Supplementary Materials to "Beyond a Video Frame Interpolator: A Space Decoupled Learning Approach to Continuous Image Transition"

Anonymous ECCV submission

Paper ID 0000

In this supplementary file, we provide the following materials:

- Detailed architecture of the synthesis network (referring to Section 3.2 in the main paper);
- More visual comparisons of different methods on video frame interpolation (referring to Section 4.2 in the main paper);
- More visual comparisons of different methods on face aging (referring to Section 4.4 in the main paper);
- More visual comparisons of different methods on face toonification (referring to Section 4.4 in the main paper);
- More visual comparisons of different methods on image morphing (referring to Section 4.4 in the main paper);
- A demo video to demonstrate the continuous image conversion results of our method.

1 Synthesis Network

We employ a GridNet architecture [4] with three rows and six columns as the synthesis network to improve the final transition results. We incorporate parametric rectified linear units [5] to improve the training and use bilinear upsampling to avoid checkerboard artifacts [12]. We do not use Batch Normalization [6] by following the recent findings in image enhancement tasks [8]. The detailed architecture is presented in Fig. 1.

2 Experiments on Video Frame Interpolation

This section shows more visual results of competing methods on the video frame interpolation task. As in the main paper, we compare our SDL method with SepConv [11], SuperSloMo [7], CAIN [3], EDSC [2], DAIN [1], and BMBC [14]. The visual comparisons are presented in Fig. 2. One can see that our method generates superior results with fine details and edge structures.

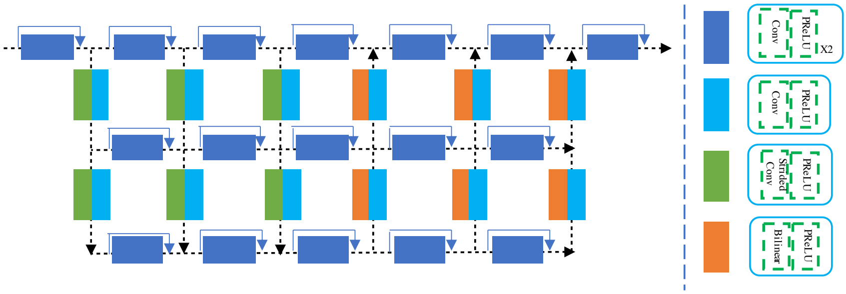


Fig. 1. The architecture of our synthesis network.

3 Experiments on Face Aging

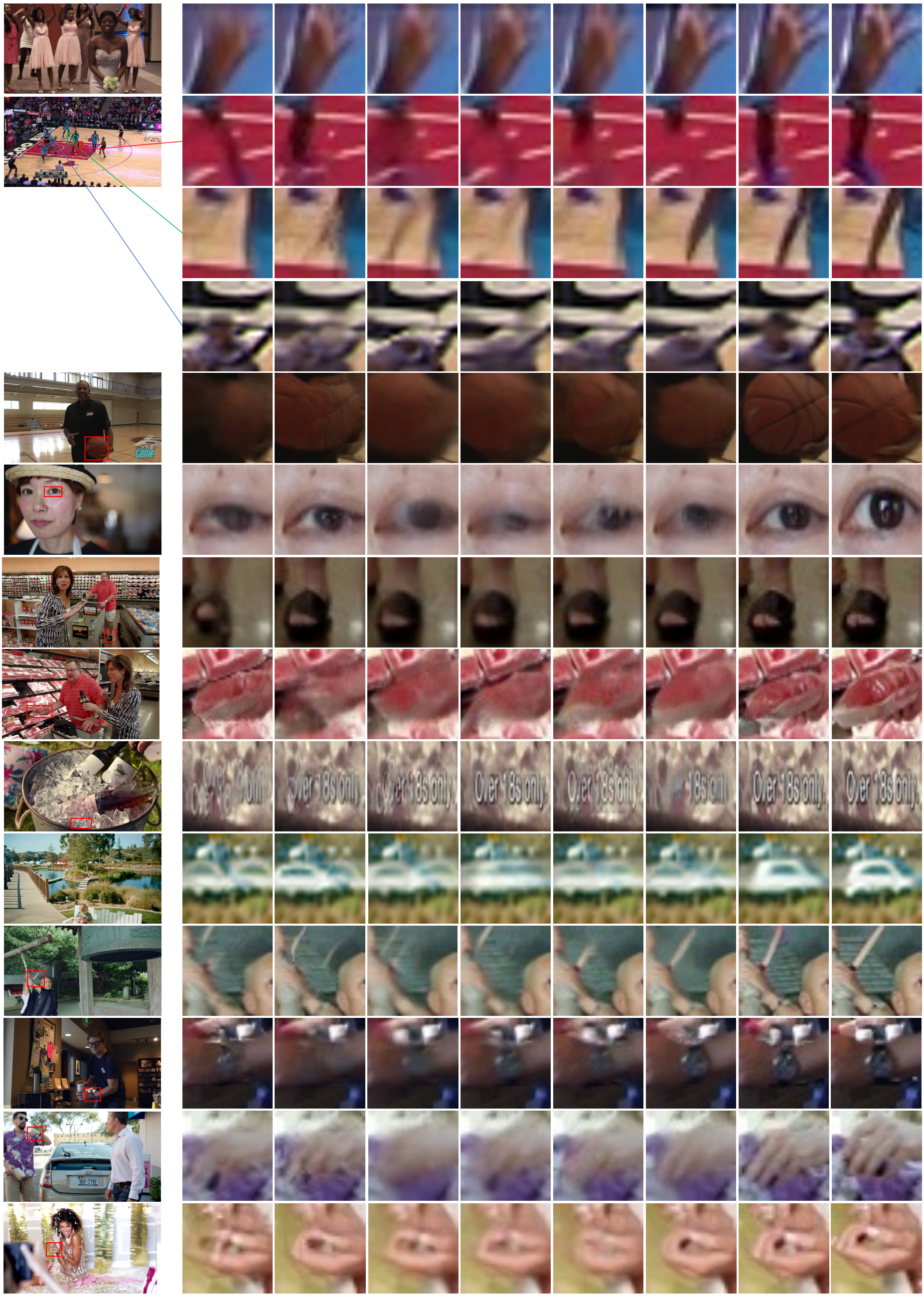
This section shows more visual results of competing methods on continuous face aging. As in Section 4.4 in the main manuscript, the methods StyleGAN2 backpropagation [17], SAVI2I [10], Lifespan [13] and DNI [18] are used in the comparison. Fig. 3 presents the visual comparisons on face aging. It can be seen that SDL achieves favorably better results than the competitors.

4 Experiments on Face Toonification

As in Section 4.4 of the main manuscript, We compare our SDL with Pinkney *et al.* [16] and SAVI2I [10] on face toonification. Fig. 4 shows the visual comparisons, demonstrating the superior performance of our method on continuous face toonification.

5 Experiments on Image Morphing

In this section, we compare SDL with StyleGAN2 backpropagation [17], CrossBreed [15], SAVI2I [10] and FUNIT [9] on image morphing. As presented in Fig. 5, SDL achieves smooth morphings with more vivid details and works much better on retaining image background.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 2. Visual comparison of competing methods on the Vimeo90K test set. (a) SepConv [11]; (b) SuperSloMo [7]; (c) CAIN [3]; (d) EDSC [2]; (e) DAIN [1]; (f) BMBC [14]; (g) SDL; (h) Ground truth.

Fig. 3. Comparison of SDL with competing methods on continuous face aging. For each example, from top to bottom are the results by SDL, StyleGAN2 backpropagation [17], SAVI2I [10], Lifespan [13] and DNI [18].

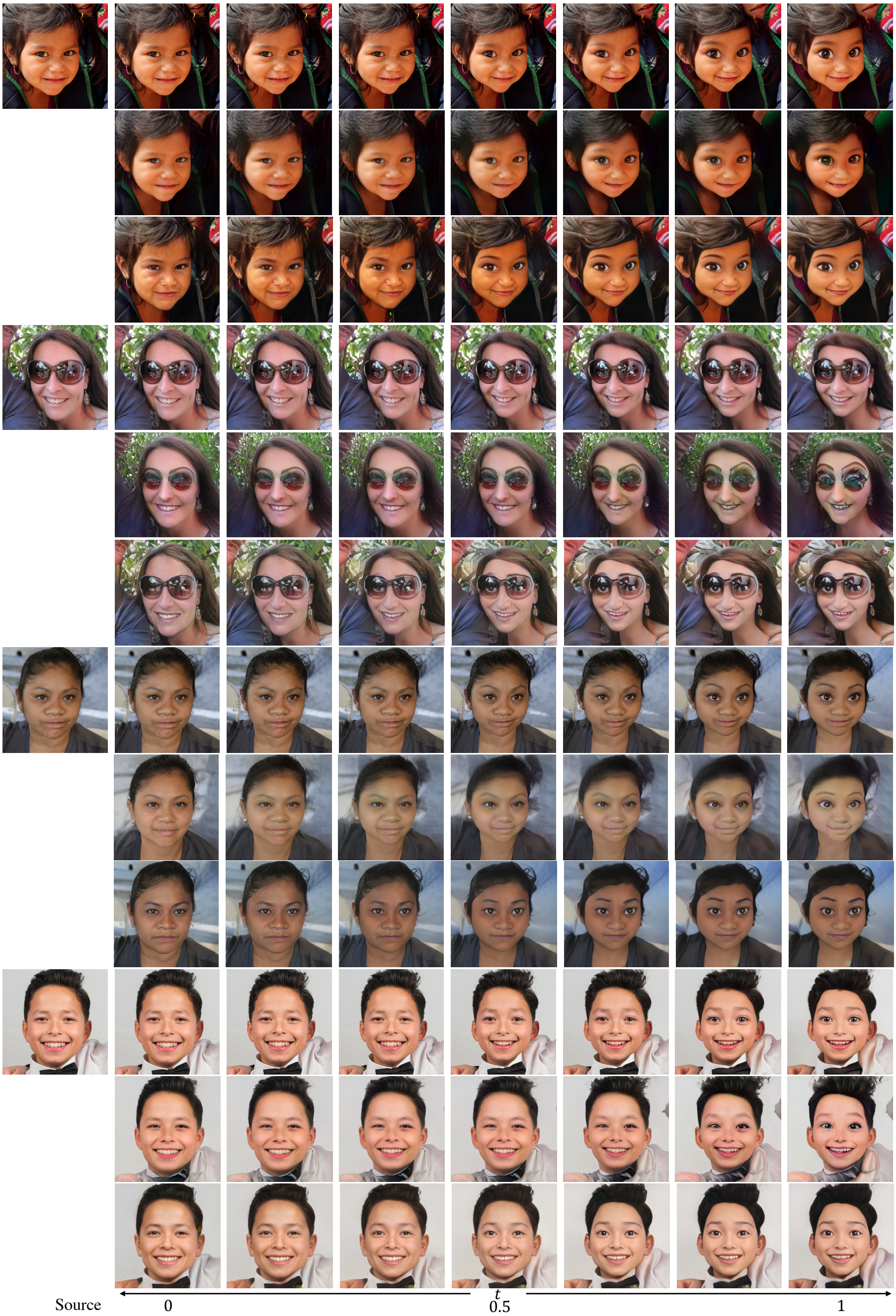


Fig. 4. Comparison of SDL with competing methods on continuous face toonification. For each example, from top to bottom are the results by SDL, Pinkney *et al.* [16] and SAVI2I [10].

Fig. 5. Comparison of SDL with competing methods on image morphing. For each example, from top to bottom are the results by SDL, StyleGAN2 backpropagation [17], CrossBreed [15], SAVI2I [10], and FUNIT [9].

References

1. Bao, W., Lai, W.S., Ma, C., Zhang, X., Gao, Z., Yang, M.H.: Depth-aware video frame interpolation. In: CVPR. pp. 3698–3707 (2019)
2. Cheng, X., Chen, Z.: Multiple video frame interpolation via enhanced deformable separable convolution. IEEE TPAMI (2021)
3. Choi, M., Kim, H., Han, B., Xu, N., Lee, K.M.: Channel attention is all you need for video frame interpolation. In: AAAI (2020)
4. Fourure, D., Emonet, R., Fromont, E., Muselet, D., Trémeau, A., Wolf, C.: Residual conv-deconv grid network for semantic segmentation. In: Proceedings of the British Machine Vision Conference (2017)
5. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: ICCV. pp. 1026–1034 (2015)
6. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. ArXiv (2015)
7. Jiang, H., Sun, D., Jampani, V., Yang, M.H., Learned-Miller, E., Kautz, J.: Super slomo: High quality estimation of multiple intermediate frames for video interpolation. In: CVPR (2018)
8. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. CVPRW pp. 1132–1140 (2017)
9. Liu, M.Y., Huang, X., Mallya, A., Karras, T., Aila, T., Lehtinen, J., Kautz, J.: Few-shot unsueprvised image-to-image translation. Arxiv (2019)
10. Mao, Q., Lee, H.Y., Tseng, H.Y., Huang, J.B., Ma, S., Yang, M.H.: Continuous and diverse image-to-image translation via signed attribute vectors. ArXiv (2020)
11. Niklaus, S., Mai, L., Liu, F.: Video frame interpolation via adaptive separable convolution. In: ICCV (2017)
12. Odena, A., Dumoulin, V., Olah, C.: Deconvolution and checkerboard artifacts (2016), http://distill.pub/2016/deconv-checkerboard
13. Or-El, R., Sengupta, S., Fried, O., Shechtman, E., Kemelmacher-Shlizerman, I.: Lifespan age transformation synthesis. In: ECCV (2020)
14. Park, J., Ko, K., Lee, C., Kim, C.S.: Bmbc: Bilateral motion estimation with bilateral cost volume for video interpolation. In: ECCV (2020)
15. Park, S., Seo, K., Noh, J.: Neural crossbreed: neural based image metamorphosis. ACM TOG **39**(6), 1–15 (2020)
16. Pinkney, J.N.M., Adler, D.: Resolution dependent gan interpolation for controllable image synthesis between domains. ArXiv (2020)
17. Viazovetskyi, Y., Ivashkin, V., Kashin, E.: Stylegan2 distillation for feed-forward image manipulation. In: ECCV (2020)
18. Wang, X., Yu, K., Dong, C., Tang, X., Loy, C.C.: Deep network interpolation for continuous imagery effect transition. In: CVPR. pp. 1692–1701 (2019)