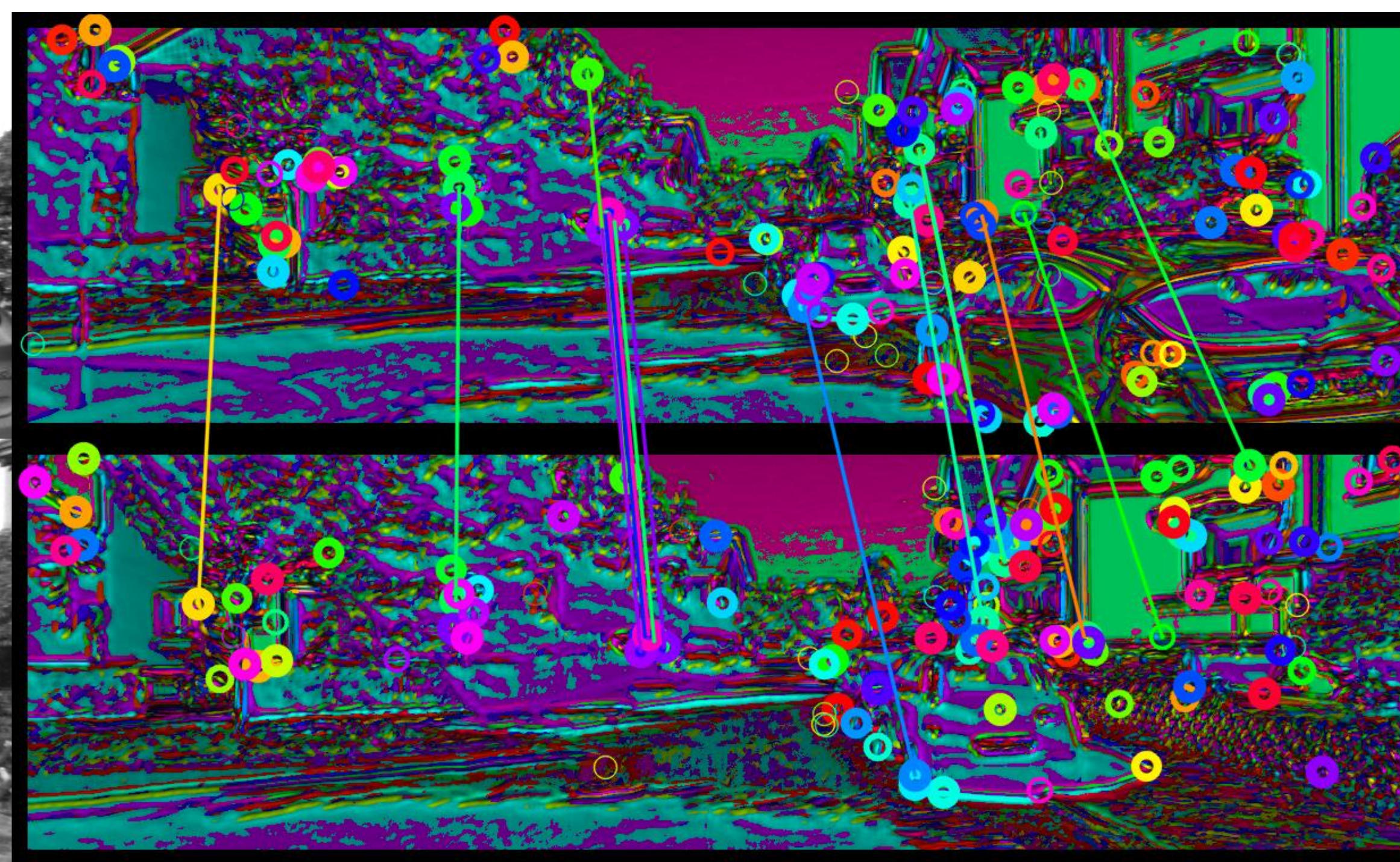


Matching Features without Descriptors: Implicitly Matched Interest Points

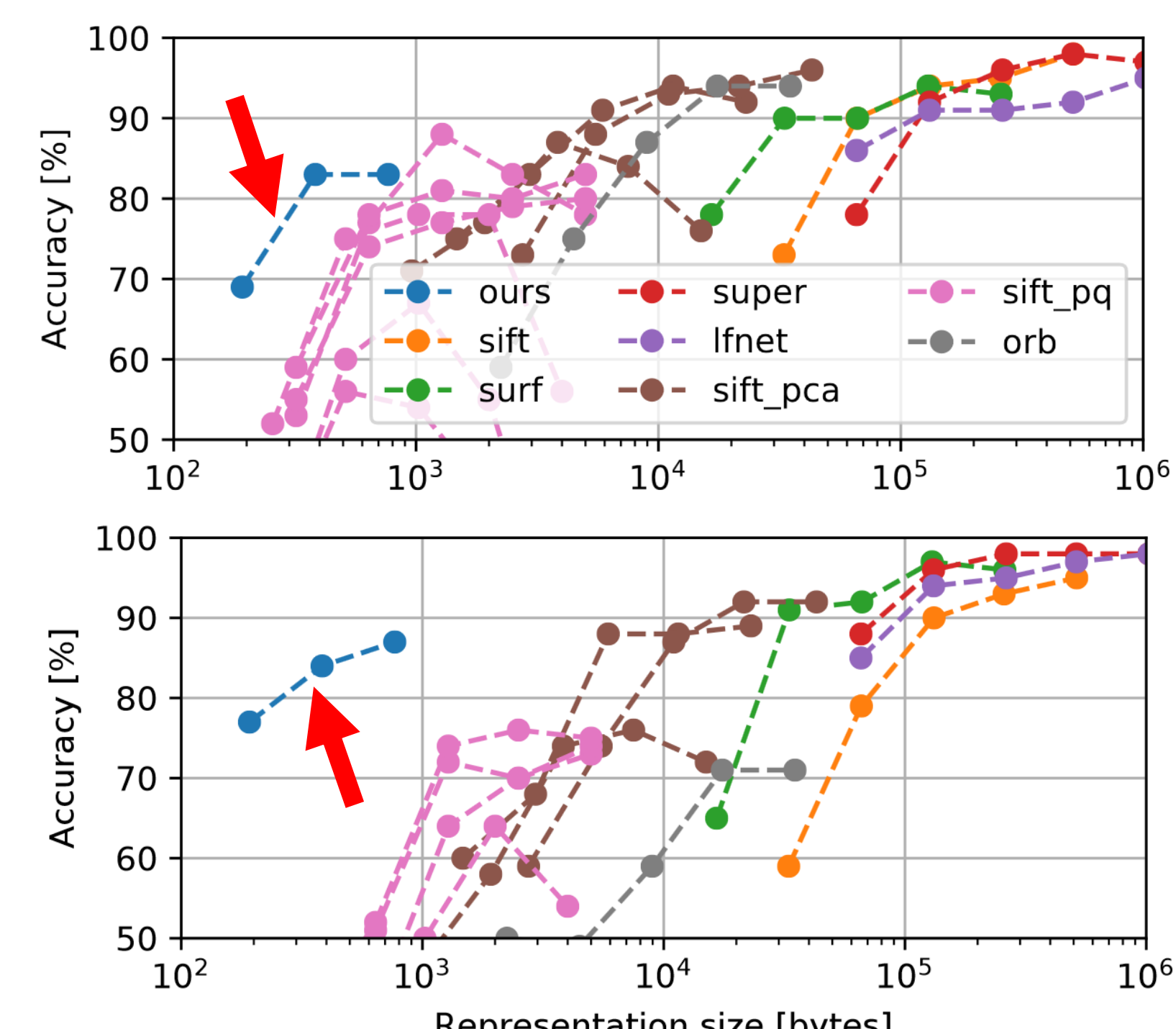
Titus Cieslewski, Michael Bloesch, Davide Scaramuzza



Input images



128-channel output, per-channel maxima, inliers

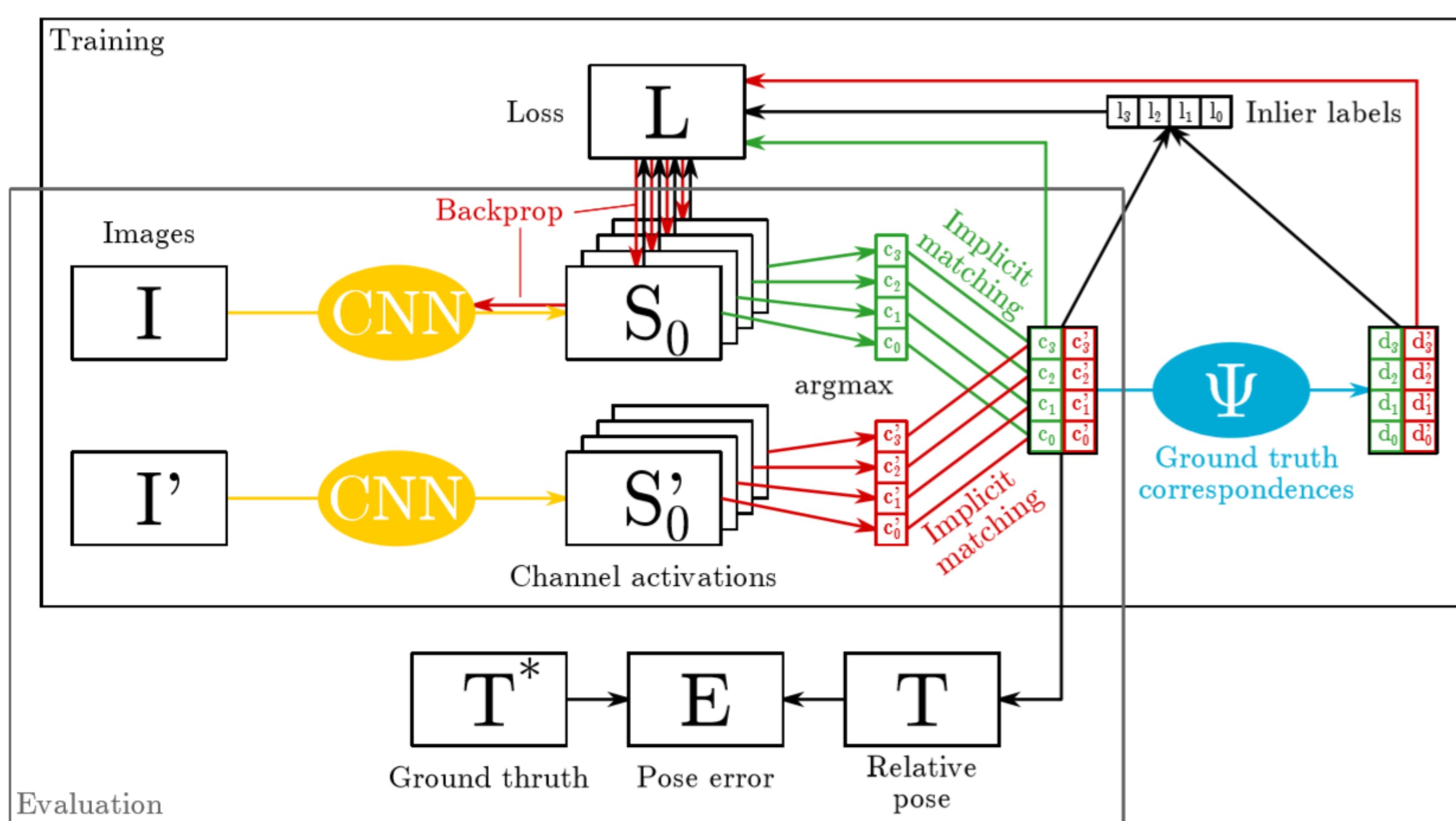


Performance VS repr. size on KITTI, EuRoC

*Features are typically matched between images using descriptors.
Can we exploit machine learning to do the same without descriptors?*

Architecture

- Input: Image; Output: 128 - channel image ●
- Argmax of the channel is "its" interest point
- Interest points of same channel implicitly matched across viewpoints ●●
- Self-supervised training, random initial weights
- Trained on pairs of images



Loss

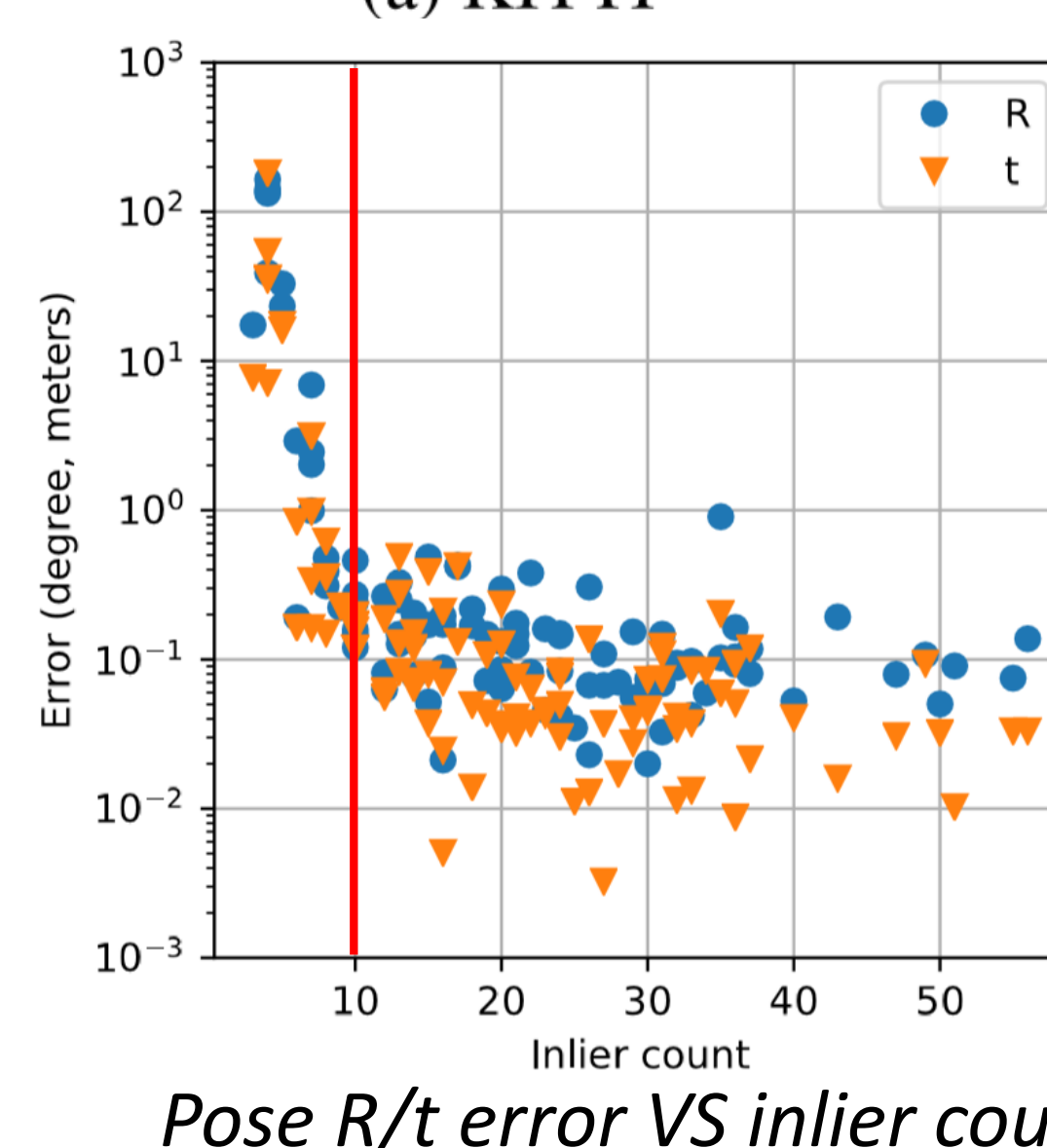
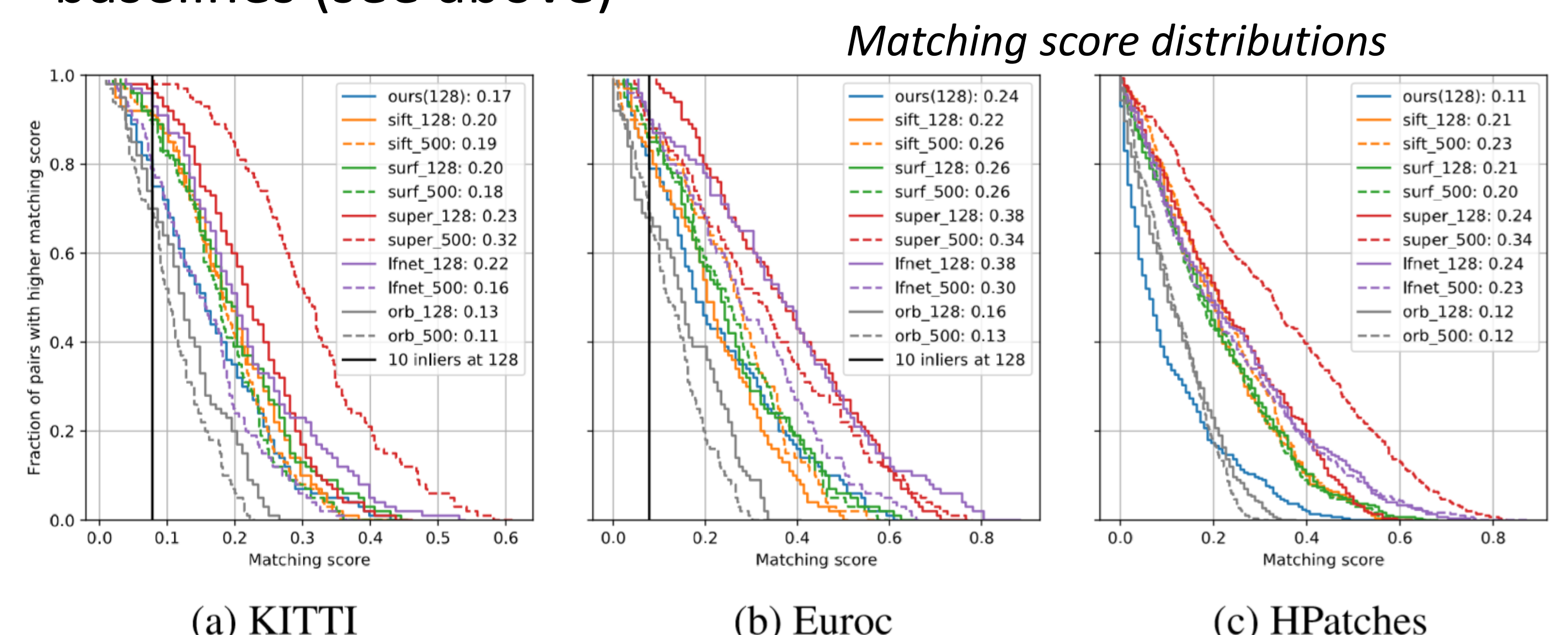
- **Inlier reinforcement:** Reinforce max if it resulted in an inlier match
- **Redundancy suppression:** Ensure an inlier argmax location is selected by only one channel
- **Correspondence reinforcement:** Reinforce true correspondences of outliers

Self-supervised Training

- From uncalibrated image sequences
- Select image pairs based on visual overlap (KLT track densely sampled points)
- Ground truth labels: KLT track argmax of one image into the other image ●
- Any other labeling would also work

Results

- Matching score: Like SIFT/SURF on KITTI and EuRoC, like ORB on Hpatches
- Relative pose estimation: 10 inliers are enough → similar results as matching score
- Best representation size – accuracy trade-off among baselines (see above)



Pose R/t error VS inlier count



Open source code:

https://github.com/uzh-rpg/imips_open

Sponsors



Swiss National Centre of Competence in Research



SWISS NATIONAL SCIENCE FOUNDATION



NVIDIA GPU Grant