Supplementary Material
BALF: Simple and Efficient Blur Aware Local Feature Detector

Zhenjun Zhao
The Chinese University of Hong Kong
ericzzj89@gmail.com

1. Architecture Details

The detailed specifications of our proposed network architecture for implementation are shown in Tab. 1. We also specify the input and output size of each block and layer. Here MLP_3_32 means channel MLP block with 3 input and 32 output channels respectively. MAB denotes multi-axis gated MLP block [23], and RMAB is our proposed residual MLP attention block.

Stage	Input size	Output size	Layers/Blocks
1	$256^2 \times 3$	$256^2 \times 32$	{ MLP_3_32, ReLU
1	$256^2 \times 32$	$256^2 \times 32$	{ MAB, RMAB
1	$256^2 \times 32$	$128^2 \times 32$	Pooling
2	$128^2 \times 32$	$128^2 \times 64$	{ MLP_32_64, ReLU
2	$128^2 \times 64$	$128^2 \times 64$	{ MAB, RMAB
2	$128^2 \times 64$	$64^2 \times 64$	Pooling
3	$64^2 \times 64$	$64^2 \times 128$	{ MLP_64_128, ReLU
3	$64^2 \times 128$	$64^2 \times 128$	{ MAB, RMAB
3	$64^2 \times 128$	$32^2 \times 128$	Pooling
4	$32^2 \times 128$	$32^2 \times 256$	{ MLP_128_256, ReLU
5	$32^2 \times 256$	$32^2 \times 64$	{ MLP_256_64, BatchNorm
6	$32^2 \times 64$	$256^2 \times 1$	{ Channel-wise softmax, Reshape

Table 1. **Detailed architecture specifications of BALF framework.** Stage 1-3 denote MLP-based encoder, while stage 4-6 correspond detection module in our proposed BALF framework.

2. Additional Ablation Study

Number of MLPCoder block. Towards understanding the BALF framework, we scaled up our architecture in terms of the number of MLPCoder. Tab. 2 suggests that using more than three MLPCoder blocks does not significantly improve the detection performance, but increases the number of parameters and computational cost. We thus use 3 MLPCoder blocks in our experiments to yield the performance and complexity tradeoff.

Num. MLPCoder block	Repeatability ↑	Params ↓	Inference time ↓
1	63.60%	23K	6.89ms
2	66.82%	111K	12.54ms
3	75.15%	381K	29.02ms
4	78.27%	1396K	112.20ms

Table 2. **Number of MLPCoder block.** The inference time here is the runtime of keypoint extraction at a VGA resolution image (*i.e.* 480×640 pixels).

Different architectures. We also re-train some classical architectures like ResNet-18 [7], VGG-16 [21], and U-Net [18] with the proposed detection module and loss function. Tab. 3 presents the performance of different architectures. Since the memory required by ViT [5] exceeds NVIDIA Geforce 2080 Ti, we did not re-train ViT in this ablation. The results demonstrate that our proposed MLP-based encoder achieves superior repeatability performance compared to these classical architectures.

Variant	Repeatability ↑	Params ↓	Inference time ↓
ResNet-18 [7]	67.90%	746K	55.12ms
VGG-16 [21]	68.52%	338K	5.25ms
U-Net [18]	67.90%	315K	3.10ms
MLP-based encoder (proposed)	75.15%	381K	29.02ms

Table 3. **Different architectures.** The performance of different network architectures on the GoPro testing dataset.

3. Complete Quantitative Results

Due to limited space in the main paper, we only present the total repeatability performance in the evaluations with blur and deblur data. We thus show the complete quantitative evaluation results here.

Evaluation with the Blur-HPatches dataset. Tabs. 4 and 5 present the average repeatability score on the viewpoint changes, illumination changes and all image sequences together with three varying levels of motion blur under blur-to-sharp and blur-to-blur configurations respectively.

Evaluation with the Blur-HPatches dataset preprocessed by deblurring network. Tabs. 6 to 9 present the complete repeatability results among all other methods on deblurred images and ours on corresponding blurred images.

4. More Qualitative Results

Detection. Figs. 1 and 2 present detection qualitative results on blurred images from RWBI dataset [25], which are captured by real cameras. It demonstrates that our network cannot only detect well distributed salient keypoints from sharp images, but also being able to detect well localized keypoints from bluured images.

Matching. To further demonstrate the performance of our method on the real blurred images, we randomly select a paired sharp and blurred images, and another sharp image (with viewpoint changes) from RealBlur dataset [17] for feature matching evaluation. Specifically, we first run our detector on the paired sharp and blurred images, and extract the correspondences between them by a pre-defined range, such as those within a circle. We then compute the correspondences between those two sharp images (with viewpoint changes), using HardNet descriptor [13] and FLANN [14] matching. Finally, we can establish the correspondences between the blurred and the second sharp images via the above two sets of correspondences. Fig. 3 demonstrates that our method can detect well localized and repeatable keypoints from both sharp and blurred images for further image matching.

References

[1] Pablo Fernández Alcantarilla, Adrien Bartoli, and Andrew J. Davison. Kaze features. In *ECCV*, 2012. 3, 4, 5, 6, 7

[2] Pablo Fernández Alcantarilla, Jesús Nuevo, and Adrien Bartoli. Fast explicit diffusion for accelerated features in nonlinear scale spaces. In *BMVC*, 2013. 3, 4, 5, 6, 7

[3] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *ECCV*, 2006. 3, 4, 5, 6, 7

[4] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *CVPRW*, pages 337–33712, 2018. 3, 4, 5, 6, 7

[5] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2020. 1

[6] Mihai Dusmanu, Ignacio Rocco, Tomás Pajdla, Marc Pollefeys, Josef Sivic, Akihiko Torii, and Torsten Sattler. D2-net: A trainable cnn for joint description and detection of local features. In *CVPR*, pages 8084–8093, 2019. 3, 4, 5, 6, 7

[7] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 1

[8] Orest Kupyn, T. Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *ICCV*, pages 8877–8886, 2019. 5

[9] Axel Barroso Laguna, Edgar Riba, Daniel Ponsa, and Krystian Mikolajczyk. Key.net: Keypoint detection by handcrafted and learned cnn filters. In *ICCV*, pages 5835–5843, 2019. 3, 4, 5, 6, 7

[10] G LoweDavid. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004. 3, 4, 5, 6, 7

[11] Jiri Matas, Ondřej Chum, Martin Urban, and Tomás Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image Vis. Comput.*, 22:761–767, 2004. 3, 4, 5, 6, 7

[12] Krystian Mikolajczyk and Cordelia Schmid. Scale & affine invariant interest point detectors. *IJCV*, 60:63–86, 2004. 3, 4, 5, 6, 7

[13] Anastasiya Mishchuk, Dmytro Mishkin, Filip Radenović, and Jiri Matas. Working hard to know your neighbor's margins: Local descriptor learning loss. In *NeurIPS*, 2017. 2

[14] Marius Muja and David G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *VISAPP*, 2009. 2

[15] Yuki Ono, Eduard Trulls, Pascal V. Fua, and Kwang Moo Yi. Lf-net: Learning local features from images. In *NeurIPS*, 2018. 3, 4, 5, 6, 7

[16] Jérôme Revaud, Philippe Weinzaepfel, César Roberto de Souza, No'e Pion, Gabriela Csurka, Yohann Cabon, and M. Humenberger. R2d2: Repeatable and reliable detector and descriptor. In *NeurIPS*, 2019. 3, 4, 5, 6, 7

[17] Jaesung Rim, Hoon Sung Chwa, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *ECCV*, 2020. 2, 8

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. 2015. 1

[19] Edward Rosten, Reid B. Porter, and Tom Drummond. Faster and better: A machine learning approach to corner detection. *IEEE TPAMI*, 32:105–119, 2010. 3, 4, 5, 6, 7

[20] Jianbo Shi and Carlo Tomasi. Good features to track. In *CVPR*, pages 593–600, 1994. 3, 4, 5, 6, 7

[21] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015. 1

[22] Xin Tao, Hongyun Gao, Yi Wang, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *CVPR*, pages 8174–8182, 2018. 4

Mehotd	Easy Viewpoint ↑	Easy Illumination ↑	Easy Total ↑	Hard Viewpoint ↑	Hard Illumination ↑	Hard Total ↑	Tough Viewpoint ↑	Tough Illumination ↑	Tough Total ↑
SIFT [10]	50.64	61.39	55.92	51.55	62.23	56.80	47.36	59.83	53.49
SURF [3]	55.77	62.09	58.88	50.98	61.66	56.23	49.36	<u>63.37</u>	56.24
Harris-Laplace [12]	16.17	57.65	36.70	17.46	58.77	37.97	18.43	51.70	34.98
Shi-Tomasi [20]	57.33	57.33	57.33	55.18	55.05	55.11	48.13	50.12	49.11
MSER [11]	43.27	45.15	44.19	40.43	43.57	41.97	34.56	39.62	37.05
KAZE [1]	46.76	53.15	49.90	43.30	50.48	46.84	34.64	45.47	39.98
AKAZE [2]	46.57	62.00	54.15	42.46	58.94	50.51	33.64	57.90	45.49
FAST [19]	60.89	63.10	61.98	61.01	62.56	61.77	47.93	54.93	51.37
LIFT [24]	46.47	55.06	50.69	45.71	54.78	50.17	42.13	52.02	46.99
Key.Net [9]	58.02	62.74	60.34	50.54	59.02	54.71	39.20	50.38	44.69
SuperPoint [4]	<u>66.06</u>	65.21	<u>65.64</u>	<u>61.96</u>	62.49	<u>62.22</u>	51.60	54.13	52.84
LF-Net [15]	59.57	<u>67.65</u>	63.54	57.64	<u>64.87</u>	61.19	<u>52.14</u>	61.57	<u>56.78</u>
D2-Net [6]	44.30	55.31	49.71	41.60	53.19	47.30	38.39	50.47	44.32
R2D2 [16]	55.10	60.99	57.99	47.37	56.25	51.73	33.46	47.92	40.57
BALF (ours)	**72.58**	**75.74**	**74.12**	**72.93**	**76.07**	**74.45**	**67.26**	**76.54**	**71.84**

Table 4. **Repeatability results (%) on Blur-HPatches dataset under blur-to-sharp configuration.** Our method achieves best performance compared to prior works on the viewpoint changes, illumination changes, and all image sequences together with three varying levels of motion blur.

Method	Easy Viewpoint ↑	Easy Illumination ↑	Easy Total ↑	Hard Viewpoint ↑	Hard Illumination ↑	Hard Total ↑	Tough Viewpoint ↑	Tough Illumination ↑	Tough Total ↑
SIFT [10]	56.67	57.31	56.99	53.15	53.85	53.49	46.23	46.63	45.94
SURF [3]	58.46	63.80	61.08	55.61	60.55	58.04	49.92	57.41	53.60
Harris-Laplace [12]	17.09	54.82	35.76	15.08	29.06	31.95	11.44	43.79	27.47
Shi-Tomasi [20]	58.16	54.35	56.29	54.72	52.73	53.75	<u>51.61</u>	51.13	51.37
MSER [11]	41.11	42.53	41.81	37.21	39.30	38.24	32.63	36.61	34.59
KAZE [1]	<u>63.31</u>	63.27	63.29	58.66	58.76	58.71	45.94	47.88	46.90
AKAZE [2]	61.81	68.63	<u>65.16</u>	<u>59.28</u>	65.24	<u>62.20</u>	48.22	55.06	51.54
FAST [19]	57.81	57.87	57.84	52.70	54.03	53.35	49.70	52.62	51.17
LIFT [24]	46.08	50.69	48.34	43.74	49.51	46.57	42.87	50.31	46.53
Key.Net [9]	61.81	63.77	62.77	56.82	59.57	58.17	45.68	52.94	49.25
SuperPoint [4]	58.01	59.22	58.60	49.69	50.37	50.03	42.34	44.25	43.28
LF-Net [15]	52.45	<u>68.74</u>	60.45	51.20	<u>67.21</u>	59.07	49.68	<u>66.02</u>	<u>57.71</u>
D2-Net [6]	46.65	57.14	51.80	45.84	56.44	51.05	45.11	56.13	50.53
R2D2 [16]	54.10	61.00	57.49	51.87	58.87	55.31	40.88	53.05	46.86
BALF (ours)	**69.44**	**71.56**	**70.48**	**67.13**	**70.22**	**68.43**	**65.90**	**69.60**	**67.71**

Table 5. **Repeatability results (%) on Blur-HPatches dataset under blur-to-blur configuration.** Our method also achieves best performance compared to prior works on the viewpoint changes, illumination changes, and all image sequences together with three varying levels of motion blur.

[23] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Conrad Bovik, and Yinxiao Li. Maxim: Multi-axis mlp for image processing. In *CVPR*, pages 5759–5770, 2022. 1

[24] Kwang Moo Yi, Eduard Trulls, Vincent Lepetit, and Pascal V. Fua. Lift: Learned invariant feature transform. In *ECCV*, 2016. 3, 4, 5, 6, 7

[25] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Bjorn Stenger, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In *CVPR*, pages 2737–2746, 2020. 2, 6, 7

Method	Easy			Hard			Tough		
	Viewpoint ↑	Illumination ↑	Total ↑	Viewpoint ↑	Illumination ↑	Total ↑	Viewpoint ↑	Illumination ↑	Total ↑
SIFT [10]	53.98	59.35	56.62	52.44	58.38	55.36	47.95	59.92	53.83
SURF [3]	60.84	62.99	61.89	57.39	60.92	59.13	49.89	60.04	54.88
Harris-Laplace [12]	14.74	19.62	17.15	14.89	18.89	16.87	17.95	23.17	20.54
Shi-Tomasi [20]	61.05	60.05	60.56	57.12	56.61	56.87	47.63	49.96	48.78
MSER [11]	46.74	46.55	46.65	42.67	43.80	43.23	35.95	39.92	37.90
KAZE [1]	65.37	64.90	65.14	62.55	63.67	63.10	56.73	63.71	60.16
AKAZE [2]	63.21	68.94	66.03	60.39	67.77	64.02	53.45	68.14	60.64
FAST [19]	60.96	62.62	61.77	58.61	60.75	59.67	58.67	64.61	61.60
LIFT [24]	52.12	57.93	54.98	49.16	56.25	52.64	41.75	51.93	46.75
Key.Net [9]	62.86	63.72	63.28	56.37	59.71	58.01	42.08	52.30	47.10
SuperPoint [4]	68.65	66.76	67.72	65.33	62.73	64.05	54.18	56.37	55.26
LF-Net [15]	54.40	70.31	62.22	52.52	67.53	59.90	47.77	61.93	54.73
D2-Net [6]	46.72	57.08	51.81	44.30	54.87	49.49	40.00	52.07	45.94
R2D2 [16]	58.36	62.31	60.31	52.58	58.38	55.43	37.40	49.32	43.26
BALF (ours)	**72.58**	**75.74**	**74.12**	**72.93**	**76.07**	**74.45**	**67.26**	**76.54**	**71.84**

Table 6. **Repeatability results (%) on deblurred images from SRN-DeblurNet [22] under deblur-to-sharp configuration.** The bottom row shows the results of our method on the corresponding blurred images.

Method	Easy			Hard			Tough		
	Viewpoint ↑	Illumination ↑	Total ↑	Viewpoint ↑	Illumination ↑	Total ↑	Viewpoint ↑	Illumination ↑	Total ↑
SIFT [10]	60.31	59.16	59.75	57.98	58.27	58.13	48.56	52.78	50.63
SURF [3]	61.33	63.58	62.44	59.84	62.74	61.26	51.76	58.90	55.27
Harris-Laplace [12]	14.88	59.55	36.98	15.22	56.66	35.73	16.37	48.31	32.23
Shi-Tomasi [20]	66.13	60.13	63.18	63.75	58.78	61.03	51.13	50.63	50.88
MSER [11]	48.52	46.85	47.70	46.11	44.66	45.40	36.33	38.70	37.49
KAZE [1]	64.87	63.51	64.20	62.78	62.12	62.45	51.57	55.31	53.41
AKAZE [2]	63.35	68.15	65.71	61.30	66.96	64.08	51.55	60.85	56.10
FAST [19]	63.74	61.67	62.72	61.86	60.40	61.14	49.11	52.16	50.61
LIFT [24]	53.88	57.95	55.88	52.12	55.22	53.64	42.85	47.87	45.31
Key.Net [9]	62.71	63.01	62.86	59.69	61.22	60.44	46.57	55.06	50.74
SuperPoint [4]	67.77	64.93	66.38	64.61	61.67	63.16	49.04	50.02	49.52
LF-Net [15]	55.36	71.03	63.06	54.60	69.72	62.03	49.17	65.68	57.28
D2-Net [6]	49.05	58.32	53.60	48.58	57.57	53.00	45.70	56.33	50.93
R2D2 [16]	55.71	60.60	58.11	52.30	57.38	54.80	40.83	50.88	45.77
BALF (ours)	**69.44**	**71.56**	**70.48**	**67.13**	**70.22**	**68.43**	**65.90**	**69.60**	**67.71**

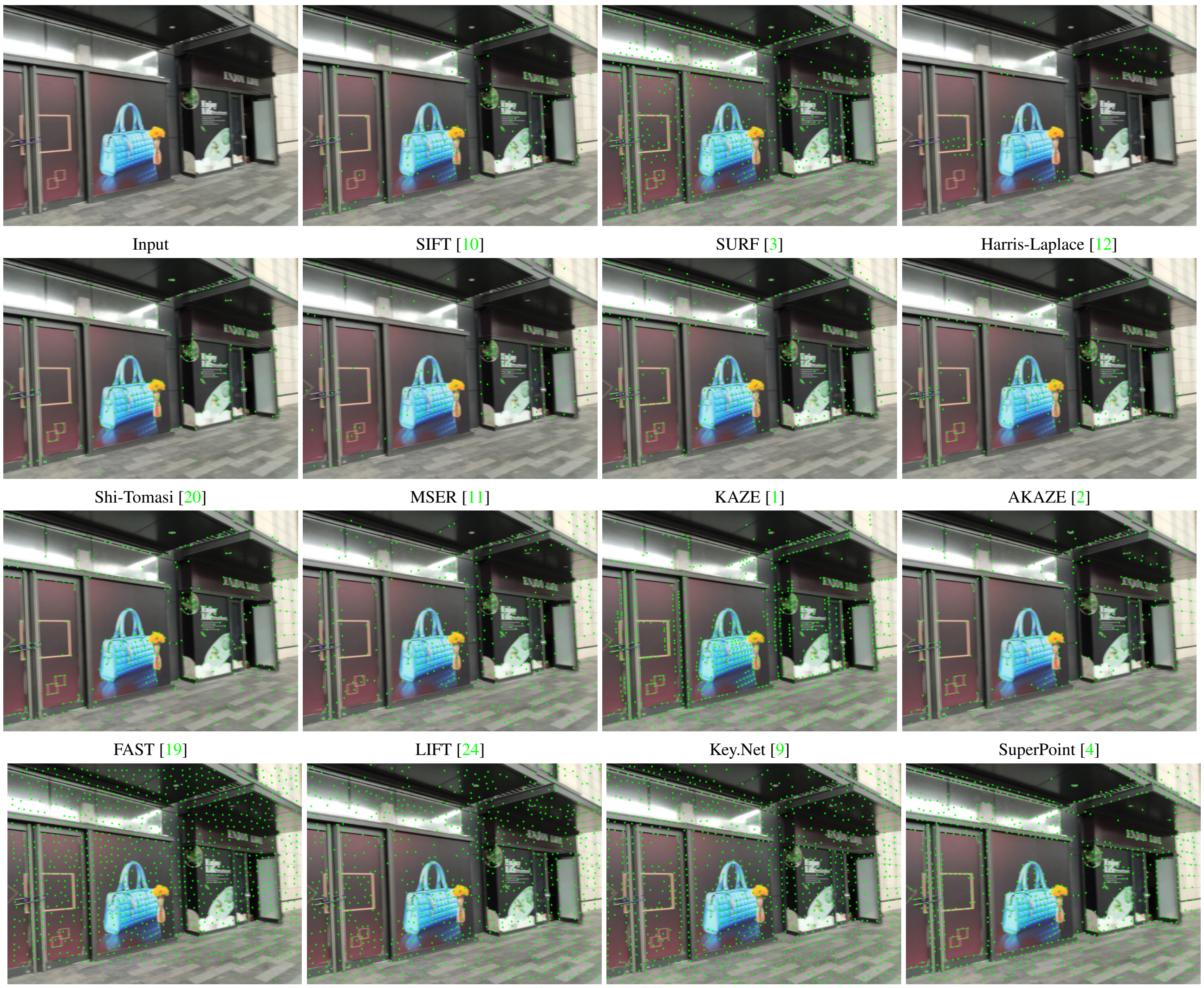
Table 7. **Repeatability results (%) on deblurred images from SRN-DeblurNet [22] under deblur-to-deblur configuration.** The bottom row shows the results of our method on the corresponding blurred images.

Method	Easy			Hard			Tough		
	Viewpoint ↑	Illumination ↑	Total ↑	Viewpoint ↑	Illumination ↑	Total ↑	Viewpoint ↑	Illumination ↑	Total ↑
SIFT [10]	56.28	59.03	57.63	54.02	59.10	56.52	51.29	61.88	56.50
SURF [3]	60.98	62.99	61.97	57.28	61.93	59.57	51.48	61.38	56.34
Harris-Laplace [12]	14.08	19.35	16.69	14.75	19.10	16.90	17.84	22.69	20.24
Shi-Tomasi [20]	62.98	60.50	61.75	59.32	58.88	59.10	51.29	51.83	51.56
MSER [11]	47.98	47.25	47.62	44.90	45.40	45.14	40.12	41.30	40.70
KAZE [1]	65.42	65.04	65.23	62.27	64.12	63.18	58.62	64.27	61.41
AKAZE [2]	63.63	69.04	66.29	60.44	68.72	64.50	56.07	69.65	62.72
FAST [19]	61.89	62.12	62.00	59.41	61.51	60.44	57.01	60.53	58.74
LIFT [24]	54.88	58.37	56.59	50.94	56.24	53.54	45.60	52.70	49.09
Key.Net [9]	63.78	64.20	63.99	57.46	60.91	59.16	44.92	53.92	49.35
SuperPoint [4]	68.85	67.02	67.95	66.83	64.84	65.86	57.37	59.10	58.22
LF-Net [15]	55.01	70.43	62.59	53.03	67.70	60.24	47.73	62.14	54.81
D2-Net [6]	47.68	57.77	52.64	44.96	55.64	50.21	40.17	51.80	45.88
R2D2 [16]	58.60	62.38	60.46	52.72	58.74	55.68	40.55	50.38	45.38
BALF (ours)	**72.58**	**75.74**	**74.12**	**72.93**	**76.07**	**74.45**	**67.26**	**76.54**	**71.84**

Table 8. **Repeatability results (%) on deblurred images from DeblurGAN-v2 [8] under deblur-to-sharp configuration.** The bottom row shows the results of our method on the corresponding blurred images.

Method	Easy			Hard			Tough		
	Viewpoint ↑	Illumination ↑	Total ↑	Viewpoint ↑	Illumination ↑	Total ↑	Viewpoint ↑	Illumination ↑	Total ↑
SIFT [10]	59.17	59.73	59.44	57.33	58.66	57.98	50.12	52.33	51.21
SURF [3]	60.57	63.63	62.07	58.95	62.74	60.81	51.82	58.47	55.09
Harris-Laplace [12]	14.49	60.16	37.09	14.73	57.67	35.97	15.02	48.35	31.54
Shi-Tomasi [20]	66.02	61.05	63.58	64.08	59.62	61.89	54.59	52.90	53.76
MSER [11]	48.43	47.23	47.84	45.93	45.18	45.56	37.45	38.59	38.01
KAZE [1]	64.77	63.47	64.13	61.92	61.82	61.87	52.93	55.48	54.19
AKAZE [2]	63.62	67.94	65.75	61.03	66.57	63.75	53.93	60.90	57.35
FAST [19]	63.86	62.92	63.40	62.16	61.23	61.70	54.60	56.30	55.43
LIFT [24]	54.44	58.99	56.68	53.01	57.70	55.31	46.95	54.05	50.44
Key.Net [9]	62.38	63.08	62.73	59.39	61.81	60.58	49.12	56.93	52.96
SuperPoint [4]	67.49	65.49	66.50	64.89	62.48	63.71	51.85	52.35	52.09
LF-Net [15]	55.57	70.69	63.00	54.58	69.25	61.79	50.48	65.49	57.85
D2-Net [6]	49.40	58.62	53.93	48.65	58.08	53.29	45.69	55.97	50.74
R2D2 [16]	55.66	60.33	57.95	52.31	57.84	55.03	43.12	52.76	47.86
BALF (ours)	**69.44**	**71.56**	**70.48**	**67.13**	**70.22**	**68.43**	**65.90**	**69.60**	**67.71**

Table 9. **Repeatability results (%) on deblurred images from DeblurGAN-v2 [8] under deblur-to-deblur configuration.** The bottom row shows the results of our method on the corresponding blurred images.



Input | SIFT [10] | SURF [3] | Harris-Laplace [12]
Shi-Tomasi [20] | MSER [11] | KAZE [1] | AKAZE [2]
FAST [19] | LIFT [24] | Key.Net [9] | SuperPoint [4]
LF-Net [15] | D2-Net [6] | R2D2 [16] | BALF (ours)

Figure 1. **Qualitative results for keypoint detection on RWBI dataset [25].** Our method generates more accurate and consistent keypoints. Best viewd in high resolution.

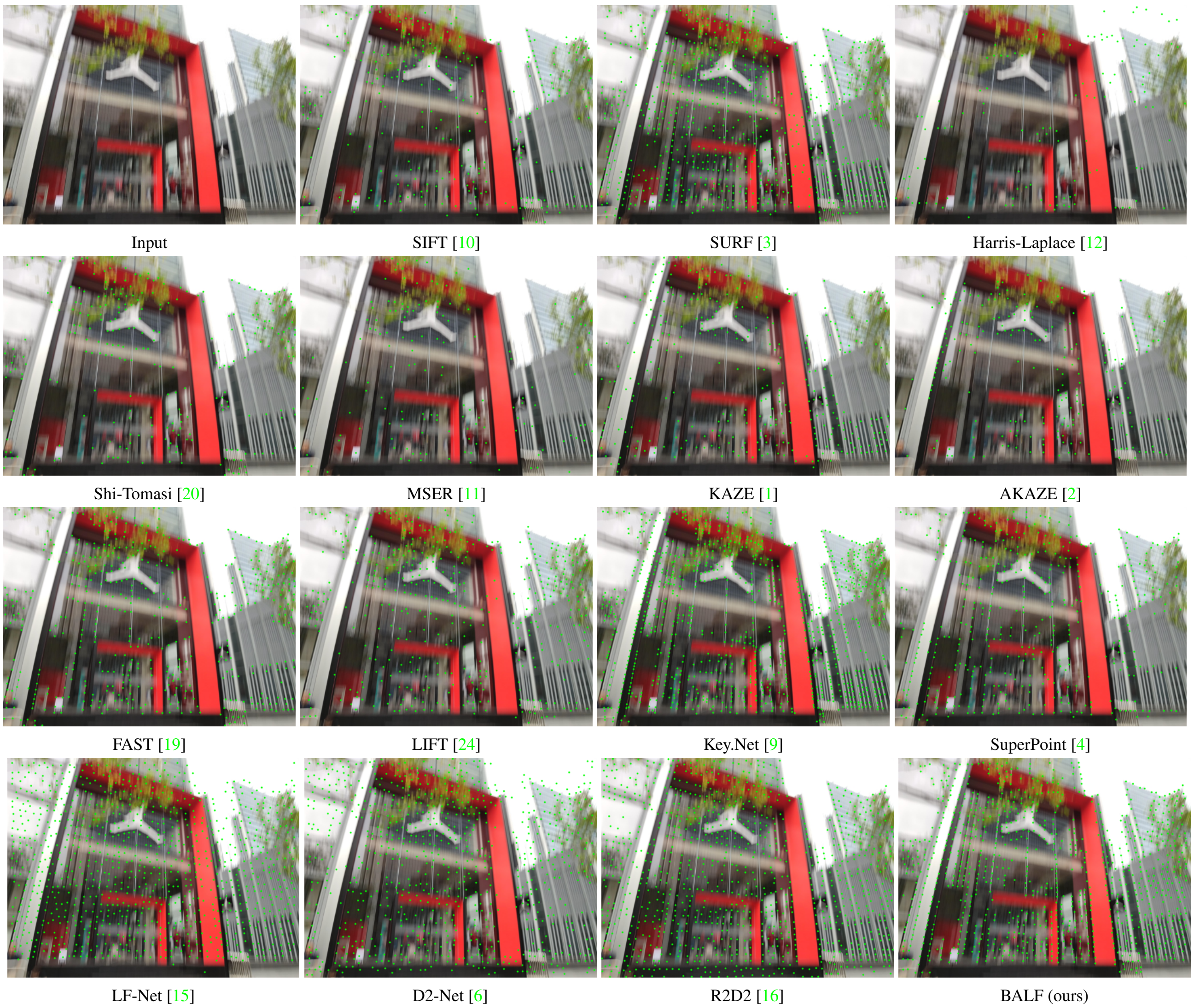


Figure 2. **Qualitative results for keypoint detection on RWBI dataset [25].** Our method generates more accurate and consistent keypoints. Best viewd in high resolution.

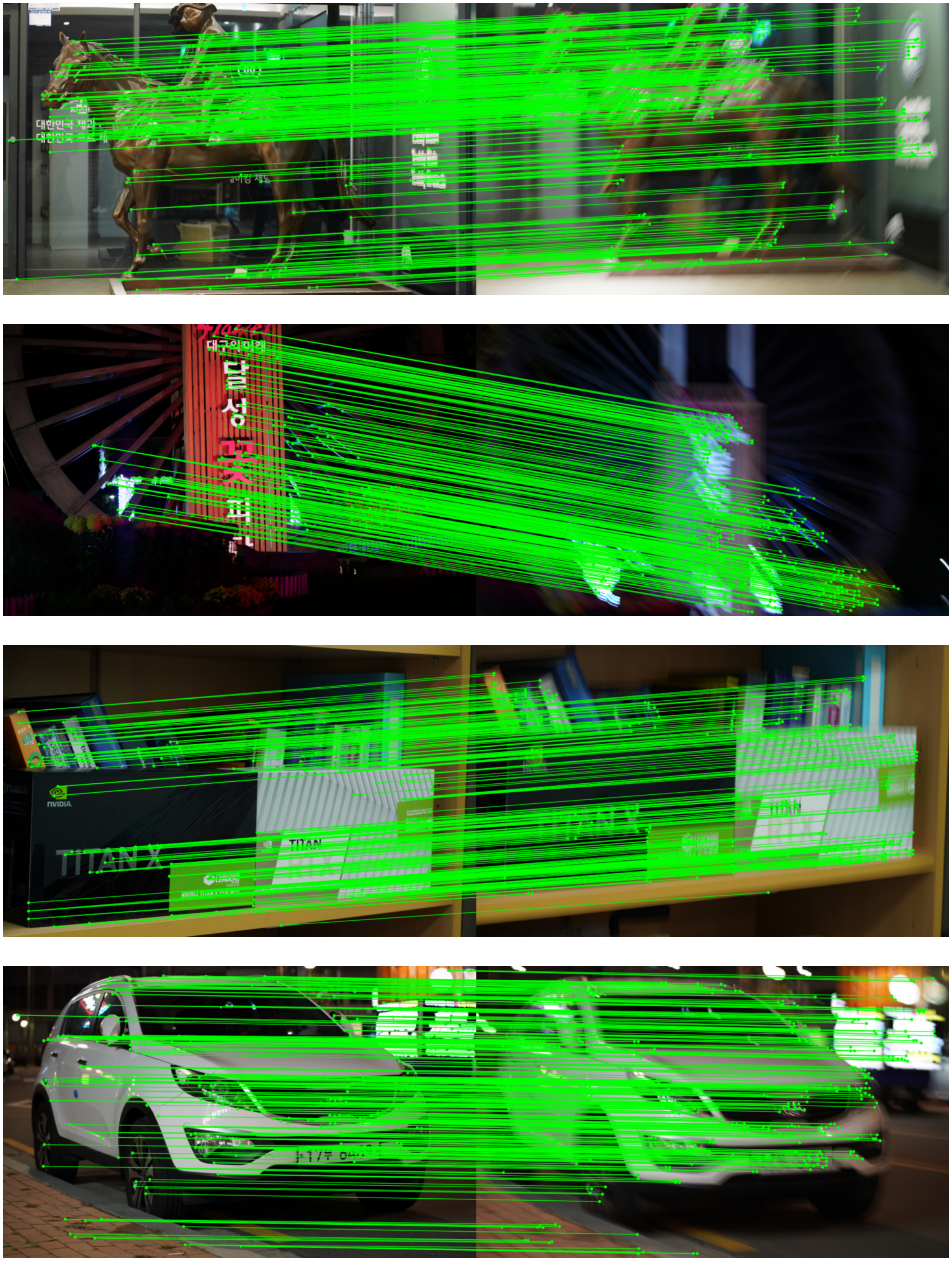


Figure 3. **Qualitative results for keypoint detection and matching on RealBlur dataset [17]. Left** Sharp image. **Right** Blurred image. Our network is able to detect well distributed and localized keypoints from both sharp and blurred images for further image matching. Best viewd in high resolution.