Supplementary Materials to "Reliable and Efficient Image Cropping: A Grid Anchor based Approach"

Hui Zeng[1] Lida Li[1] Zisheng Cao[2] Lei Zhang[1]
[1]The Hong Kong Polytechnic University [2]DJI Co.,Ltd
{cshzeng, cslli}@comp.polyu.edu.hk, zisheng.cao@dji.com, cslzhang@comp.polyu.edu.hk

This supplementary file provides the following materials:

(1) the interface of our annotation toolbox;

(2) more statistical information of the GAICD;

(3) performance by using different numbers of training images;

(4) quantitative comparison on previous cropping databases;

(5) more qualitative comparison of different methods.

1. Annotation toolbox

The interface of our annotation toolbox is shown in Fig. 1. Each time, it displays one source image on the left side and 4 crops generated from it on the right side. The crops are displayed in ordered aspect ratio to alleviate the influence of dramatic changes of aspect ratio on human perception. Specifically, we choose six common aspect ratios (including 16:9, 3:2, 4:3, 1:1, 3:4 and 9:16) and group crops into six sets based on their closest aspect ratios. The top-right corner displays the approximate aspect ratio of current crops. Two horizontal and two vertical guidelines can be optionally used to assist judgement during the annotation. For each crop, we provide five scores (from 1 to 5, representing "bad," "poor," "fair," "good," and "excellent" crops) to rate by annotators. The annotators can either scroll their mouse or click the "Previous" and "Next" buttons to change page. In the bottom-left of the interface, we show the score distribution of rated crops for the current image as a reference for annotators. The bottom-right corner shows the progress of the annotation and the elapsed time.

2. Statistics of the GAICD

Our GAICD contains a total of 106,860 annotated crops from 1,236 images. Each image contains on average 86.5 crops and each crop was scored by 7 different annotators, who are are either experienced photographers from photography communities or senior students from the art department of two universities. The histograms of the MOS and standard deviation are plotted in Fig. 2, and some statistical details are summarized in Table 1. It can be seen that most crops have ordinary or poor quality, while about 10% crops have MOS larger than 4. Regarding to the standard deviation, only 5.75% crops are larger than 1, which indicates the consistency of annotations under our grid anchor based formulation.

3. Results by using different numbers of training images

This part presents the results by using five different numbers of training images, including 200, 400, 600, 800 and 1000. Note that each image contains on average 86.5 annotated candidate crops in our database. The testing set, which contains 200 images, was fixed for all cases. The two average accuracy metrics ($\overline{Acc_5}$ and $\overline{Acc_{10}}$) and the average SRCC ($\overline{SRCC}$) are plotted in Fig. 3. The three curves clearly show that the performance constantly increases with the number of training images, indicating that training with more annotated images could improve the model accuracy.

4. Results on previous databases

We also evaluated our trained model on the ICDB [7] and FCDB [1] using the IoU as metric. Since some groundtruth crops on these two databases have uncommon aspect ratios, we did not employ the aspect ratio constraint when generating

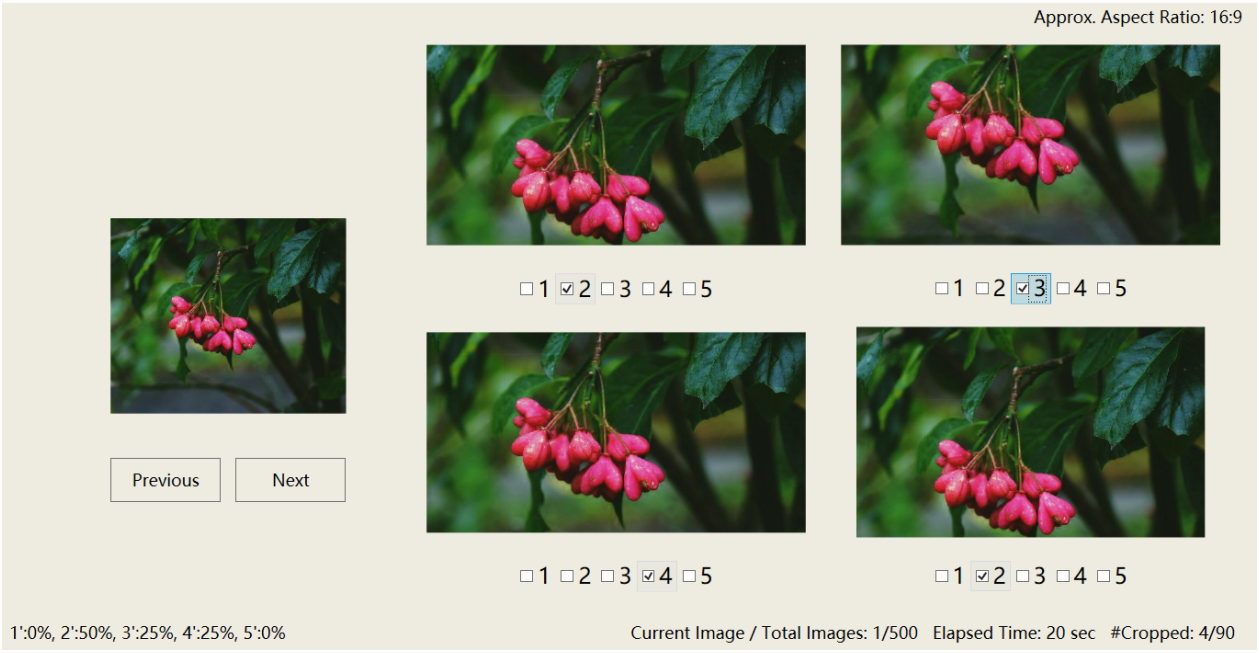

Figure 1: Interface of our annotation toolbox.

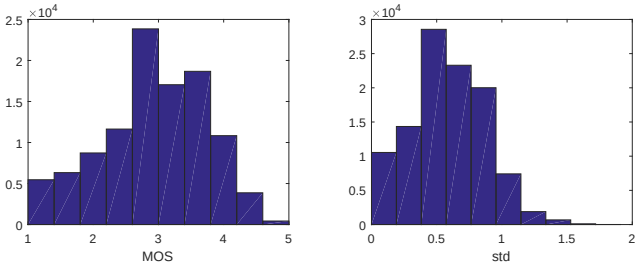

Figure 2: Histograms of the MOS and standard deviation on the GAICD.

	MOS				std.	
range	1-2	2-3	3-4	4-5	≤ 1	>1
ratio (%)	13.59	27.28	49.20	9.93	94.25	5.75

Table 1: Statistics of the MOS and standard deviation on the GAICD.

candidate crops. In practice, we found that the value of λ, which is defined in Eq.(1) as one content preservation constraint in our main paper, affects the performance on the ICDB. The results of our model using five different values of λ to generate crops are reported in Table 2. As can be seen, like most previous methods, our model also obtains comparable or even smaller IoU than the baselines on these two databases. Since the groundtruth crops in the ICDB have large overlap with the source image, using a large λ can directly discard the candidate crops which have having small IoU with the source image, thus improving the IoU on this database.

In contrast, as shown in the main paper, a well trained model on our GAICD can obtain much better performance than the baseline. These results further prove the advantages of our new database as well as the associated metrics compared to the previous ones.

5. More qualitative comparison

This section shows more qualitative comparison of different methods on four typical scenes: single object, multi-objects, building and landscape. Following the main paper, we first compare our method with VFN [2], A2-RL [4] and VEN [6] under the setting of returning top-1 crop using the default candidate crops of each method. The results are shown in Fig. 4. We then compare our method with VFN and VEN under the setting of return crops having fixed aspect ratios: 16:9, 4:3 and 1:1. The results on four typical scenes are shown in Fig. 5, Fig. 6, Fig. 7 and Fig. 8, respectively.

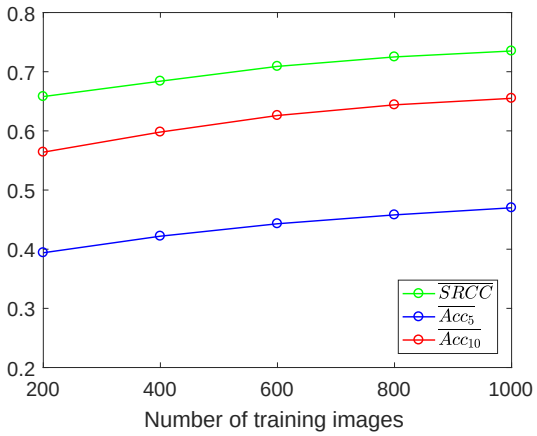

Figure 3: Performance by using different numbers of training images. The $\overline{Acc_5}$ and $\overline{Acc_{10}}$ are scaled into the range [0,1] for display.

Table 2: Performance comparison on previous databases using IoU as the metric. For Baseline_N, we simply calculate the IoU between the groundtruth and source image without cropping. For Baseline_C, we crop the central part whose width and height are 0.9 time of the source image.

Method	ICDB[7]			FCDB[1]
	Set 1	Set 2	Set 3	
Yan *et al*. [7]	0.7487	0.7288	0.7322	–
Chen *et al*. [1]	0.6683	0.6618	0.6483	0.6020
VFN [2]	0.7640	0.7529	0.7333	0.6802
Wang *et al*. [5]	0.8130	0.8060	0.8160	–
Li *et al*. [4]	0.8019	0.7961	0.7902	0.6633
Guo *et al*. [3]	**0.8500**	**0.8370**	**0.8280**	–
VEN [6]	–	–	–	**0.7349**
Baseline_N	0.8237	0.8299	0.8079	0.6379
Baseline_C	0.7843	0.7599	0.7636	0.6647
Ours ($\lambda = 0.5$)	0.7329	0.7123	0.7188	0.6645
Ours ($\lambda = 0.6$)	0.7491	0.7286	0.7340	0.6681
Ours ($\lambda = 0.7$)	0.7703	0.7507	0.7528	**0.6734**
Ours ($\lambda = 0.8$)	0.7988	0.7812	0.7788	0.6680
Ours ($\lambda = 0.9$)	**0.8237**	**0.8299**	**0.8079**	0.6379

References

[1] Y.-L. Chen, T.-W. Huang, K.-H. Chang, Y.-C. Tsai, H.-T. Chen, and B.-Y. Chen. Quantitative analysis of automatic image cropping algorithms: A dataset and comparative study. In *WACV*, pages 226–234, 2017. 1, 3

[2] Y.-L. Chen, J. Klopp, M. Sun, S.-Y. Chien, and K.-L. Ma. Learning to compose with professional photographs on the web. In *ACM Multimedia*, pages 37–45, 2017. 2, 3

[3] G. Guo, H. Wang, C. Shen, Y. Yan, and H.-Y. M. Liao. Automatic image cropping for visual aesthetic enhancement using deep neural networks and cascaded regression. *arXiv preprint arXiv:1712.09048*, 2017. 3

[4] D. Li, H. Wu, J. Zhang, and K. Huang. A2-RL: Aesthetics aware reinforcement learning for image cropping. In *CVPR*, pages 8193–8201, 2018. 2, 3

[5] W. Wang and J. Shen. Deep cropping via attention box prediction and aesthetics assessment. In *ICCV*, 2017. 3

[6] Z. Wei, J. Zhang, X. Shen, Z. Lin, R. Mech, M. Hoai, and D. Samaras. Good view hunting: Learning photo composition from dense view pairs. In *CVPR*, pages 5437–5446, 2018. 2, 3

[7] J. Yan, S. Lin, S. Bing Kang, and X. Tang. Learning the change for automatic image cropping. In *CVPR*, pages 971–978, 2013. 1, 3

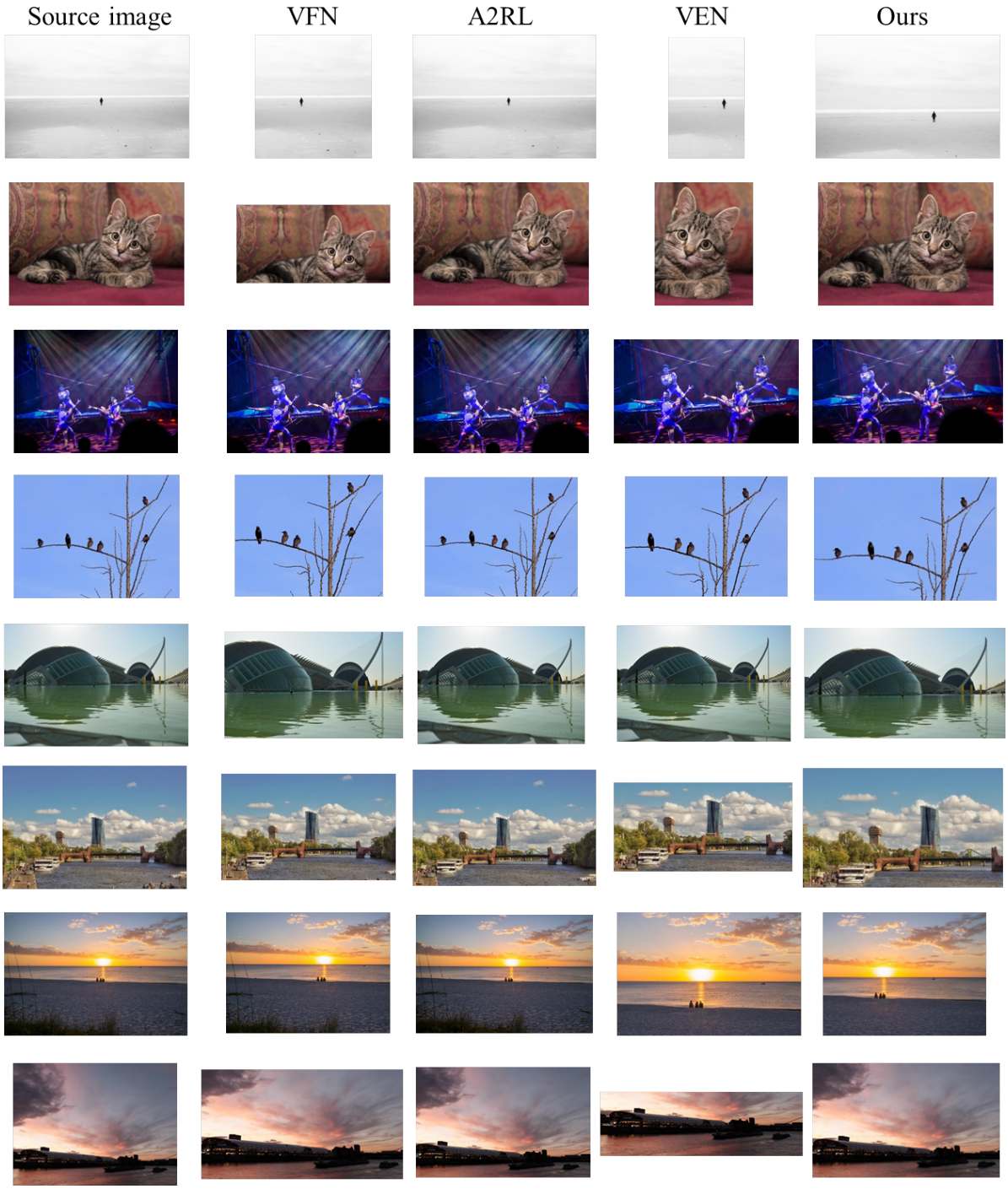

Figure 4: Qualitative comparison of returned top-1 crop by different methods using their default candidate crops except for VFN, which does not provide source code to generate candidate crops thus uses the same candidates as our method.

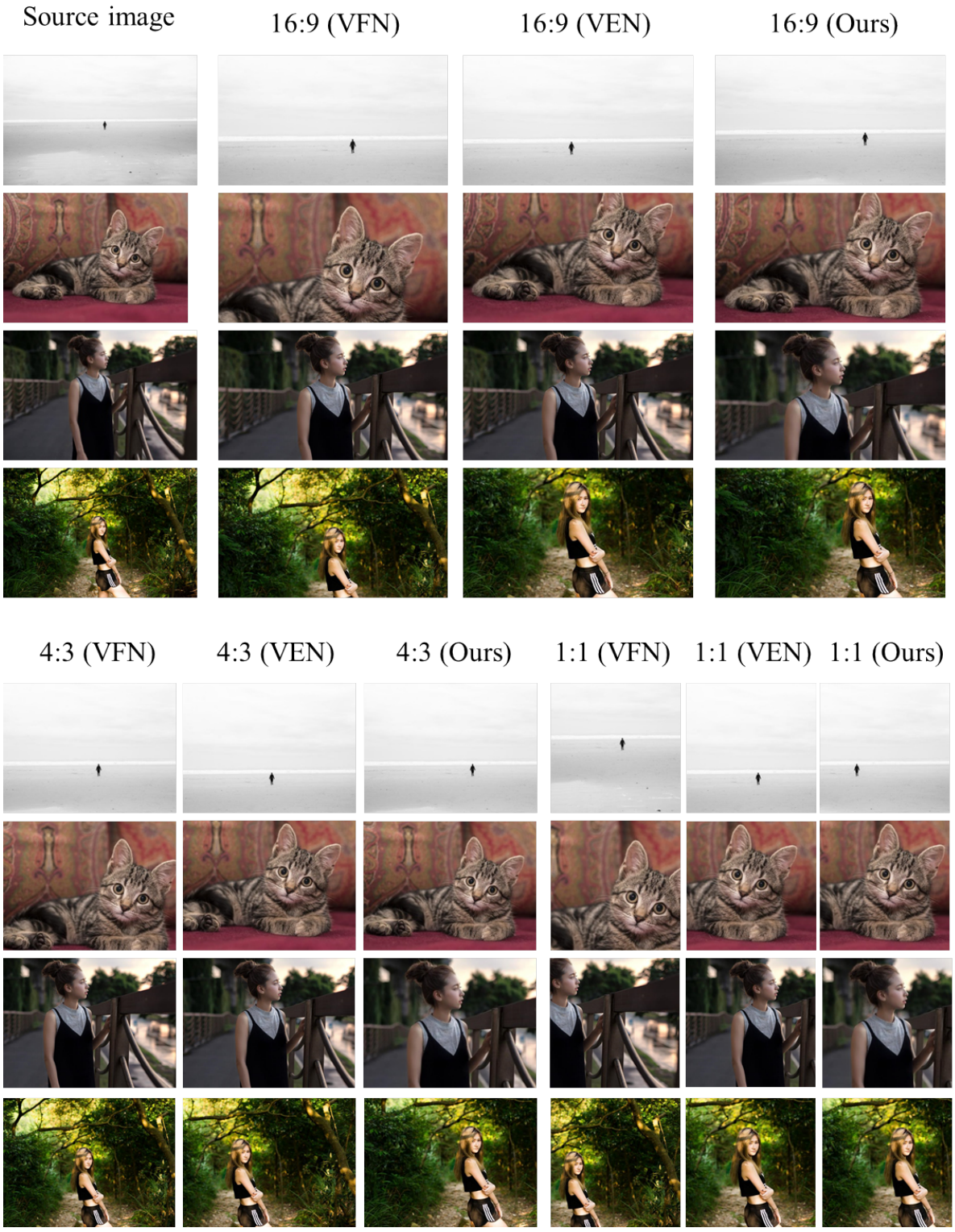

Figure 5: Qualitative comparison of different methods on single-object images under the setting of returning crops having fixed aspect ratios: 16:9, 4:3 and 1:1.

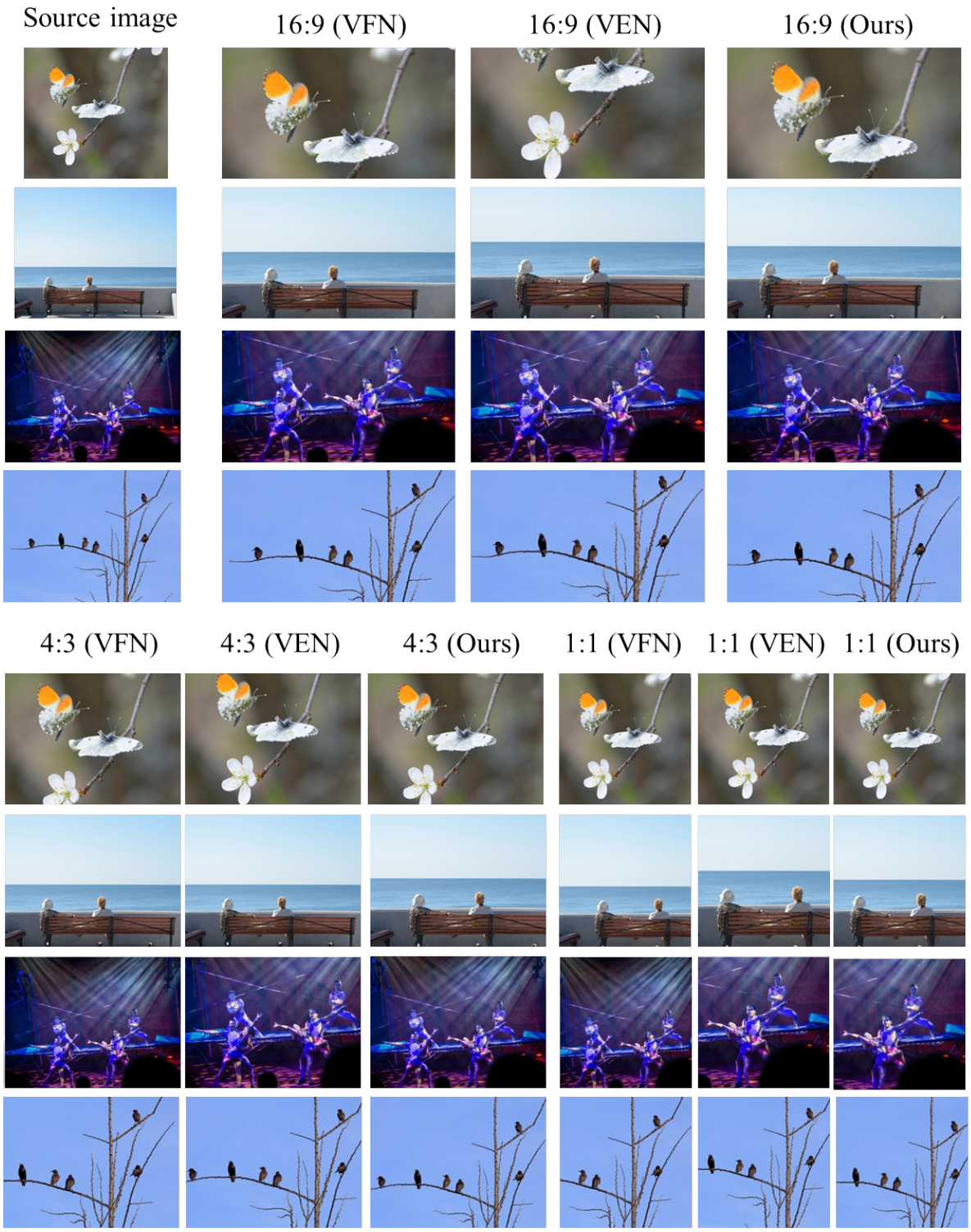

Figure 6: Qualitative comparison of different methods on multi-object images under the setting of returning crops having fixed aspect ratios: 16:9, 4:3 and 1:1.

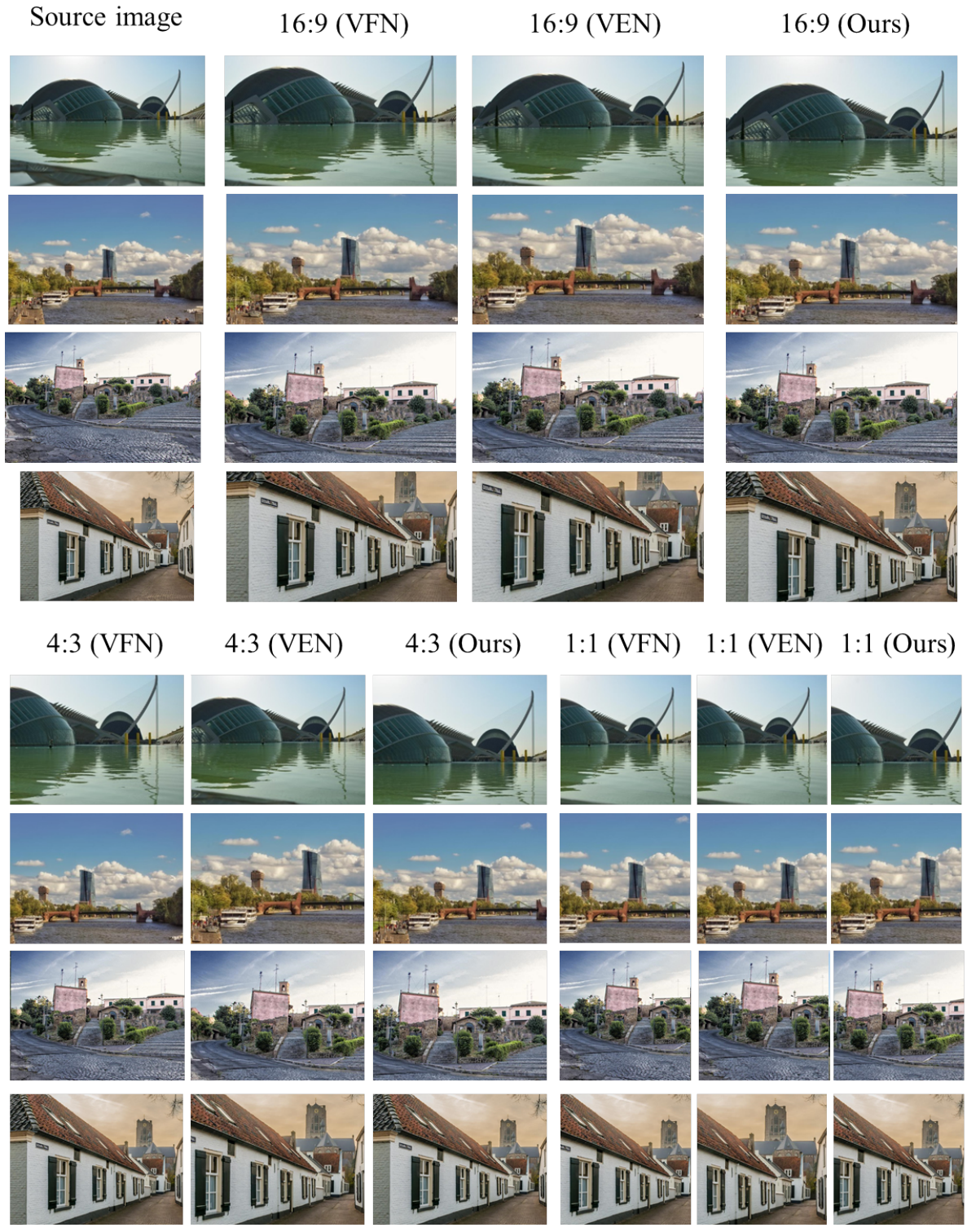

Figure 7: Qualitative comparison of different methods on building images under the setting of returning crops having fixed aspect ratios: 16:9, 4:3 and 1:1.

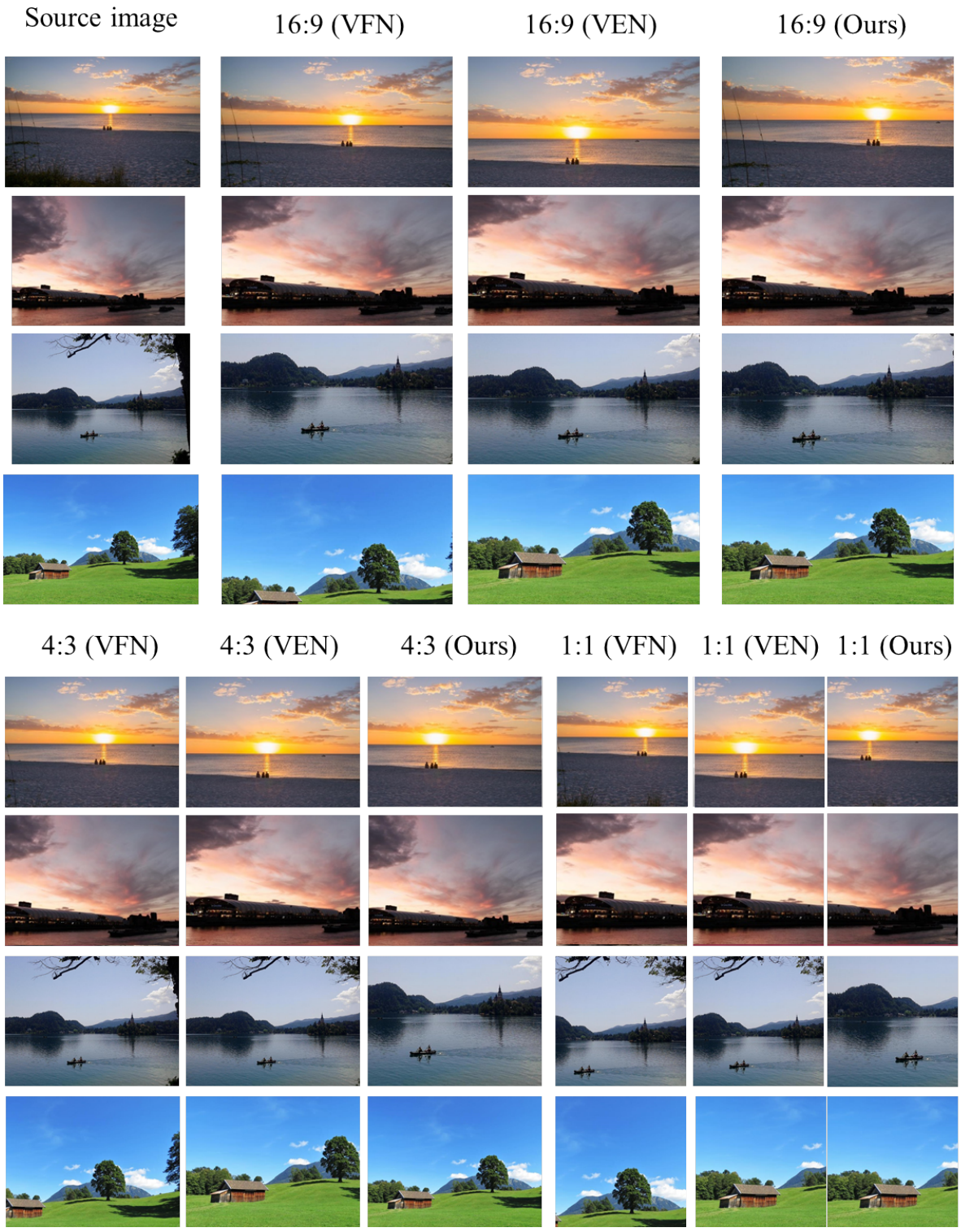

Figure 8: Qualitative comparison of different methods on landscape images under the setting of returning crops having fixed aspect ratios: 16:9, 4:3 and 1:1.