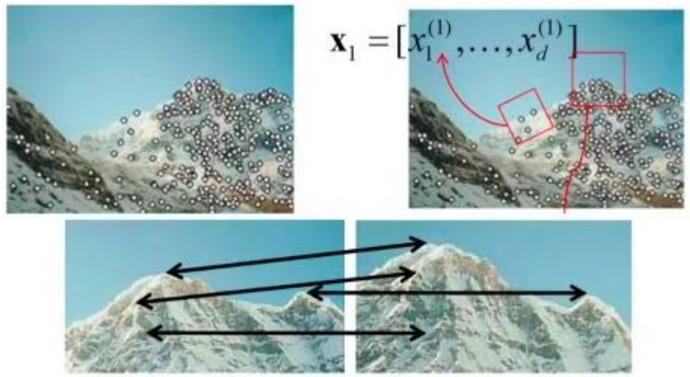
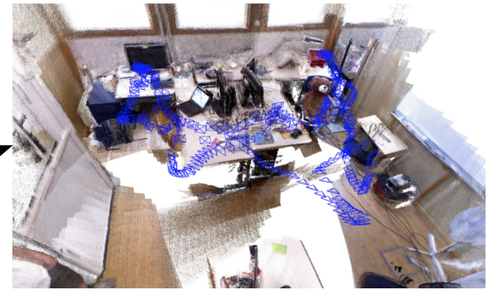
PoGOSS-ST: Policy Gradient for Optimizing the Stability Score for the Shi-Tomasi Detector

Local features are ...



corners, **edges** serve as the unique identifiers of an image

used for



An ORB-SLAM2

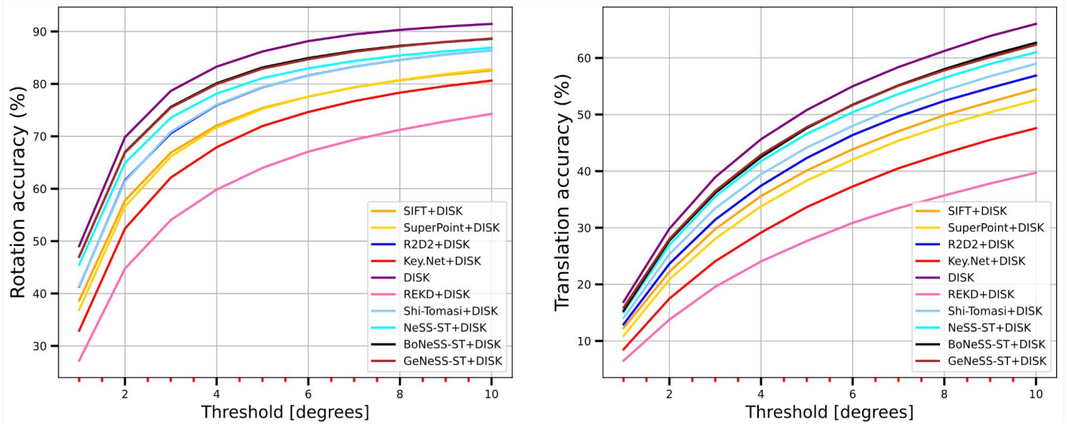
From "ORB-SLAM2: …" by Raúl Mur-Artal, 2017.

used for

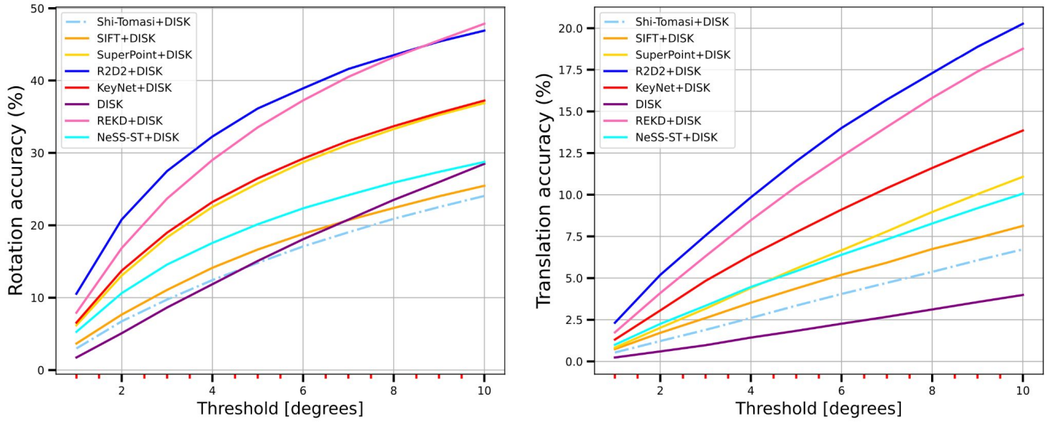
A COLMAP scene reconstruction.

From "Structure-from-Motion Revisited" by J. L. Schönberger, 2016

2048 points



128 points



comparing different methods with different numbers of points

Our proposal → $V^{\pi_\theta}(s_o) = \mathbb{E}_{s'}\, \mathbb{E}_{\{a', a'', \ldots\} \sim \pi(*|s)} R(s', \{a', a'', \ldots\})$

Value function — using multi-action Reward

Skoltech

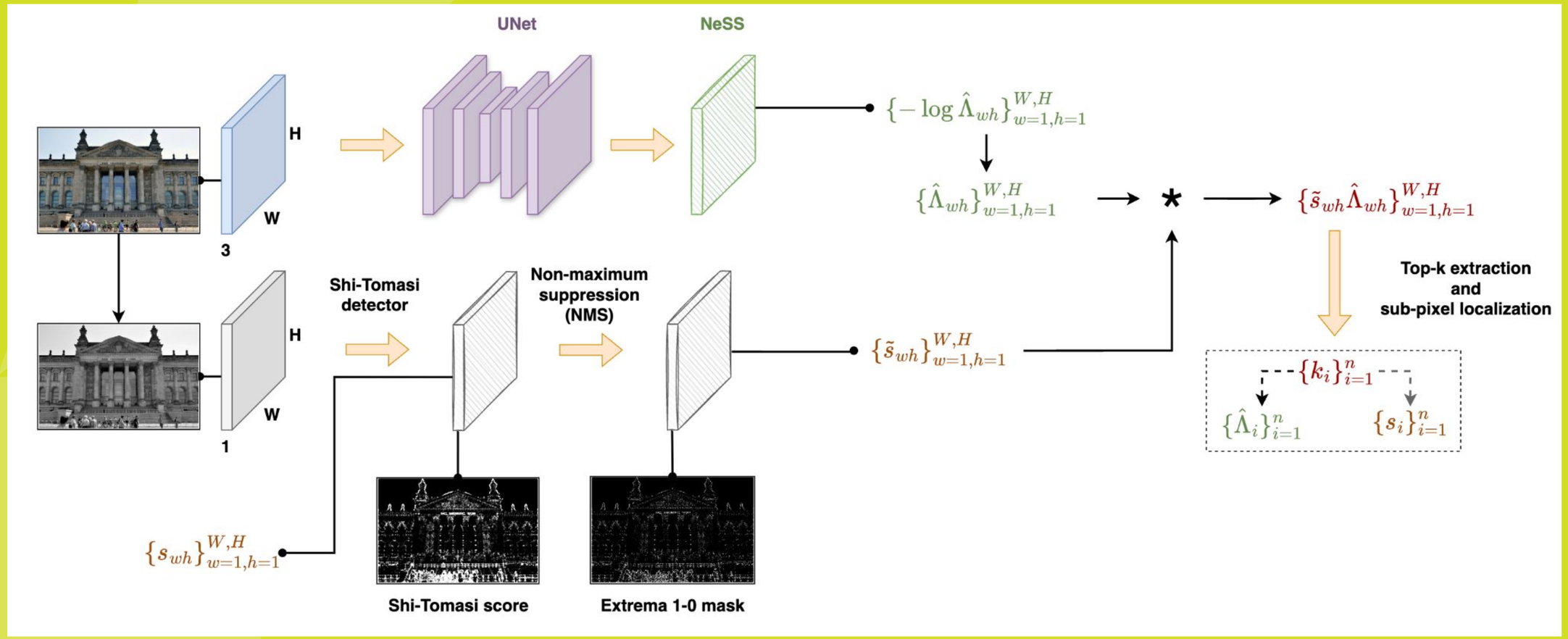
Machine Learning, Skoltech, 2025

Team

Denis Fatykhoph
German Devchich
Vladislav Sarmatin

PoGOSS-ST: Policy Gradient for Optimizing the Stability Score for the Shi-Tomasi Detector

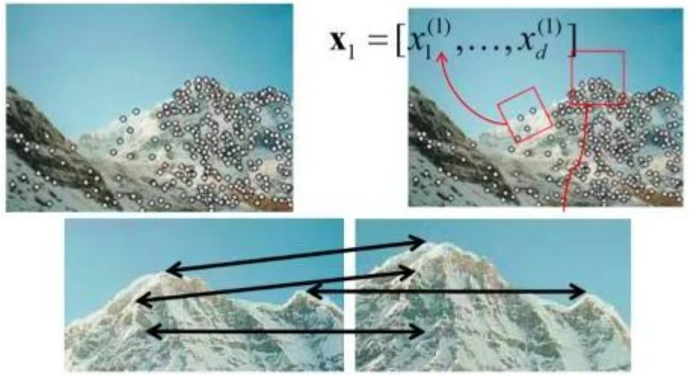
MOBILE ROBOTICS LAB



Skoltech

Keypoints is All You Need

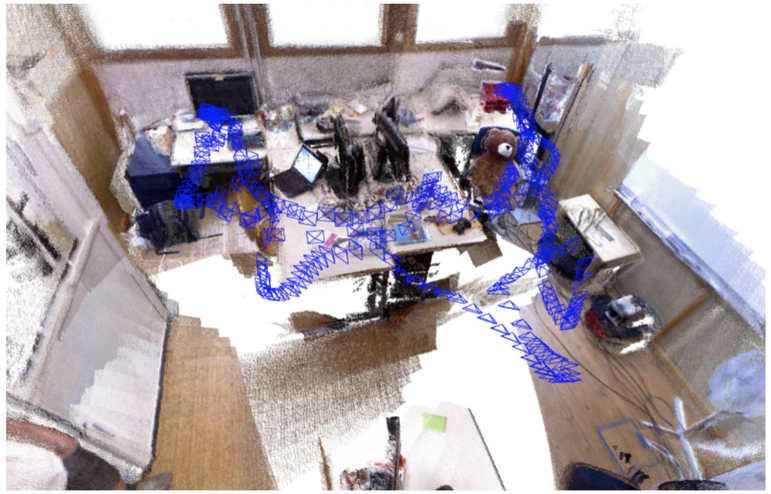
Local features are ...



corners, **edges**, and **texture patterns**, serve as the unique identifiers of an image

A COLMAP scene reconstruction.

From "Structure-from-Motion Revisited" by J. L. Schönberger, 2016, CVF Open Access.

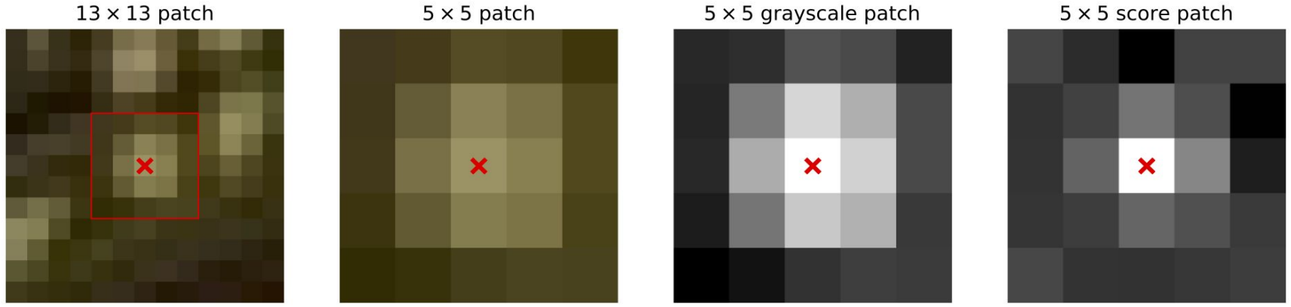


An ORB-SLAM2 visual localization and scene reconstruction.

From "ORB-SLAM2: an Open-Source SLAM System …" by Raúl Mur-Artal, 2017, arXiv.

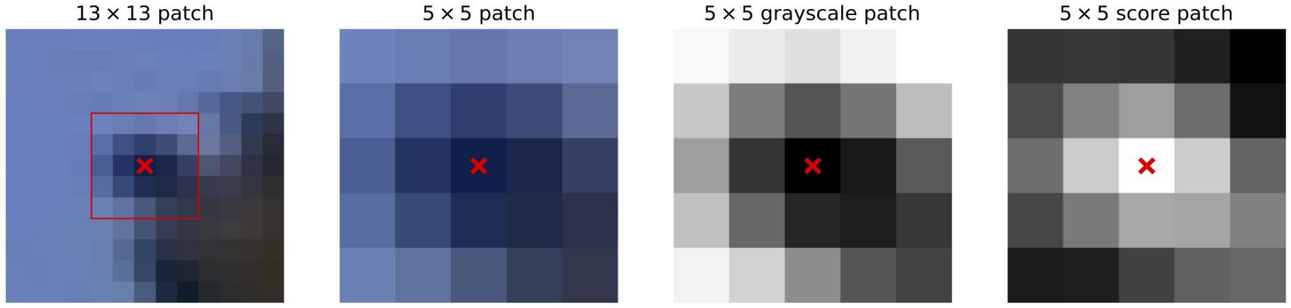
Local feature extraction is a **cornerstone** of computer vision, enabling tasks like 3D reconstruction, object recognition, and SLAM.

Background: Keypoints Detection



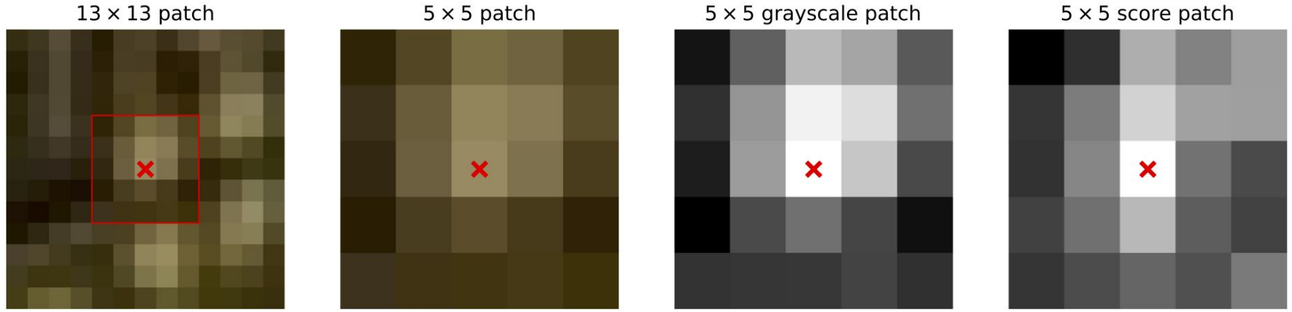
A keypoint detected by the Determinant-of-the-Hessian detector (DoH).

$$\det L = L_{xx}L_{yy} - L_{xy}^2$$



A keypoint detected by the Laplacian-of-the-Gaussian detector (LoG).

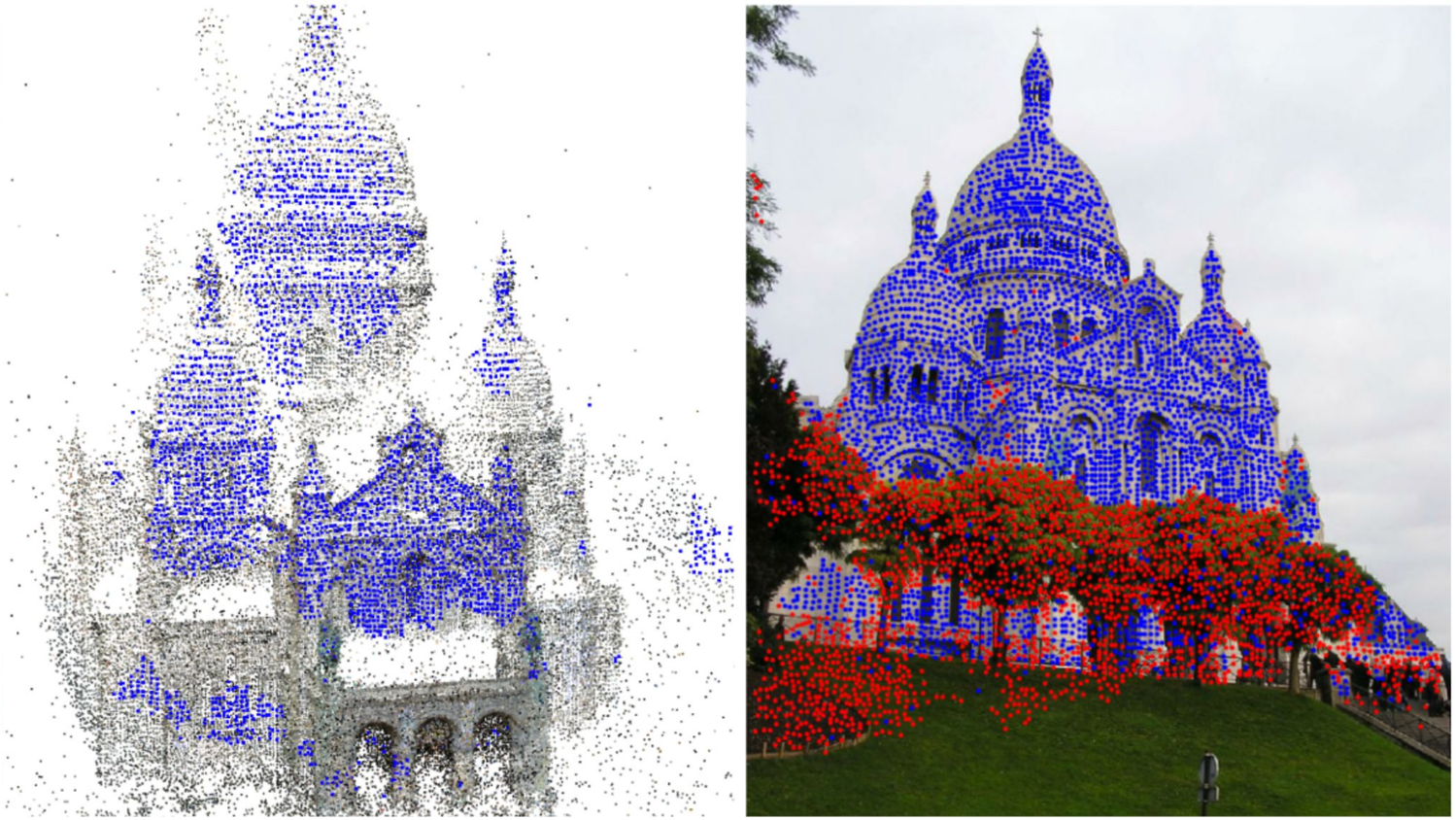
$$\nabla^2 L = L_{xx} + L_{yy}$$



A keypoint detected by the Shi-Tomasi detector.

Looks for points where the grayscale intensity undergoes a rapid change

DISK: learning local features with policy gradient

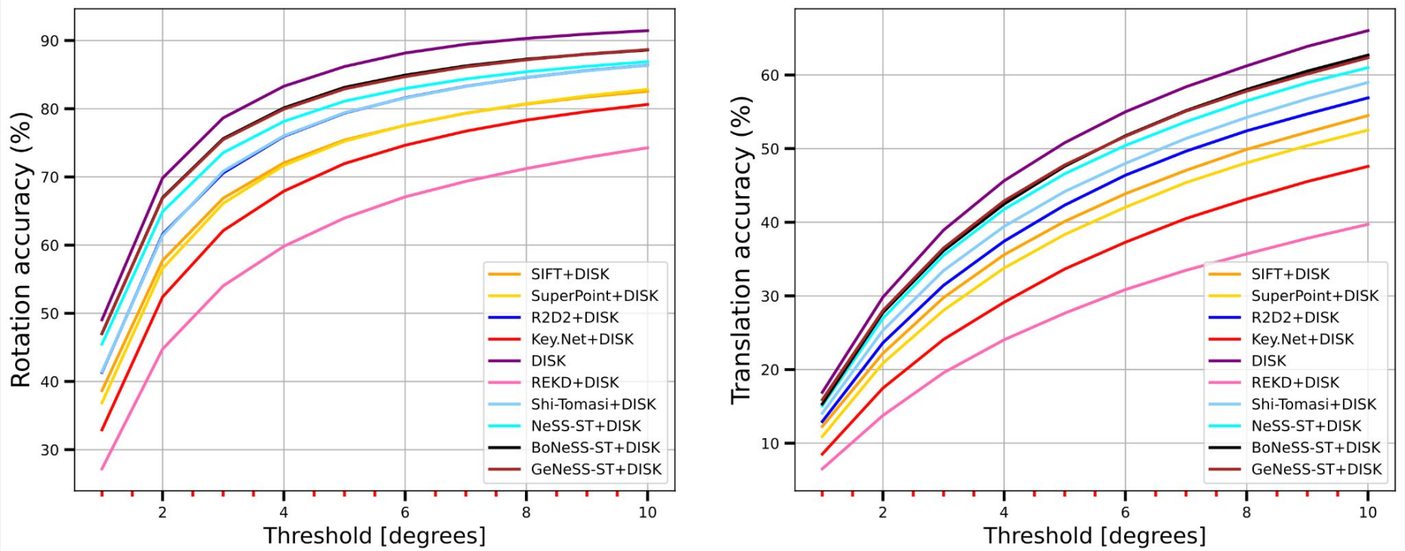


“Sacre Coeur” reconstruction from 1179 images 190k landmarks
From “DISK: Learning Local Features with Policy Gradient” by Tyszkiewicz, 2020

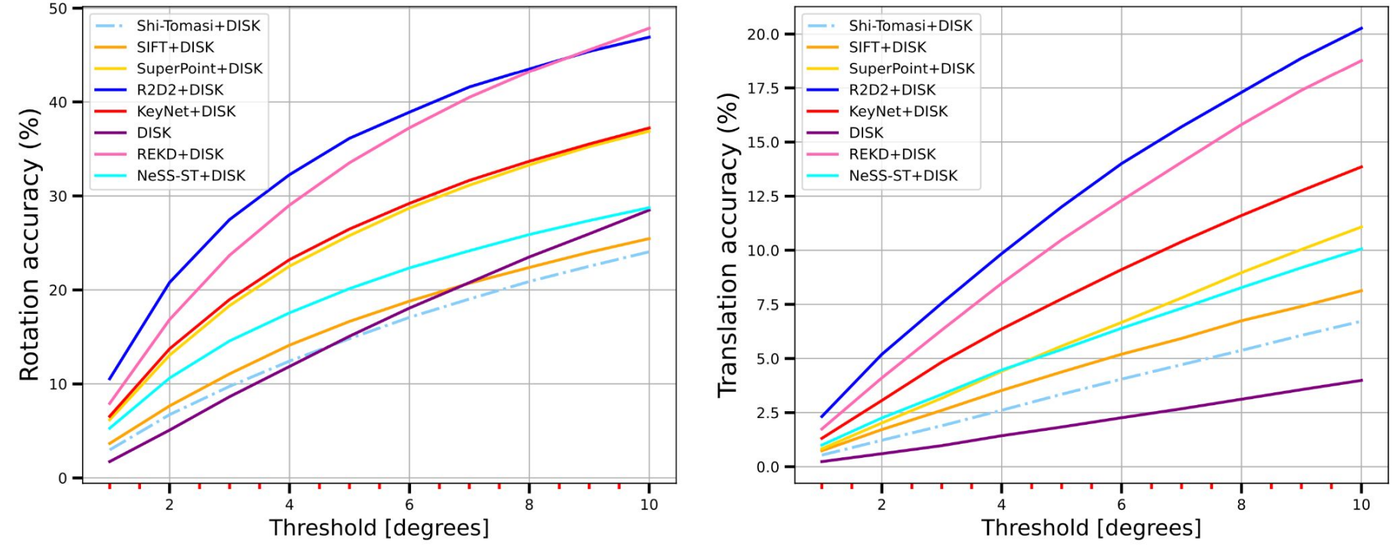
Local feature frameworks are difficult to learn in an end-to-end fashion, due to the discreteness inherent to the selection and matching of sparse keypoints.DISK (DIScrete Keypoints), a novel method that overcomes these obstacles by leveraging principles from Reinforcement Learning (RL), optimizing end-to-end for a high number of correct feature matches.

DISK: learning local features with policy gradient

2048 points



128 points



With a **sufficient number of points**, DISK demonstrates the best results among all methods in determining rotation and translation displacement.

However, with their lack, its effectiveness is significantly reduced, inferior to other approaches. This is because DISK, as a purely **geometric method**, requires a large number of points to reliably establish geometric correspondences and filter out false matches. With a small number of points, methods using **additional features** (for example, descriptors in R2D2+DISK) gain an advantage.

Skoltech

Suggested Method

Thx!

Skoltech