Learning to Make Keypoints Sub-Pixel Accurate

ETH zürich[1] CVG
Microsoft[2]

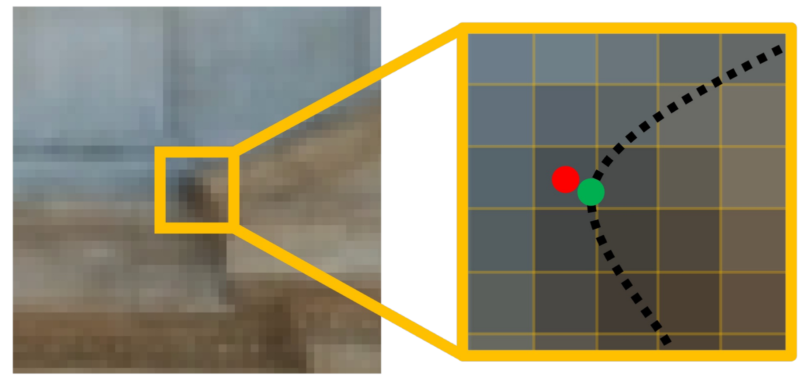
Shinjeong Kim[1], Marc Pollefeys[1, 2], Daniel Barath[1]

ECCV

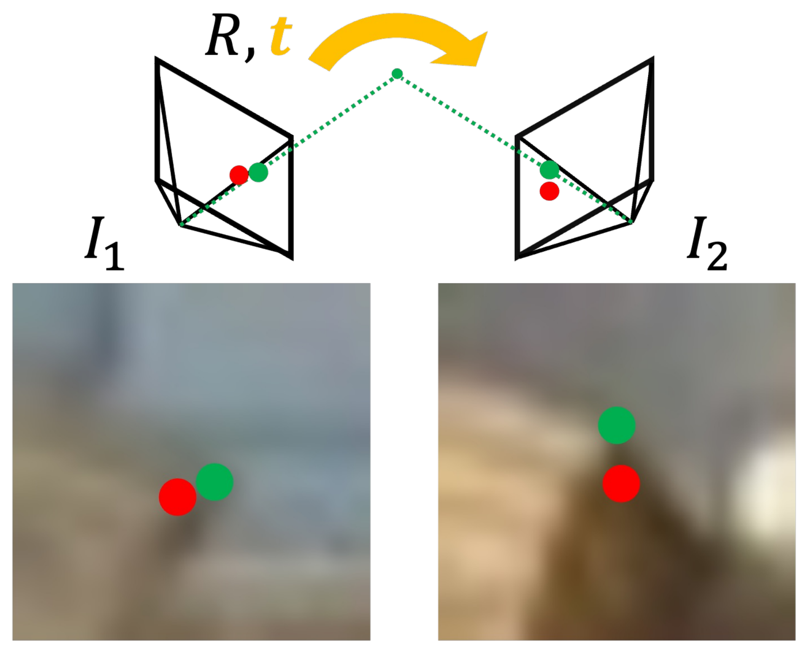
Introduction

Many learned detectors are

1. Only up to pixel-level accurate



2. Weak invariance with camera baseline



Ideal / detected keypoints

Such learned detectors include

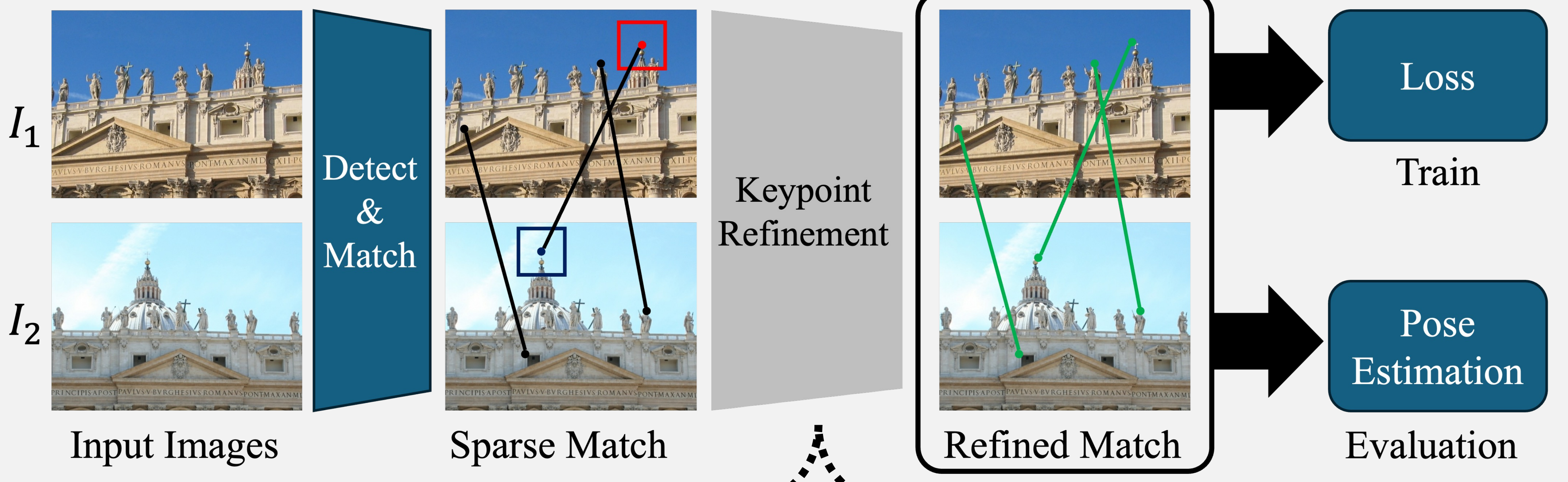
	Subpixel-accurate	Baseline-invariant
SuperPoint	✗	✗
Key.Net	✓	✗
ALIKED	✓	🤔
DeDoDe	✗	🤔
XFeat	✗	🤔

🤔: objective on point reprojection, not pose

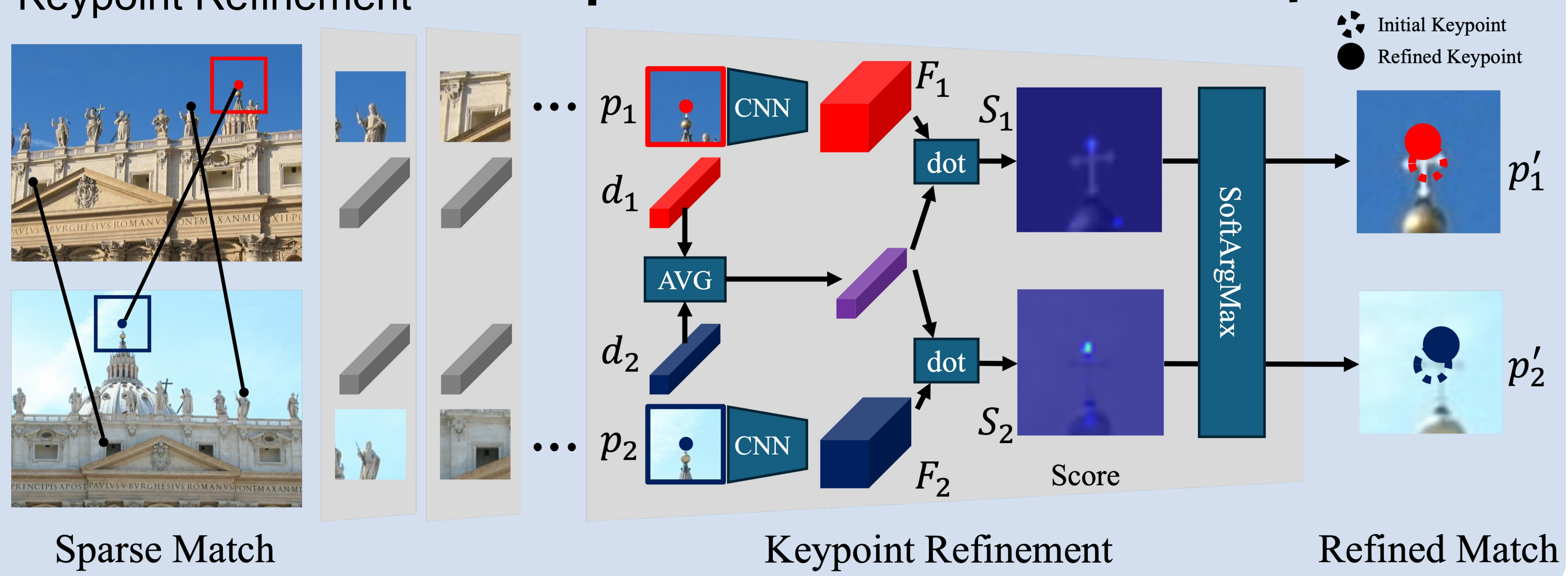
Method

Goal: Enhance any keypoint detectors with sub-pixel precision by learning offset vector in a two-view consistent way, to improve relative pose estimation.

Pipeline

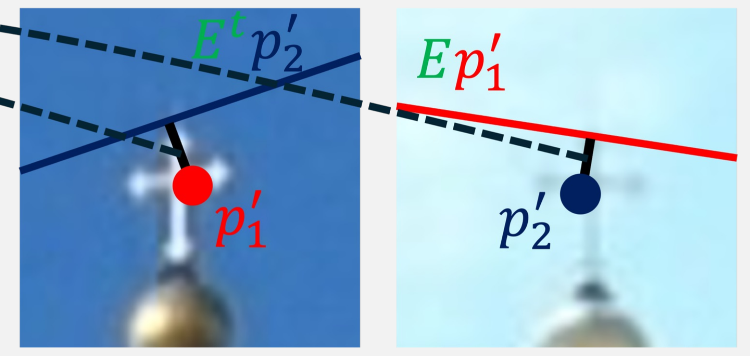


Keypoint Refinement



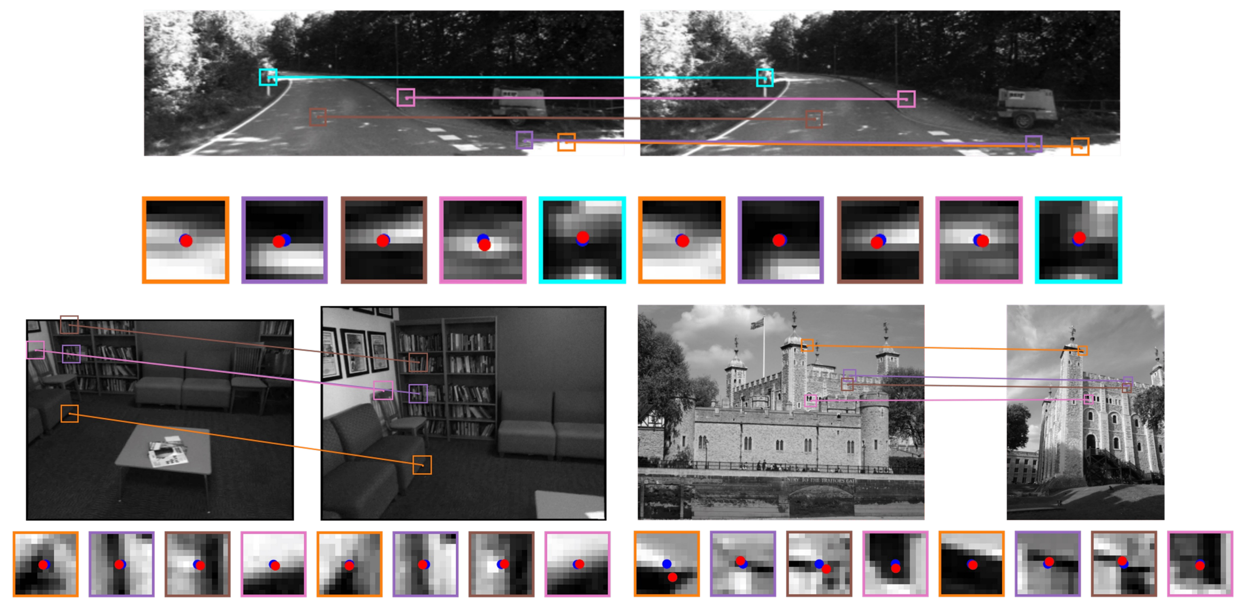
Sampson Loss

$$\ell(p'_1, p'_2, E) = \frac{\left({p'}_2^{t} E p'_1\right)^2}{[Ep'_1]_0^2 + [Ep'_1]_1^2 + [E^t p'_2]_0^2 + [E^t p'_2]_1^2}$$

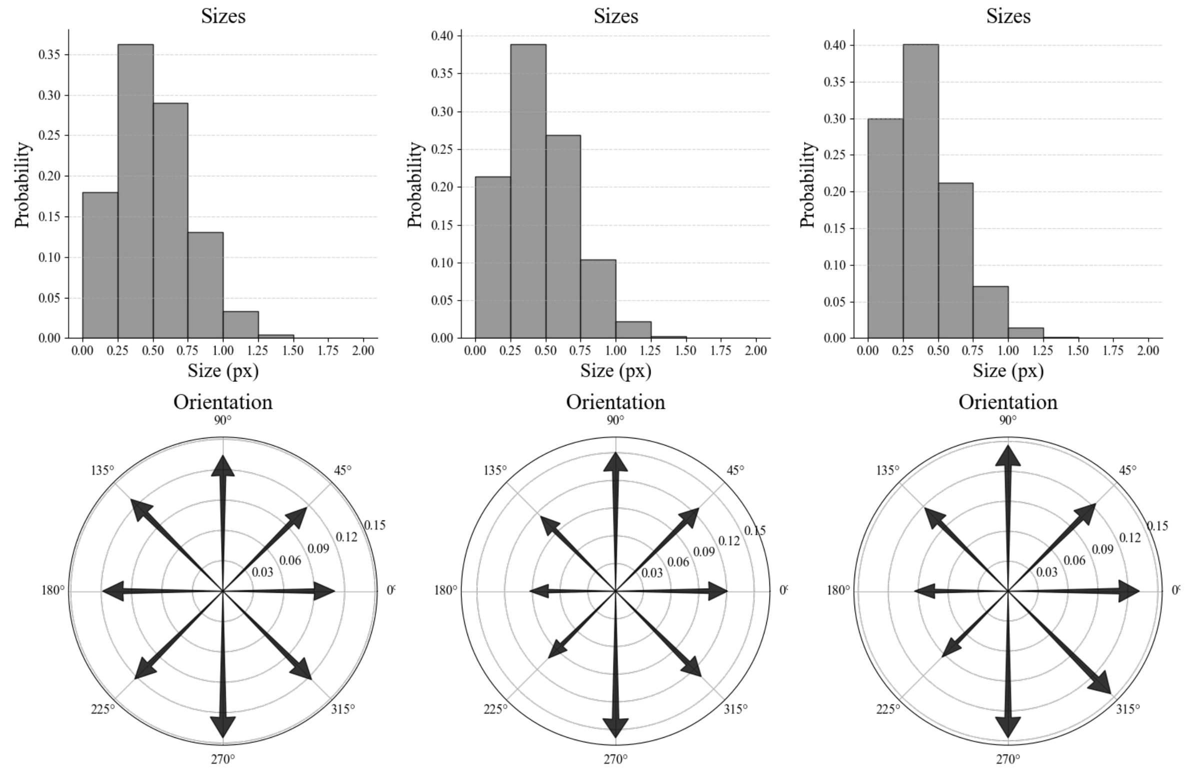


Results and Discussion

Keypoint Refinement Visualization



Offset Distribution Analysis



Relative Pose Estimation (vs. baseline)

Dataset	Method	AUC@5°	AUC@10°	AUC@20°	Inlier Ratio (%)	Mean (°)	Median (°)
MegaDepth	SP + MNN	35.34	45.37	54.22	33.78	27.89	3.94
	Reinforced SP + MNN	36.61	47.56	56.71	33.90	27.91	3.92
	SP + MNN + Ours	**37.16**	**48.07**	**57.15**	**34.42**	**27.69**	**3.82**
KITTI	SP + MNN	80.92	89.49	94.41	73.27	1.20	0.62
	Reinforced SP + MNN	80.97	89.53	94.43	73.27	1.20	0.62
	SP + MNN + Ours	**81.36**	**89.69**	**94.49**	**73.47**	**1.16**	**0.61**
ScanNet	SP + MNN	13.92	28.16	43.11	37.56	**28.58**	9.76
	Reinforced SP + MNN	13.83	28.11	43.17	37.55	28.60	9.76
	SP + MNN + Ours	**14.16**	**28.57**	**43.56**	**37.66**	28.68	**9.59**

Relative Pose Estimation (generalization)

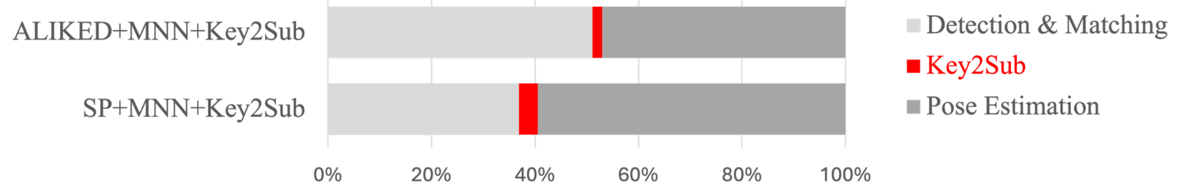
Dataset	Method	AUC@5°	AUC@10°	AUC@20°	Inlier Ratio (%)	Mean (°)	Median (°)
MegaDepth	ALIKED+LG	64.69	76.77	85.17	87.10	6.65	1.01
	ALIKED+LG+Ours	**65.34**	**77.27**	**85.52**	**87.73**	**6.57**	**0.98**
	DeDoDe+DSM	48.83	61.70	71.76	78.08	16.26	1.93
	DeDoDe+DSM+Ours	**52.57**	**65.15**	**74.68**	**81.94**	**15.06**	**1.63**
	DeDoDev2+DSM	55.49	67.37	76.23	82.66	14.73	1.41
	DeDoDev2+DSM+Ours	**56.08**	**67.90**	**76.56**	**83.44**	**14.58**	**1.38**
	Key.Net+MNN	52.76	64.98	74.24	42.18	12.50	1.63
	Key.Net+MNN+Ours	**53.78**	**65.83**	**74.92**	**42.86**	**12.15**	**1.52**
	R2D2+MNN	51.64	64.49	74.26	63.23	11.96	1.72
	R2D2+MNN+Ours	**52.14**	**64.90**	**74.55**	**63.53**	**11.93**	**1.71**
	XFeat+MNN	41.37	52.88	62.42	33.83	18.38	2.85
	XFeat+MNN+Ours	**43.11**	**54.40**	**63.62**	**35.15**	**17.92**	**2.58**
	SP+LG	59.82	72.88	82.31	79.34	7.39	1.25
	SP+LG+Ours	**61.82**	**74.46**	**83.41**	**81.24**	**7.22**	**1.15**
KITTI	ALIKED+LG	84.24	91.45	**95.36**	79.88	1.24	**0.50**
	ALIKED+LG+Ours	**84.27**	**91.46**	**95.36**	**80.02**	**1.20**	**0.50**
	SP+LG	81.50	89.68	94.37	81.28	1.40	0.59
	SP+LG+Ours	**81.99**	**89.95**	**94.51**	**81.59**	**1.34**	**0.57**
	DeDoDe+DSM	71.21	83.01	90.10	67.28	**3.18**	0.88
	DeDoDe+DSM+Ours	**71.88**	**83.39**	**90.29**	**67.29**	**3.18**	**0.85**
ScanNet	ALIKED+LG	20.99	38.68	54.81	70.92	**20.35**	5.86
	ALIKED+LG+Ours	**21.21**	**38.90**	**55.00**	**70.98**	20.39	**5.82**
	SP+LG	19.40	37.96	55.46	85.51	20.07	5.97
	SP+LG+Ours	**19.71**	**38.26**	**55.80**	**85.67**	**19.84**	**5.89**
	DeDoDe+DSM	3.18	6.34	10.16	89.32	**131.2**	**180.0**
	DeDoDe+DSM+Ours	**3.25**	**6.45**	**10.31**	**89.92**	**131.2**	**180.0**

Fundamental Matrix Estimation

Dataset	Method	F1-score	Epi. Inlier Ratio (%)	Epi. Mean	Epi. Med
MegaDepth	SP + MNN	47.64	18.64	**84.15**	1.69
	Reinforced SP + MNN	47.62	18.65	84.32	1.70
	SP + MNN + Ours	**48.42**	**19.64**	84.41	**1.67**
KITTI	SP + MNN	74.15	48.48	0.91	0.81
	Reinforced SP + MNN	74.12	48.47	0.91	0.81
	SP + MNN + Ours	**75.48**	**51.82**	**0.87**	**0.77**
ScanNet	SP + MNN	18.83	15.83	**45.86**	4.96
	Reinforced SP + MNN	18.76	15.82	46.42	4.95
	SP + MNN + Ours	**18.91**	**16.23**	47.11	**4.83**

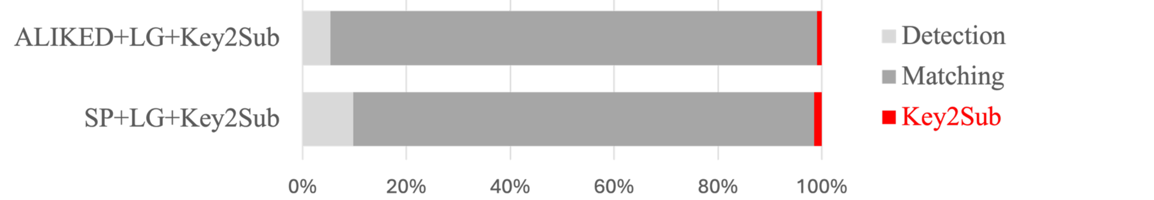
Computational Analysis

Inference Time



Method	Detection&Matching	Ours	Pose Estimation	Total
SP + MNN + Ours	72ms (37%)	7ms (4%)	116ms (60%)	195ms
ALIKED + MNN + Ours	163ms (51%)	6ms (2%)	150ms (47%)	319ms

Parameters



Method	Detection	Matching	Ours	Total
SP+LG+Ours	1.30M (10%)	11.85M (89%)	0.20M (2%)	13.35M
ALIKED+LG+Ours	0.68M (5%)	11.85M (94%)	0.12M (1%)	12.65M