XRefine: Attention-Guided Keypoint Match Refinement

Supplementary Material

Extract+Match	Refinement	AUC5	AUC10	AUC20
ALIKED+LG		10.71	21.16	34.02
ALIKED+LG	Keypt2Subpx	12.32	23.21	36.14
ALIKED+LG	PixSfM	23.23	34.34	45.82
ALIKED+LG	XRefine general	29.31	41.01	52.06
ALIKED+LG	XRefine specific	**29.41**	**41.07**	**52.07**
DISK+LG		62.24	74.13	82.67
DISK+LG	Keypt2Subpx	63.77	75.35	83.53
DISK+LG	PixSfM	63.52	75.14	83.36
DISK+LG	XRefine general	64.13	75.58	83.67
DISK+LG	XRefine specific	**64.93**	**76.27**	**84.20**
DeDoDe2+DSM		35.90	50.22	62.76
DeDoDe2+DSM	Keypt2Subpx	46.58	59.91	70.66
DeDoDe2+DSM	PixSfM	48.45	61.16	71.28
DeDoDe2+DSM	XRefine general	**51.28**	**63.87**	**73.63**
DeDoDe2+DSM	XRefine specific	50.94	63.58	73.42
R2D2+MNN		45.13	57.83	68.02
R2D2+MNN	Keypt2Subpx	46.03	58.61	68.61
R2D2+MNN	PixSfM	46.63	58.77	68.39
R2D2+MNN	XRefine general	47.88	60.01	69.54
R2D2+MNN	XRefine specific	**49.38**	**61.45**	**70.79**

Table 9. Pose estimation result on MegaDepth [17]. Bold indicates best performance and underscores second best per feature.

Extract+Match	Refinement	AUC5	AUC10	AUC20
ALIKED+LG		20.73	38.26	54.84
ALIKED+LG	Keypt2Subpx	20.93	38.31	54.81
ALIKED+LG	PixSfM	20.87	38.27	54.60
ALIKED+LG	XRefine general	21.70	39.37	55.74
ALIKED+LG	XRefine specific	**21.82**	**39.50**	**55.79**
DISK+LG		18.48	34.24	49.12
DISK+LG	Keypt2Subpx	18.77	34.63	49.63
DISK+LG	PixSfM	18.50	33.98	48.98
DISK+LG	XRefine general	19.10	**35.23**	**50.29**
DISK+LG	XRefine specific	**19.39**	35.21	50.17
DeDoDe2+DSM		14.48	28.91	43.80
DeDoDe2+DSM	Keypt2Subpx	16.43	31.64	46.84
DeDoDe2+DSM	PixSfM	16.10	30.93	45.97
DeDoDe2+DSM	XRefine general	**17.21**	32.40	47.61
DeDoDe2+DSM	XRefine specific	17.06	**32.47**	**47.70**
R2D2+MNN		12.14	24.68	38.56
R2D2+MNN	Keypt2Subpx	12.60	25.23	38.98
R2D2+MNN	PixSfM	11.84	24.00	37.59
R2D2+MNN	XRefine general	12.60	25.59	39.46
R2D2+MNN	XRefine specific	**12.63**	**25.72**	**39.57**

Table 10. Pose estimation result on ScanNet [4]. Bold indicates best performance and underscores second best per feature.

6. Results

Additional feature extractors Tables 9 to 11 present relative pose estimation results on MegaDepth [17], ScanNet [4], and KITTI [10] for four additional combinations of feature extractor and matcher: ALIKED [34] with LightGlue (LG) [19] matching, DISK [31] with LightGlue (LG) matching, DeDoDev2 [7] with Double Soft Max (DSM) matching, and R2D2 [26] with Mutual Nearest Neighbor (MNN) matching. Besides for ALIKED+LG on KITTI, where all refinement approaches perform very similarly, XRefine achieves the best performance. Quite striking is the performance improvement that can be achieved for ALIKED+LG on MegaDepth, where the AUC5 without refinement is 10.71% and 29.41% when using XRefine specific.

Comparison of Keypt2Subpx weights Table 12 presents the results that we obtained for Keypt2Subpx when using the original weights provided by the authors (https://github.com/KimSinjeong/keypt2subpx/tree/master/pretrained) and the weights that we obtained using the same training procedure as for XRefine. At this moment in time, from the combinations of feature extractors and matchers that are considered by us, only weights for SuperPoint with LightGlue matching, DeDoDe with Double Soft Max matching, XFeat with Mutual Nearest Neighbor matching, and ALIKED with LightGlue matching are available. We observe very similar results for Keypt2Subpx with the original weights and with our weights. Only in one case (XFeat+MNN on MegaDepth) the AUC5 reached with our weights is more than 0.1 percentage points lower than with the original weights, while in 8 out of the 12 evaluations our weights have slightly higher AUC5 than the original weights. The small differences in performance can be explained by the stochastic nature of the training process.

Varying numbers of keypoints for DeDoDe Similarly to Tab. 7, the Tab. 13 presents the pose estimation performance for varying numbers of extracted keypoints per image, but for DeDoDe [8] features. In our evaluation, the best performance without refinement is reached at 16384 keypoints, while the best performance with refinement is reached at 8192 keypoints. We observe diminishing perfor-

Extract+Match	Refinement	AUC5	AUC10	AUC20
ALIKED+LG		82.14	91.07	95.54
ALIKED+LG	Keypt2Subpx	**84.73**	**91.70**	95.63
ALIKED+LG	PixSfM	84.34	91.52	95.54
ALIKED+LG	XRefine general	84.70	**91.70**	**95.64**
ALIKED+LG	XRefine specific	84.67	91.69	95.63
DISK+LG		84.14	91.47	**95.58**
DISK+LG	Keypt2Subpx	84.48	91.52	95.42
DISK+LG	PixSfM	84.19	91.38	95.35
DISK+LG	XRefine general	84.55	91.57	95.43
DISK+LG	XRefine specific	**84.63**	**91.63**	95.48
DeDoDe2+DSM		83.59	91.22	95.45
DeDoDe2+DSM	Keypt2Subpx	84.16	91.49	95.56
DeDoDe2+DSM	PixSfM	84.15	91.46	95.57
DeDoDe2+DSM	XRefine general	84.44	91.67	95.67
DeDoDe2+DSM	XRefine specific	**84.52**	**91.70**	**95.69**
R2D2+MNN		83.37	90.99	95.25
R2D2+MNN	Keypt2Subpx	83.49	91.12	95.33
R2D2+MNN	PixSfM	84.28	91.51	95.60
R2D2+MNN	XRefine general	84.38	91.57	95.64
R2D2+MNN	XRefine specific	**84.56**	**91.66**	**95.68**

Table 11. Pose estimation result on KITTI [10] odometry. Bold indicates best performance and underscores second best per feature.

mance gains from refinement as the number of keypoints increases. This is expected because, when fewer matches are available, the accuracy of individual keypoint correspondences plays a more critical role in pose estimation. At 32768 keypoints the pose estimation accuracy when using Keypt2Subpx even becomes slightly worse than without refinement. With XRefine, on the other hand, accuracy is still increased, *e.g.* the AUC5 with XRefine specific is about 7% higher than without refinement.

Qualitative results Figure 5 presents visualizations of four refinement examples for our XRefine, Keypt2Subpx [14], and PixSfM [18].

Dataset	Extract+Match	Refinement	AUC5	AUC10	AUC20
MegaDepth	SP+LG		58.48	71.41	80.83
MegaDepth	SP+LG	Keypt2Subpx (original weights)	60.06	72.70	81.78
MegaDepth	SP+LG	Keypt2Subpx (our weights)	60.16	72.73	81.78
MegaDepth	DeDoDe+DSM		34.88	48.64	60.84
MegaDepth	DeDoDe+DSM	Keypt2Subpx (original weights)	44.02	57.27	68.14
MegaDepth	DeDoDe+DSM	Keypt2Subpx (our weights)	44.86	58.08	68.84
MegaDepth	XFeat+MNN		36.45	47.89	57.81
MegaDepth	XFeat+MNN	Keypt2Subpx (original weights)	38.34	49.56	59.15
MegaDepth	XFeat+MNN	Keypt2Subpx (our weights)	38.01	49.06	58.52
MegaDepth	ALIKED+LG		10.71	21.16	34.02
MegaDepth	ALIKED+LG	Keypt2Subpx (original weights)	11.37	22.01	34.87
MegaDepth	ALIKED+LG	Keypt2Subpx (our weights)	12.32	23.21	36.14
ScanNet1500	SP+LG		19.48	37.40	54.79
ScanNet1500	SP+LG	Keypt2Subpx (original weights)	20.11	38.14	55.38
ScanNet1500	SP+LG	Keypt2Subpx (our weights)	20.31	38.21	55.43
ScanNet1500	DeDoDe+DSM		10.13	21.04	32.42
ScanNet1500	DeDoDe+DSM	Keypt2Subpx (original weights)	11.11	22.71	34.92
ScanNet1500	DeDoDe+DSM	Keypt2Subpx (our weights)	11.20	23.02	34.99
ScanNet1500	XFeat+MNN		10.28	22.04	35.77
ScanNet1500	XFeat+MNN	Keypt2Subpx (original weights)	11.33	23.58	37.39
ScanNet1500	XFeat+MNN	Keypt2Subpx (our weights)	11.27	23.32	37.08
ScanNet1500	ALIKED+LG		20.73	38.26	54.84
ScanNet1500	ALIKED+LG	Keypt2Subpx (original weights)	21.01	38.54	55.04
ScanNet1500	ALIKED+LG	Keypt2Subpx (our weights)	20.93	38.31	54.81
KITTI	SP+LG		83.37	90.84	95.12
KITTI	SP+LG	Keypt2Subpx (original weights)	83.59	90.90	95.12
KITTI	SP+LG	Keypt2Subpx (our weights)	83.63	90.93	95.14
KITTI	DeDoDe+DSM		83.98	91.31	95.42
KITTI	DeDoDe+DSM	Keypt2Subpx (original weights)	84.16	91.40	95.49
KITTI	DeDoDe+DSM	Keypt2Subpx (our weights)	84.21	91.42	95.50
KITTI	XFeat+MNN		81.55	89.99	94.79
KITTI	XFeat+MNN	Keypt2Subpx (original weights)	82.21	90.38	94.97
KITTI	XFeat+MNN	Keypt2Subpx (our weights)	82.42	90.47	95.00
KITTI	ALIKED+LG		82.14	91.07	95.54
KITTI	ALIKED+LG	Keypt2Subpx (original weights)	84.77	91.73	95.64
KITTI	ALIKED+LG	Keypt2Subpx (our weights)	84.73	91.70	95.63

Table 12. Comparison of Keypt2Subpx results using the original weights from the authors and our retrained weights.

#KP per image	Refinement	AUC5	AUC10	AUC20
2048		38.54	53.92	67.75
2048	Keypt2Subpx	47.75	62.87	74.68
2048	XRefine specific	**52.34**	**66.65**	**77.56**
2048	XRefine general	52.32	66.38	77.04
4096		45.05	59.90	72.44
4096	Keypt2Subpx	51.65	65.94	77.21
4096	XRefine specific	**55.67**	**69.10**	**79.18**
4096	XRefine general	55.20	68.73	78.79
8192		49.63	63.57	74.75
8192	Keypt2Subpx	52.82	66.62	77.25
8192	XRefine specific	**56.51**	**69.80**	**79.33**
8192	XRefine general	**56.51**	69.29	78.84
16384		51.68	65.03	75.49
16384	Keypt2Subpx	52.66	65.90	76.10
16384	XRefine specific	56.29	**69.06**	**78.29**
16384	XRefine general	**56.32**	68.69	77.99
32768		51.26	63.95	73.53
32768	Keypt2Subpx	50.49	63.62	73.70
32768	XRefine specific	**54.95**	**67.13**	**76.08**
32768	XRefine general	54.69	66.91	75.81

Table 13. Results for varying numbers of extracted keypoints (KPs) per image on MegaDepth1500 [17] with DeDoDe [8] features and double soft max matching.

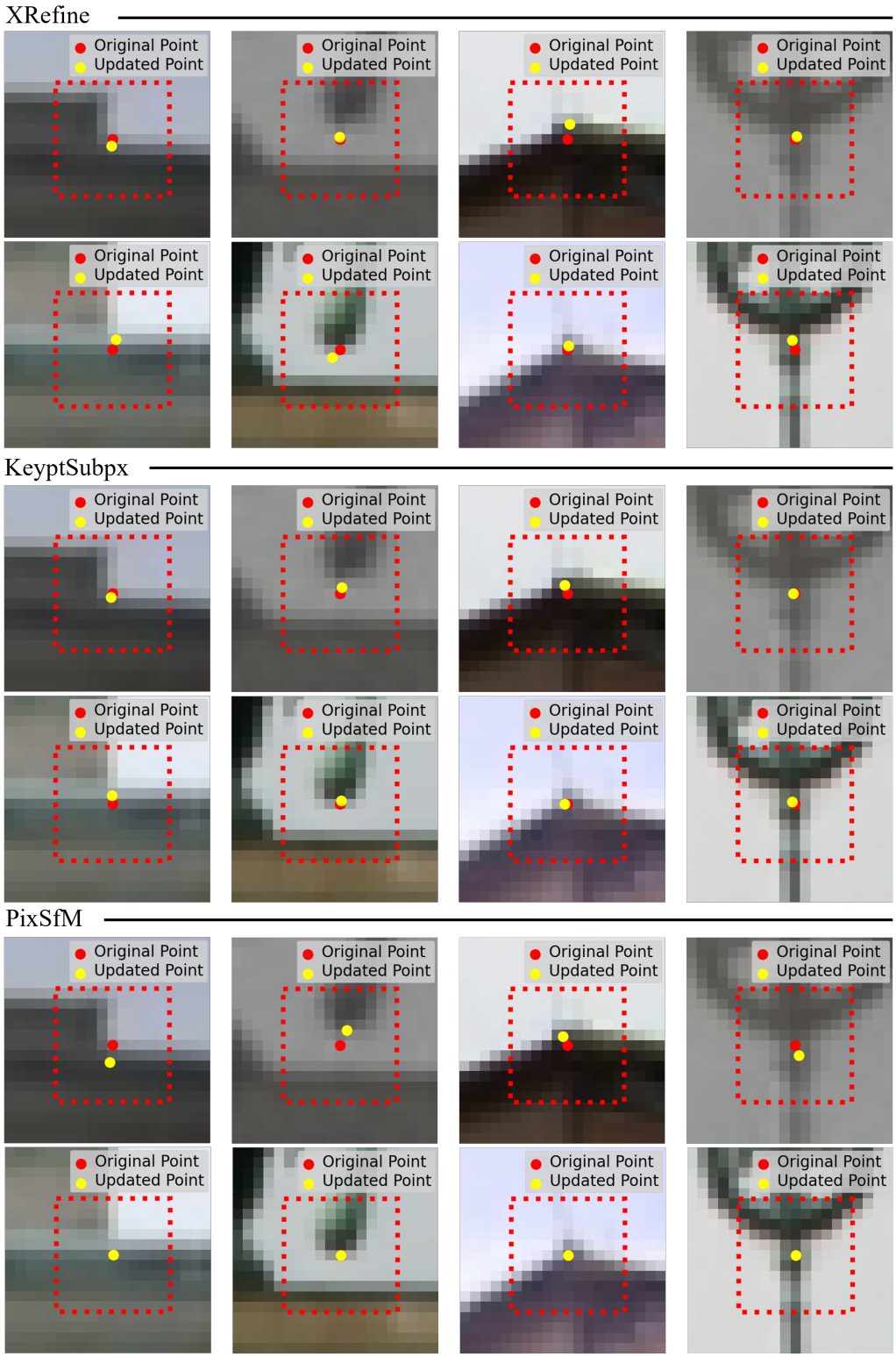

Figure 5. Example keypoint refinements for XRefine (top two rows), Keypt2Subpx (middle two rows), and PixSfM (bottom two rows). Keypoints are extracted from MegaDepth, using SuperPoint and LightGlue. Each column represents the extracted patches for a given pair of matched keypoints. The same four extracted keypoint matches are refined by the three refinement methods. The presented patches have a size of 21×21 pixel, while the 11×11 area that is given as input to XRefine and Keypt2Subpx is highlighted by the red dotted square.