Composing Good Shots by Exploiting Mutual Relations
Supplementary Material

Debang Li [1,2], Junge Zhang[1,2], Kaiqi Huang[1,2,3], Ming-Hsuan Yang[4,5]
[1] CRISE, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
[3] CAS Center for Excellence in Brain Science and Intelligence Technology, Beijing, China
[4] University of California, Merced [5] Google Research
{debang.li,jgzhang, kaiqi.huang}@nlpr.ia.ac.cn, mhyang@ucmerced.edu

1. More Visual Examples of the Graph Module

In the paper, we analyze how the graph-based reasoning helps find good results by showing a visual example, which demonstrates that the graph module can make the distance between the features of good views and bad views get much more substantial so that the model can find good views more easily. In this section, we show more such visual examples that reflect how the graph module performs in Figure 1. In Figure 1(b), the edge weights between different nodes are highly correlated with the similarities between the annotated scores. The closer the order (*Top-K*) is, the bigger the weight, demonstrating that the graph (adjacency matrix) can reflect the relations of the image composition quality between different regions. By comparing Figure 1(c) and 1(d), we can find that the distances between the features of the good views and bad views get much larger after the graph-based reasoning. The reason for this is already discussed in the paper, that the edge weights between the good views and bad views are very small, at the same time, the good views are connected with large weights, and the bad views are also connected with substantial weights. The feature propagation on the graph makes the distance between the aggregated features of good views and bad views much more substantial.

2. More Qualitative Results

To more intuitively compare the capabilities of different methods, we show more qualitative results in the supplementary material. We compare the qualitative results of the proposed method with results of other state-of-the-art methods, including the A2RL [3], VFN [1], VPN [5], VEN [5], and GAIC [7] methods. We show the qualitative comparison results on the GAICD [7], HCDB [2], and ICDB [6] datasets in Figures 2, 3, 4, 5, 6, and 7. These qualitative results show that the proposed method can remove the unpleasant outer areas of images under different conditions and find views with a better composition compared to most existing methods.

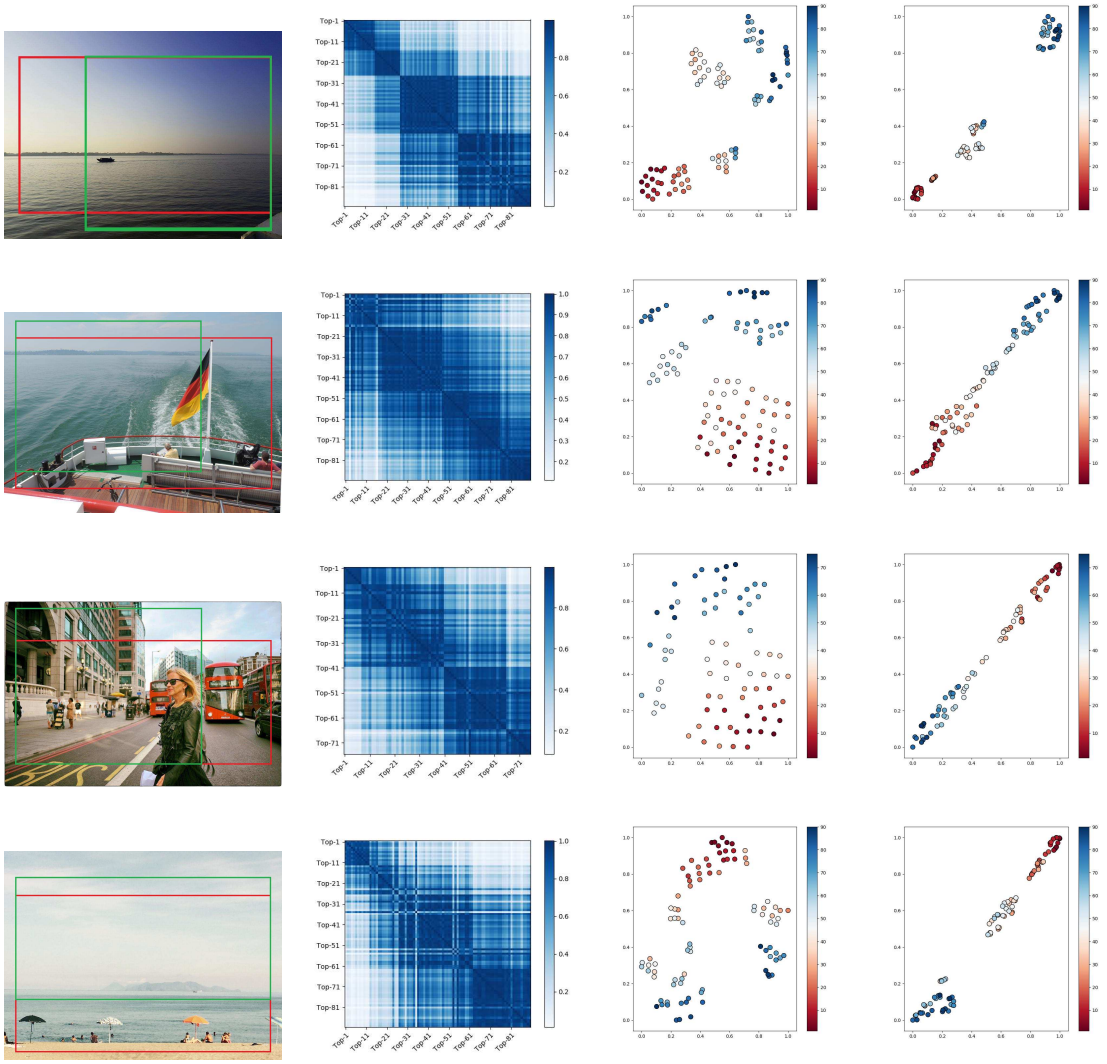
We also compare the results of different models under different aspect ratio requirements because a practical model should be able to handle varying needs. We generate the candidates with the required aspect ratio following [7] and use different models to return the best one. In this experiment, we choose three standard aspect ratios, *i.e.*, 1:1, 4:3, and 16:9, to evaluate different models. Because the A2RL [3] model directly generates the results and the VPN [5] model employs fixed anchors, they do not apply to this experimental setting. So we only compare the proposed method with the GAIC [7], VEN [5], and VFN [1] models in this setting. The results in Figures 8 and 9 show that the proposed model can find the most pleasing views under different requirements in most cases.

References

[1] Yi-Ling Chen, Jan Klopp, Min Sun, Shao-Yi Chien, and Kwan-Liu Ma. Learning to compose with professional photographs on the web. In *ACM Multimedia*, 2017.

[2] Chen Fang, Zhe Lin, Radomir Mech, and Xiaohui Shen. Automatic image cropping using visual composition, boundary simplicity and content preservation models. In *ACM Multimedia*, 2014.

[3] Debang Li, Huikai Wu, Junge Zhang, and Kaiqi Huang. A2-rl: Aesthetics aware reinforcement learning for image cropping. In *CVPR*, 2018.



(a) (b) (c) (d)

Figure 1. **More visual examples of how the graph-based reasoning performs.** (a) The source image, (b) the adjacency matrix of the constructed graph for the source image, (c) the t-SNE [4] visualization of the feature distribution for different candidates before the graph-based reasoning, (d) the feature distribution after the graph-based reasoning. In (b), the *Top-K* indicates that the region has the K-th highest annotated score among all regions. In (c) and (d), the number K in the color bar also indicates the region with the K-th highest annotated score among all regions, so red nodes represent regions with high annotated scores, and blue nodes represent regions with low annotated scores. Zoom in for the best view.

[4] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *JMLR*, 2008.

[5] Zijun Wei, Jianming Zhang, Xiaohui Shen, Zhe Lin, Radomír Mech, Minh Hoai, and Dimitris Samaras. Good view hunting: Learning photo composition from dense view pairs. In *CVPR*, 2018.

[6] Jianzhou Yan, Stephen Lin, Sing Bing Kang, and Xiaoou Tang. Learning the change for automatic image cropping. In *CVPR*, 2013.

[7] Hui Zeng, Lida Li, Zisheng Cao, and Lei Zhang. Reliable and efficient image cropping: A grid anchor based approach. In *CVPR*, 2019.

Figure 2. **Qualitative comparisons on the GAICD [7] dataset.** We compare the proposed method with the state-of-the-arts methods, including the A2RL [3], VFN [1], VPN [5], VEN [5], and GAIC [7] methods.

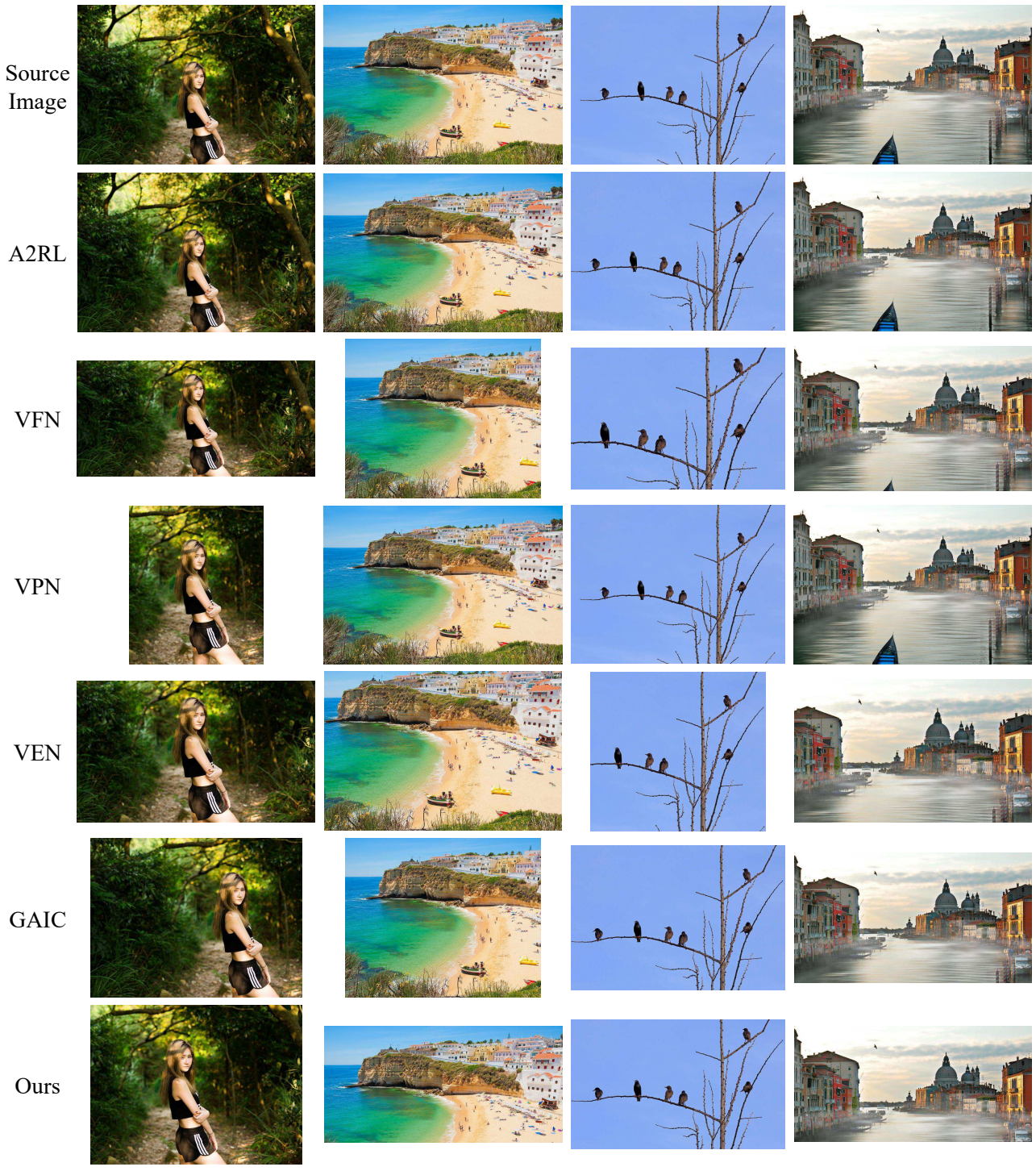


Figure 3. **Qualitative comparisons on the GAICD [7] dataset.** We compare the proposed method with the state-of-the-arts methods, including the A2RL [3], VFN [1], VPN [5], VEN [5], and GAIC [7] methods.

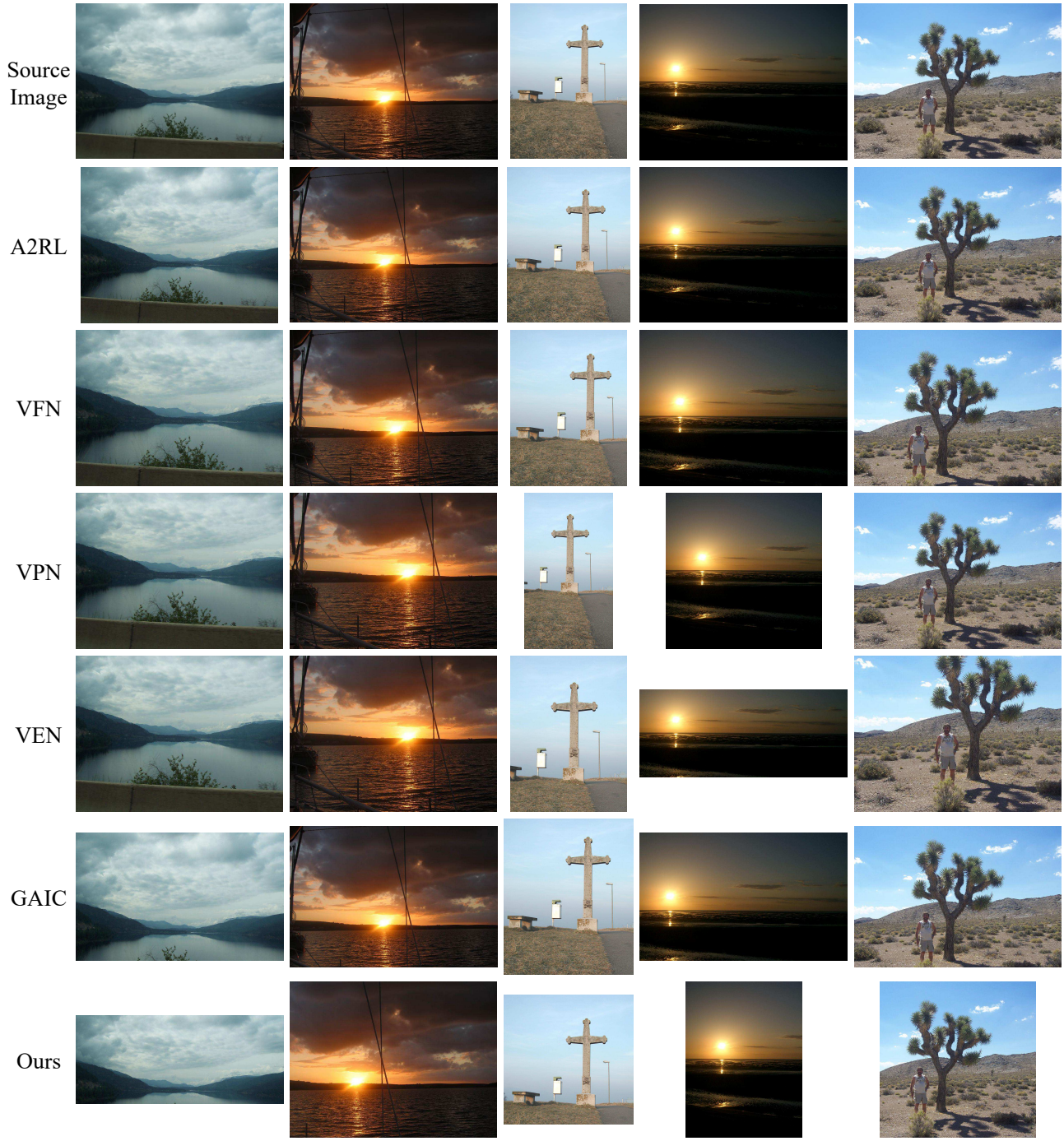


Figure 4. **Qualitative comparisons on the HCDB [2] dataset.** We compare the proposed method with the state-of-the-arts methods, including the A2RL [3], VFN [1], VPN [5], VEN [5], and GAIC [7] methods.

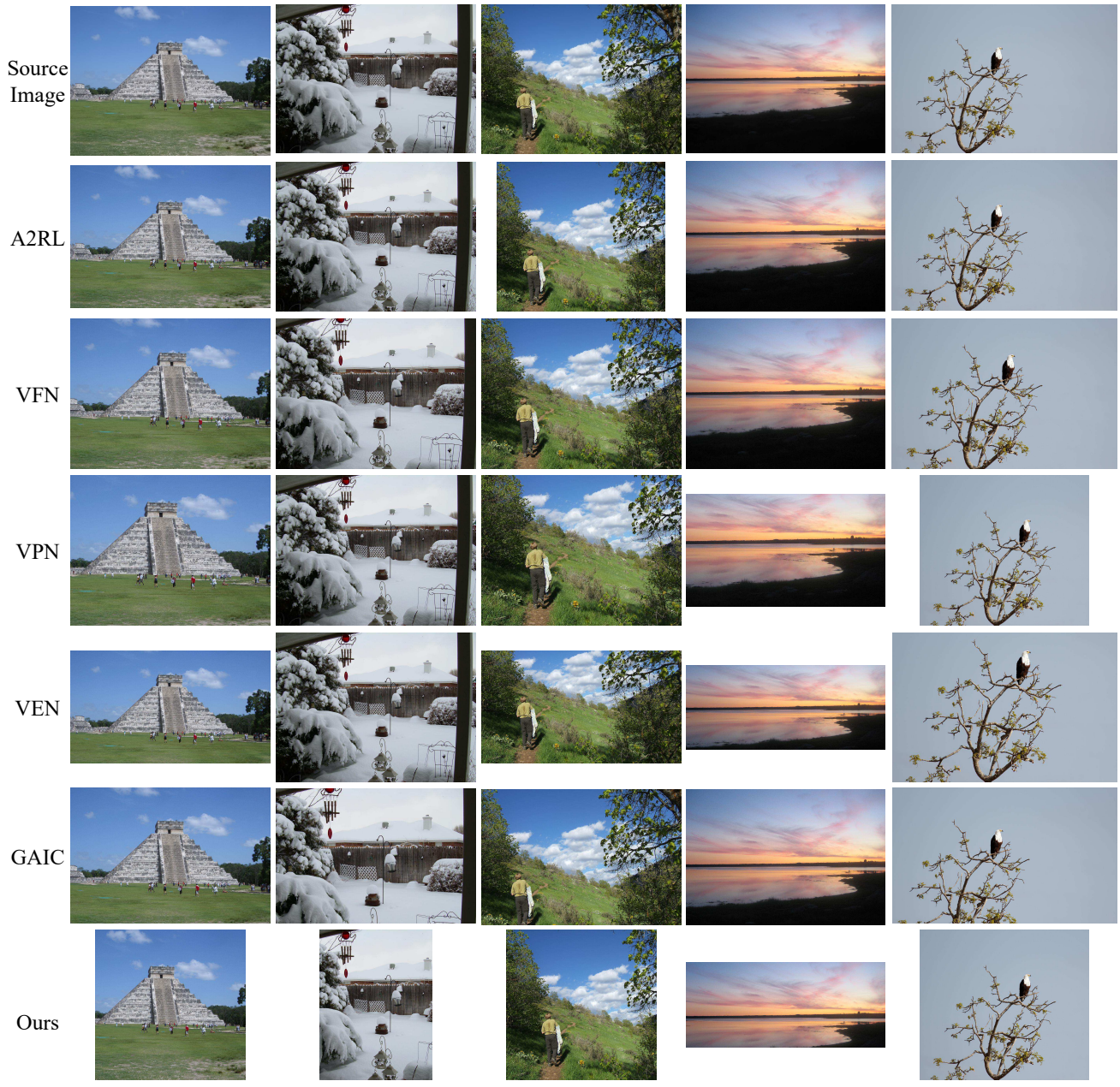


Figure 5. **Qualitative comparisons on the HCDB [2] dataset.** We compare the proposed method with the state-of-the-arts methods, including the A2RL [3], VFN [1], VPN [5], VEN [5], and GAIC [7] methods.

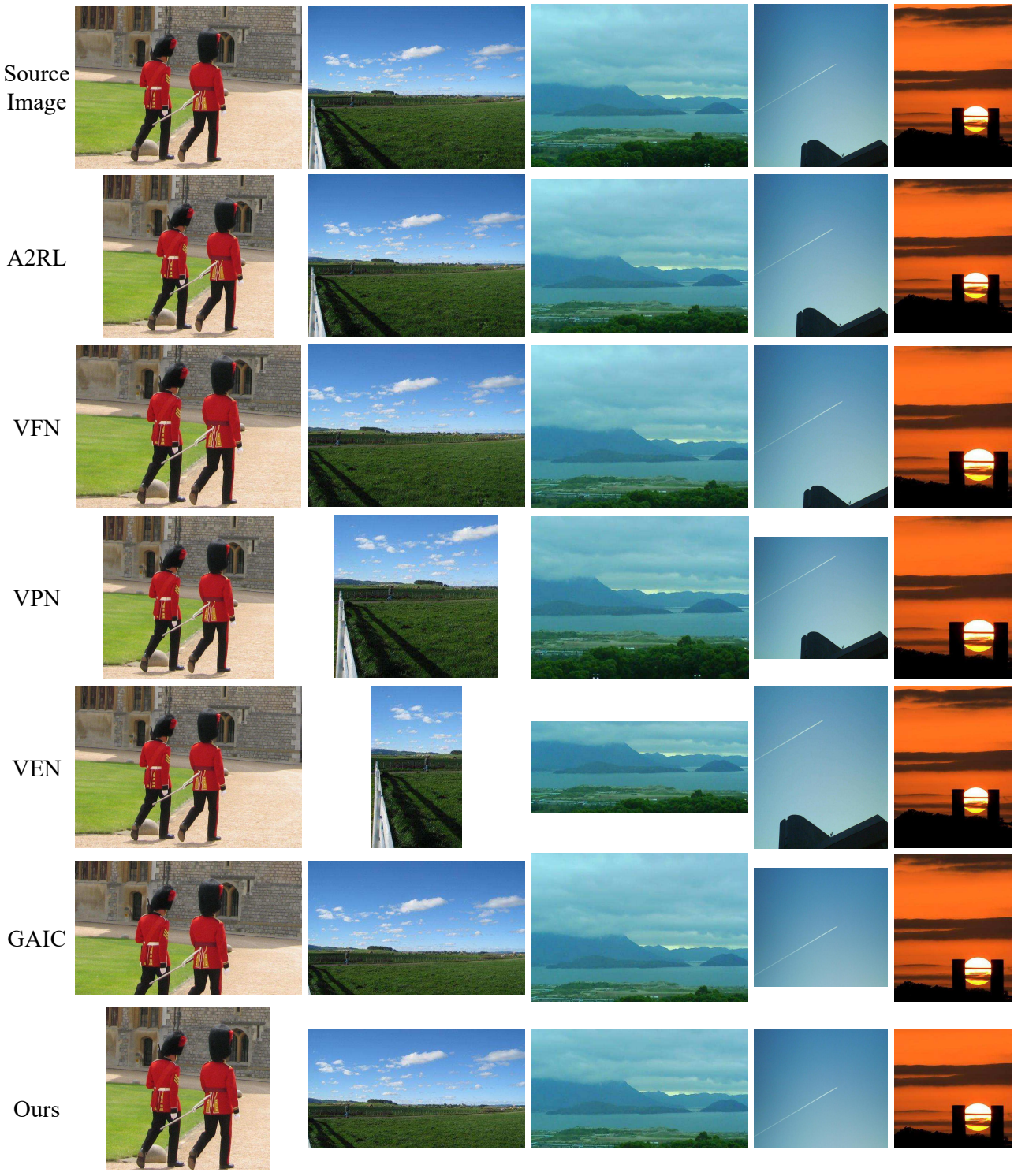


Figure 6. **Qualitative comparisons on the ICDB [6] dataset.** We compare the proposed method with the state-of-the-arts methods, including the A2RL [3], VFN [1], VPN [5], VEN [5], and GAIC [7] methods.

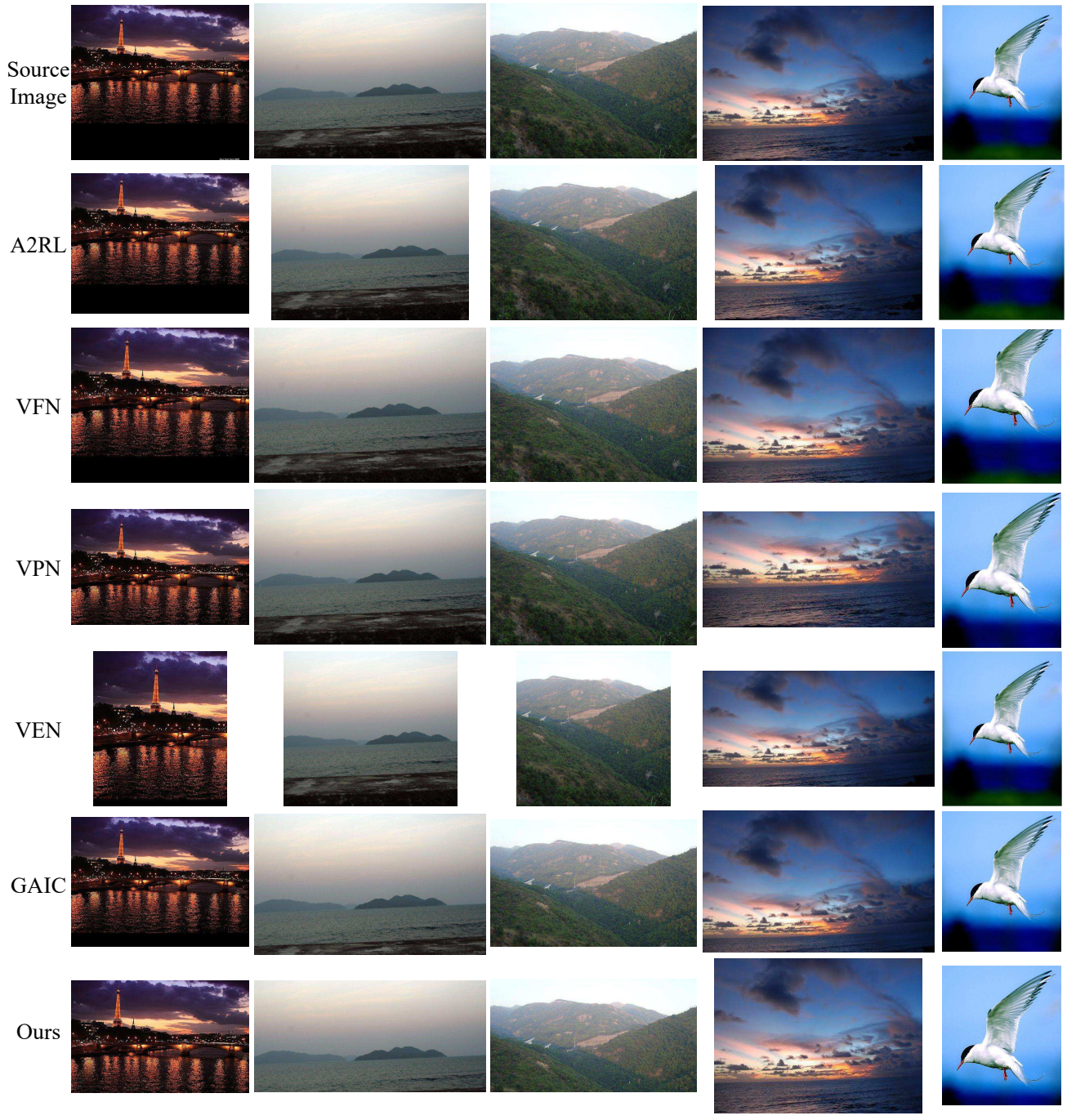


Figure 7. **Qualitative comparisons on the ICDB [6] dataset.** We compare the proposed method with the state-of-the-arts methods, including the A2RL [3], VFN [1], VPN [5], VEN [5], and GAIC [7] methods.

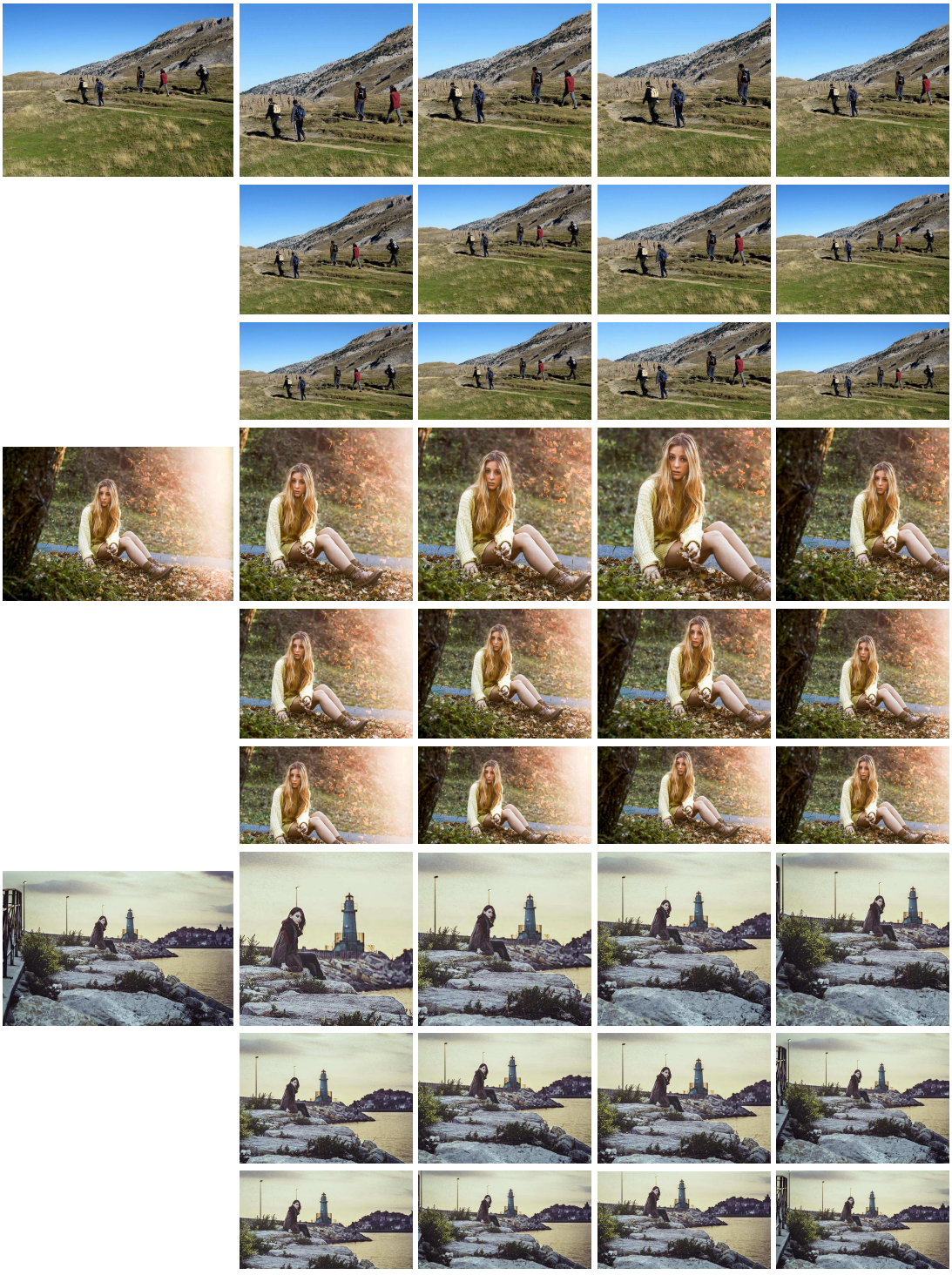


Figure 8. **Qualitative comparisons of different methods on returning crops having fixed aspect ratios.** From each group of images, the aspect ratio is 1:1, 4:3, and 16:9 for the first, second, and third row, respectively.

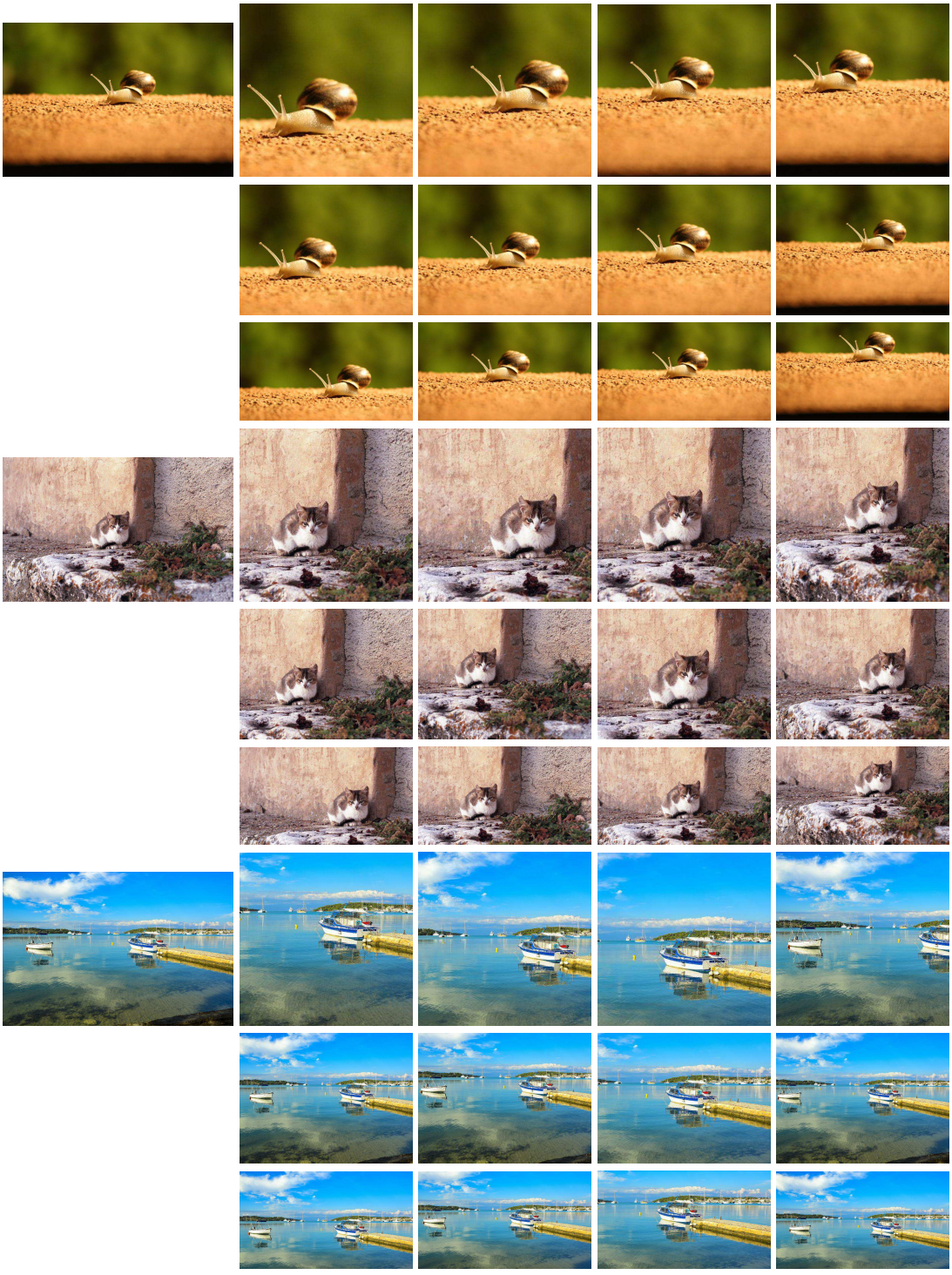


Figure 9. **Qualitative comparisons of different methods on returning crops having fixed aspect ratios.** From each group of images, the aspect ratio is 1:1, 4:3, and 16:9 for the first, second, and third row, respectively.