Supplementary

1 Performance on image retargeting

We exhibit more results on image retargeting in Fig. 1 to further demonstrate the capability of our method. Alike video content, our approach also performs well on different image contents, *i.e.*, portrait image - Fig. 1(a), shaped object - Fig. 1(b), multiple objects - Fig. 1(c), line structure - Fig. 2(d). Images with reflection symmetry are challenging when retargeted by seam carving operator [2]. Fig. 1(d) consists of line structure and symmetrical structure. Our method still works well and produces appealing results without distortion.

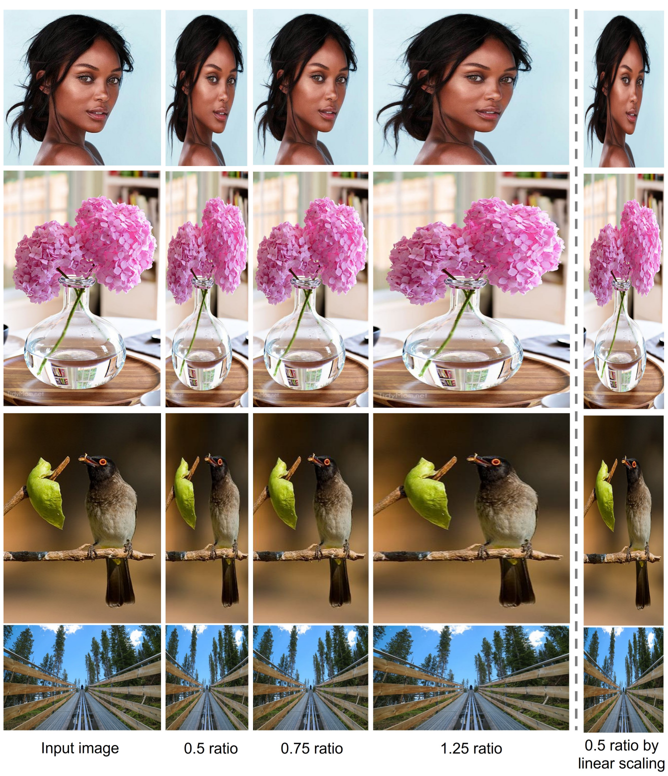

Fig. 1: Demonstrate our performance on retargeting images with diverse image content.

Apart from the comparisons with video retargeting methodologies in the manuscript, we further discuss the ability of our method via visual comparison with four recent state-of-the-art image retargeting systems, WSSDCNN [3], SAMIR [7], grid encoding model [4] and Cycle-IR [6]. The visualization is presented in Fig. 2. In each comparison, we include the results generated from MultiOP [5] and linear scale to facilitate readers inferring the quality of the results. The experimental results in a single method demonstrate their effectiveness with their own advance on a certain data. However, all these methods have their downfalls. WSSDCNN [3] is the first work approaching image retargeting by deep learning technique. In the result presented in Fig.2(a), it can be seen that noticeable distortions occur in their results (the table and the chairs). Meanwhile there is no distortion in our result. With the second competitor SAMIR [7], Fig.2(b), our method outperforms while their result suffers from cropping significantly. In Fig.2(c), the image "*Housefense*" is challenging since the important contents are mostly distributed throughout the entire image. As observed, Cycle-IR [6] fails to preserve the line structure encompassed in this image, *i.e.*, the road or the fence. This is also the limitation mentioned in their work, *i.e.*, they fail to work on images with different complex scenes. The final competitor is a novel grid encoding model for content-aware image retargeting [4]. Generally, the bird's shape in Kim's result is preserved pretty well and slightly better than ours. However, the wire is distorted heavily in their result. Our result is probably not perfect in this competition but there is no artifact or distortion appears in this case.

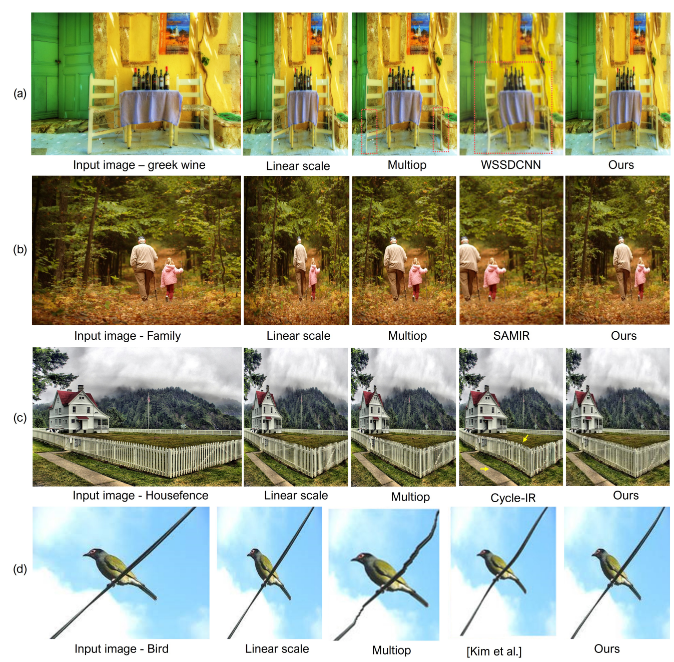

Fig. 2: Comparisons with WSSDCNN [3] (a), SAMIR [7] (b), Cycle-IR [6] (c), and a grid-warping by Kim et al. [4] (d).

2 Discussion on cropping effect

As we discussed in the manuscript, when we construct frames R, we set a window $[(\frac{S_w}{2} - \frac{T_w}{2}) : (\frac{S_w}{2} + \frac{T_w}{2})]$, as we visualize in Fig. 3, in which the black is the source size, the red is the estimated size, and the blue is the target size. Therefore, when constructing the final resized frame by this equation, the pixels, which are out of the blue window, will be cropped. It's worth noting that although cropping may occur in almost all cases, this is just a light cropping. In the cases that the main objects are at the leftmost or rightmost of frames, the resized results are not appealing such as the example in Fig. 4.

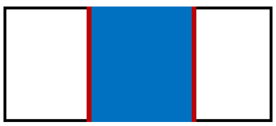

Fig. 3: Visualization the reason of cropping effect.

Fig. 4: Sample of resizing result with main object not in the middle of frame.

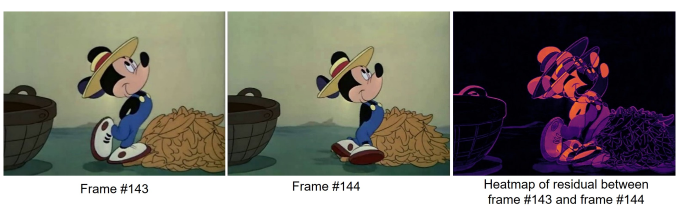

Fig. 5: Heatmap of residual image.

3 More visualizations

In Fig. 4 of the main manuscript, we exhibit our results with the same configuration of equation (6) to demonstrate our performance in different resizing ratios. The results we put in Fig. 4 are scaled down to make them fit the page width, it is therefore not clear for reader's observation. We showcase the Fig. 6 for the comparisons between our results and those by linear scaling in 0.5 and 0.3 ratios. Zooming in the regions we highlight in yellow, we can see that our results are less deformed than those by linear scaling. Particularly the body (yellow arrows) and the legs (yellow circles) of main objects. Inferring these regions in the results of 0.3 resizing ratio, we can see the same effect.

Should we consider more cropping? - Actually, we do not need to integrate with cropping. By adjusting value of ϑ, as we present in equation (16), we can generate resizing results according to our perceptual expectations. Besides, our method by itself has the cropping effect. Adjusting ϑ helps us to control the content in the window. Since the ratios 0.5 and 0.3 are the common use cases for mobile phones, to apply it on mobile phones, we will run equation (16) to make the content in resized videos more appealing. A visualization can be seen in the below figure. Obviously, our results are much better than linear scale and Adobe Express [1] in 0.3 ratio.

Fig. 5 shows the heatmap of residual between two frames discussed in equation (18) of the manuscript. In this equation, given two frames $\mathbf{F}_t$ and $\mathbf{F}_{t-1}$ of height H and width W, the numerator returns the residual of the two frames, as we visualize the heatmap of residual in the figure below. Thereafter, we convert it to an array and average on the total of pixels by $H \times W$. We divide it by 100 to normalize it to range of [0, 1].

Besides the results in the manuscript, Fig. 8 exhibits more video retargeting results. Visualization on video of these results could be seen on our project website[1].

References

[1] Adobe express, 2022. URL https://www.adobe.com/express/feature/video/resize/instagram. Accessed: March 24, 2023.

[2] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. In *ACM SIGGRAPH 2007 papers*, pages 10–es. 2007.

[3] D. Cho, J. Park, T.-H. Oh, Y.-W. Tai, and I. So Kweon. Weakly-and self-supervised learning for content-aware deep image retargeting. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4558–4567, 2017.

[4] Y. Kim, H. Eun, C. Jung, and C. Kim. A quad edge-based grid encoding model for content-aware image retargeting. *IEEE Transactions on Visualization and Computer Graphics*, 25(12):3202–3215, 2018.

[5] M. Rubinstein, A. Shamir, and S. Avidan. Multi-operator media retargeting. *ACM Transactions on Graphics (TOG)*, 28(3):1–11, 2009.

[6] W. Tan, B. Yan, C. Lin, and X. Niu. Cycle-ir: Deep cyclic image retargeting. *IEEE Transactions on Multimedia*, 22 (7):1730–1743, 2019.

[7] Y. Zhou, Z. Chen, and W. Li. Weakly supervised reinforced multi-operator image retargeting. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(1): 126–139, 2020.

1. http://graphics.csie.ncku.edu.tw/RETVI/

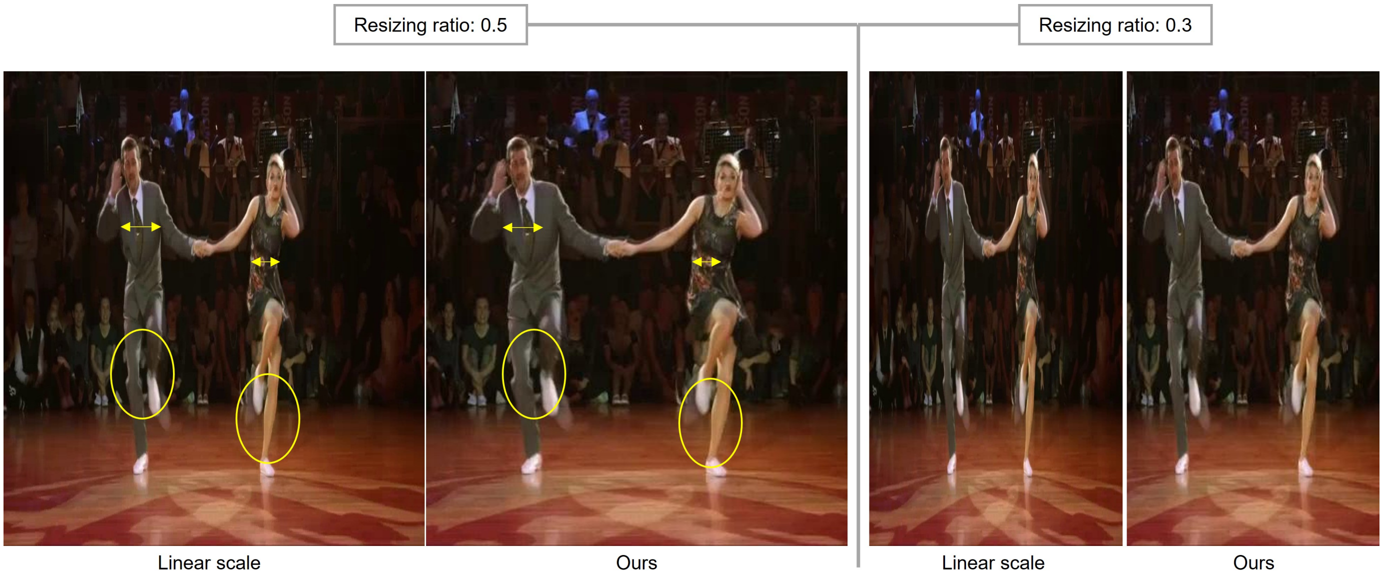

Fig. 6: Zoom in comparison between our our RETVI versus linear scale on 0.5 and 0.3 ratios.

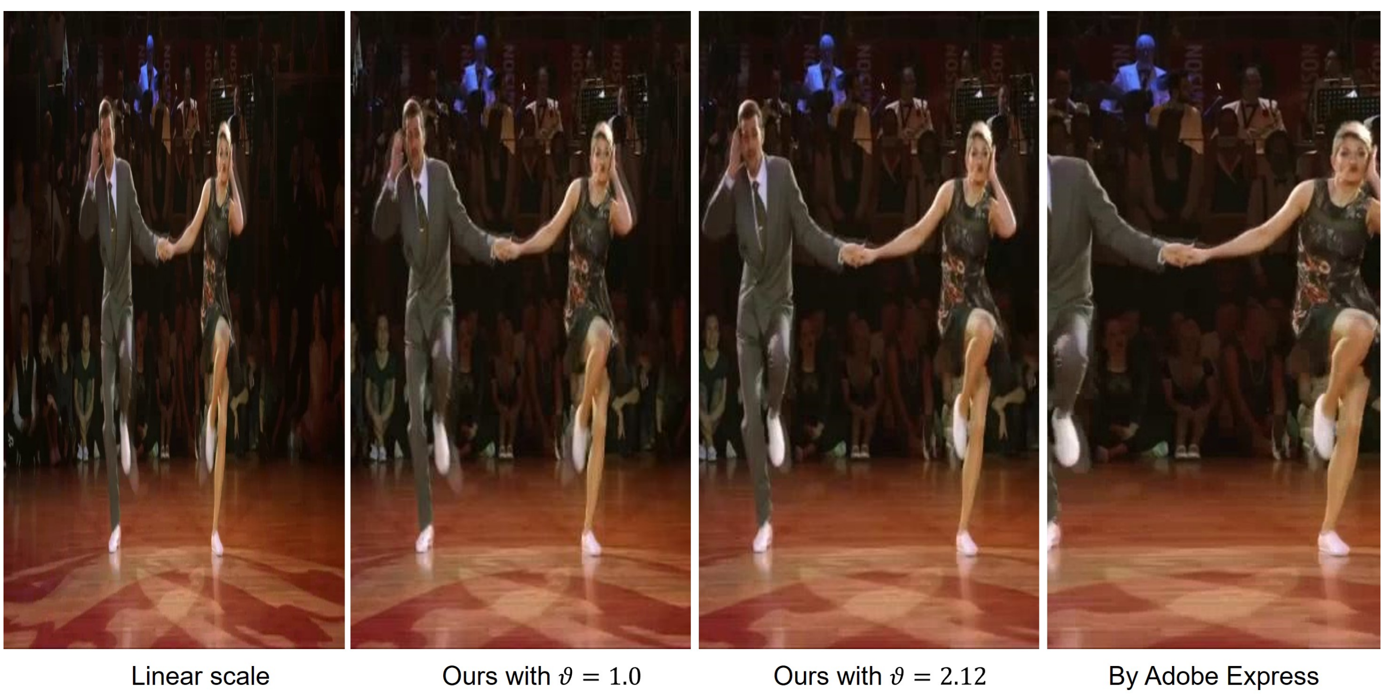

Fig. 7: Our result competes with linear scale and Adobe Express on 0.3 ratio.

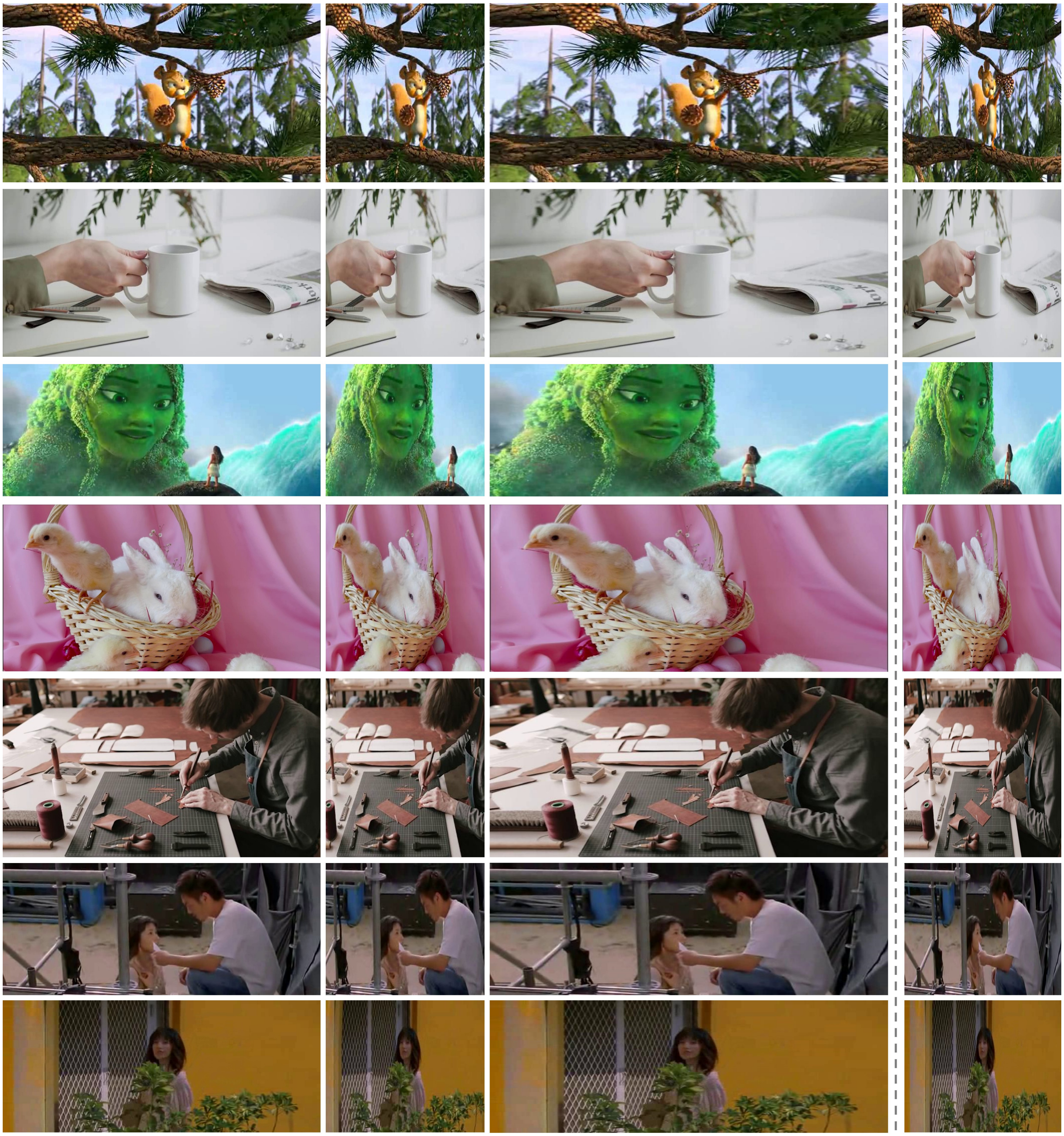

Fig. 8: Left to right: input video frame, resized frame of 0.5 ratio, resized frame of 1.25 ratio. The right most column is the 0.5 ratio results by linear scaling.