

# CoDeF: Content Deformation Fields for Temporally Consistent Video Processing

## Supplementary Materials

Anonymous CVPR submission

Paper ID 7925

Table 1. Reconstruction PSNR on DAVIS [2]

Video Name	LNA [1]	CoDeF
Blackswan	29.92	31.51
Boat	31.51	34.13
Car-turn	28.35	30.69
Kite-surf	28.37	34.26
Libby	28.29	30.05
Motorbike	29.85	32.99

001 The supplementary materials are structured as follows:  
 002 We first outline the detailed configurations of the training  
 003 process. Subsequently, we present additional quantitative  
 004 results pertaining to the reconstruction. Lastly, we supply  
 005 further qualitative results, expanding on various applica-  
 006 tions and a diverse array of video sequences.

### 007 1. Implementation Details

008 Our training networks comprise two MLPs equipped with  
 009 HashEncoding. For the hash settings, we configure the  
 010 number of features per level to 2 and set the level to 16.  
 011 The base resolution starts from 16 and the log hashmap size  
 012 is designated as 19. For the 3D deformable MLP, the scale  
 013 per level is set to 1.38, and for the 2D MLP, it is set to  
 014 1.44. Pertaining to the MLP configurations, we employ  
 015 eight layers for the 3D variant and two layers for the 2D  
 016 variant. The activation function used in the MLP is ReLU.  
 017 The initial learning rate is established at 1e-3 and is halved  
 018 every 2,500 iterations. Our experiments' default parameters  
 019 have the anneal begin and end steps set at 4,000 and 8,000,  
 020 respectively. The total iteration step is limited at 10,000.  
 021 The flow loss coefficient is set to 1, and the background  
 022 loss is set at 0.03.

### 023 2. Quantitative Comparison

024 We follow the evaluation settings in LNA [1] and report  
 025 more sequences in DAVIS dataset as shown in Table. 1

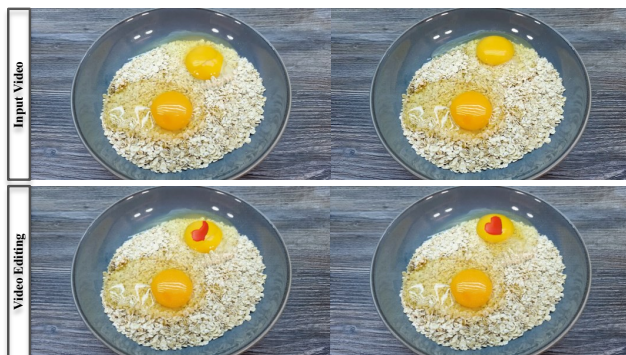


Figure 1. **User interactive video editing** achieved by editing *only one* image and propagating the outcomes along the time axis using our CoDeF. We strongly encourage the readers to see the supplementary videos to appreciate the temporal consistency.

### 026 3. More Results

027 In this section, we first show another application of CoDeF  
 028 which is *User interactive Video Editing*. Our representation  
 029 allows for user editing on objects with unique styles without  
 030 influencing other parts of the image. As exemplified  
 031 in Fig. 1, users can manually adjust content on the canonical  
 032 image to perform precise edits in areas where the automatic  
 033 editing algorithm may not be achieving optimal results.

034 We present further results generated using CoDeF for a  
 035 variety of video sequences in Fig. 2, Fig. 3 and Fig. 4. Addi-  
 036 tionally, we supply videos that more effectively demonstrate  
 037 the consistency within the videos. We strongly recommend  
 038 that readers review these materials for a more comprehen-  
 039 sive understanding.

### 040 References

- 041 [1] Yoni Kasten, Dolev Ofri, Oliver Wang, and Tali Dekel.  
 042 Layered neural atlases for consistent video editing. *ACM*  
 043 *Trans. Graph.*, 40(6):1–12, 2021. 1  
 044 [2] Jordi Pont-Tuset, Federico Perazzi, Sergi Caelles, Pablo Ar-  
 045 belaez, Alexander Sorkine-Hornung, and Luc Van Gool. The  
 046 2017 DAVIS challenge on video object segmentation. *arXiv*  
 047 *preprint arXiv:1704.00675*, 2017. 1



Prompt: CG Style, Pink Hair



Prompt: Tifa from the Final Fantasy



Figure 2. More results

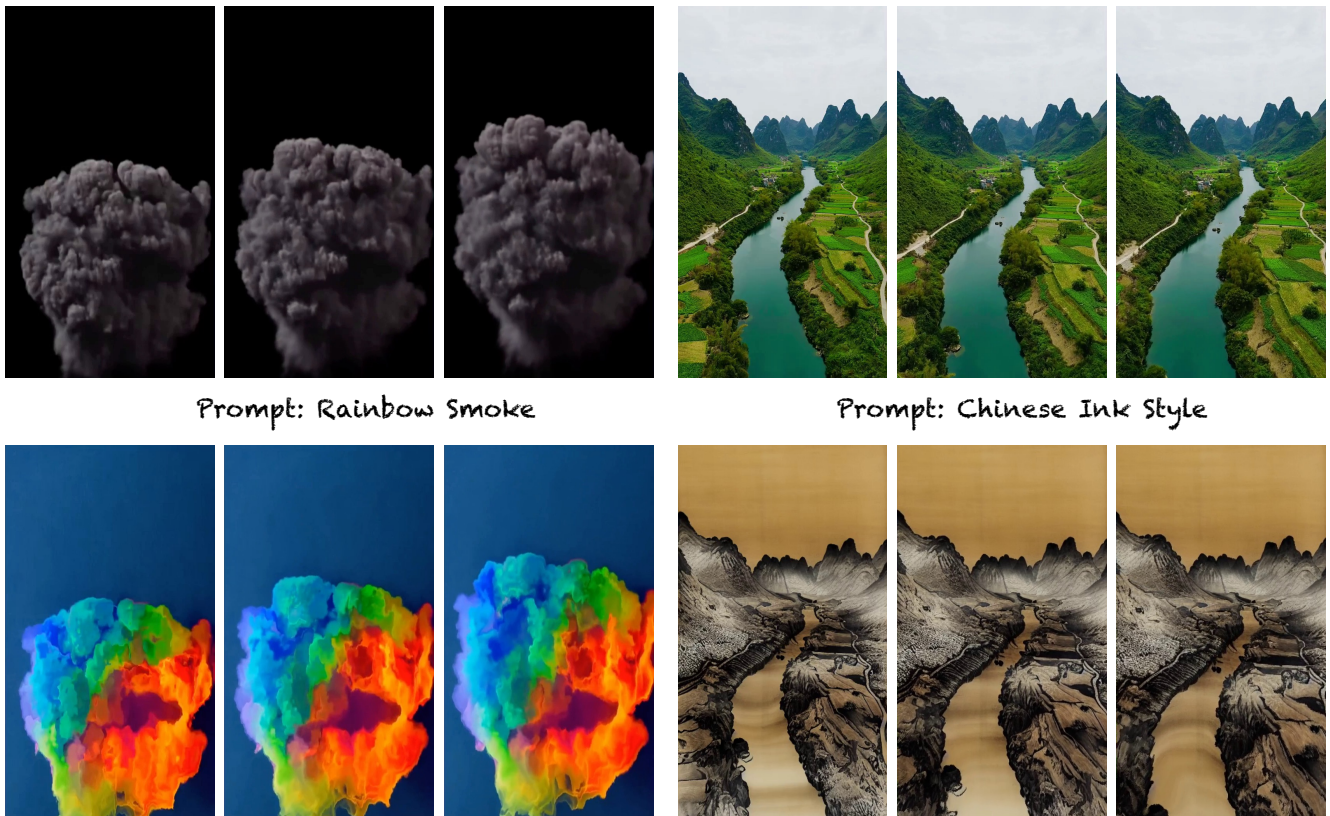


Figure 3. More results

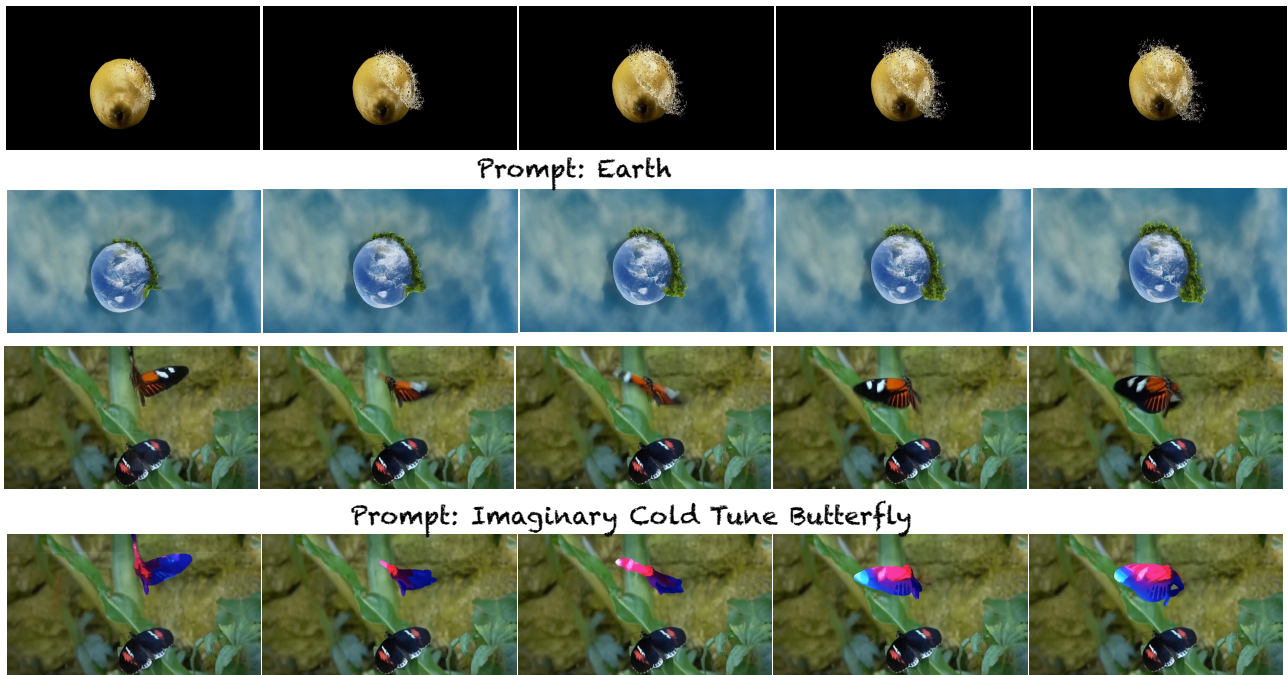


Figure 4. More results