AI Capstone Final Project Report - Editable Video Colorization

Team ID : 17 Team Name : Nuguseyo
Team member : 110550090 王昱力, 110550093 蔡師睿, 110550110 林書愷

1 Motivation

Speaking of low-level vision techniques, such as image colorization, image inpainting, and image super-resolution, there are many methods that already exist to meet these needs. Building on these advancements, we want to extend this concept to video processing. Video processing often encounters issues with temporal consistency, leading to flickering and other artifacts. Additionally, existing methods offer limited functionality for user editing, restricting the ability to customize the colorization process according to personal preferences.

Therefore, the main goal of this project is to develop a method for converting grayscale video into colorized video, addressing the temporal consistency issues and enhancing the functionality for user editing. By enabling users to colorize the video themselves, we aim to provide a more flexible tool that combines automated processes with manual adjustments. This approach not only leverages the power of advanced algorithms but also incorporates the user's creative input.

2 Techniques used

2.1 RAFT: Recurrent All-Pairs Field Transforms for Optical Flow [1]

Optical flow is the task of estimating per-pixel motion between video frames. The feature encoder extracts per-pixel features from input images, and these features can construct a set of 4D correlation volumes. RAFT then iteratively updates the estimated optical flow using the correlation volumes and features from the context encoder.

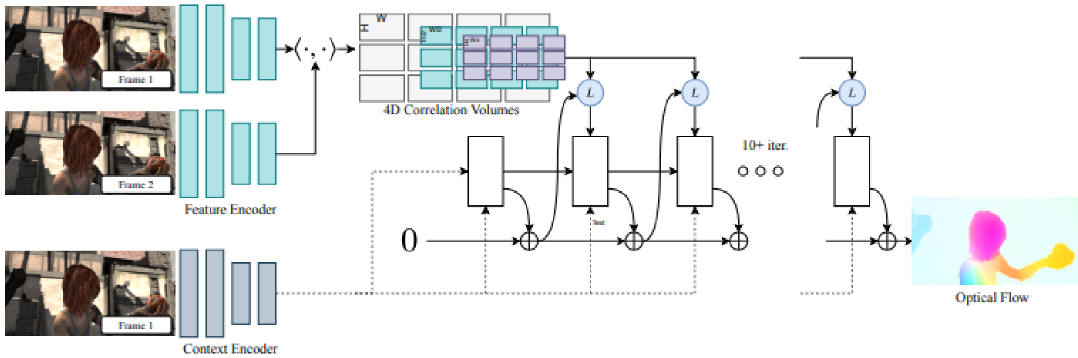

Figure 1: Pipeline of RAFT.

To achieve the desired video results, we first need to extract the optical flow of video sequences using RAFT as a preprocessing step.

2.2 CoDeF: Content Deformation Fields for Temporally Consistent Video Processing [2]

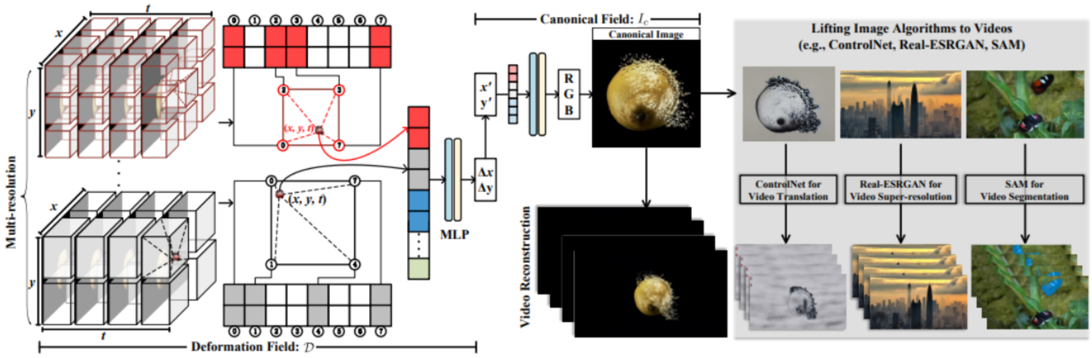

Figure 2: Pipeline of CoDeF.

CoDeF enables the factorization of an arbitrary video into a 2D content canonical field and a 3D deformation field. Each field is implemented with a multi-resolution 2D or 3D hash table using an MLP. We can thus apply image algorithms to the canonical content field, which is the canonical image, and then propagate the results along the time axis through the deformation field.

We selected this as our backbone model to convert the input video into a 2D canonical image. Following application of any image algorithm, the video can be reconstructed using the canonical image.

2.3 DDColor: Towards Photo-Realistic Image Colorization via Dual Decoders [3]

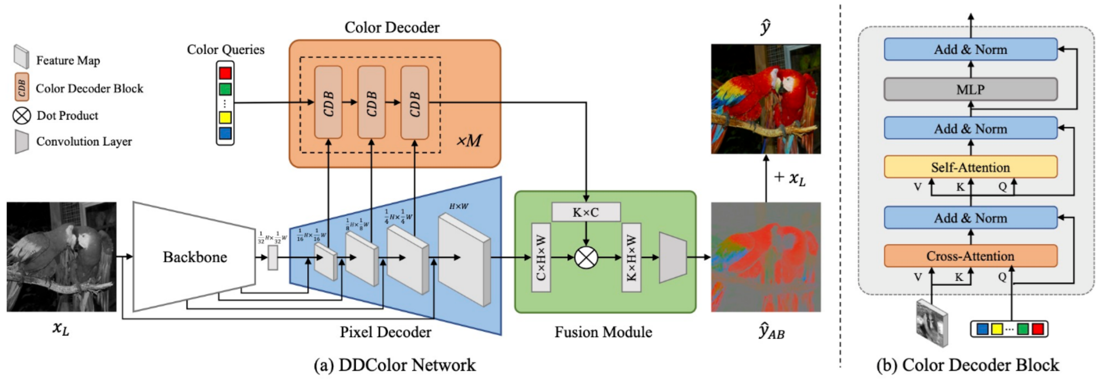

Figure 3: Pipeline of DDColor.

DDColor, to our best knowledge, is currently the SOTA (state of the art) in the field of auto image colorization. It utilizes multi-scale visual features to optimize learnable color tokens (i.e. color queries). As a result, we choose this model for image colorization task.

3 Our Method

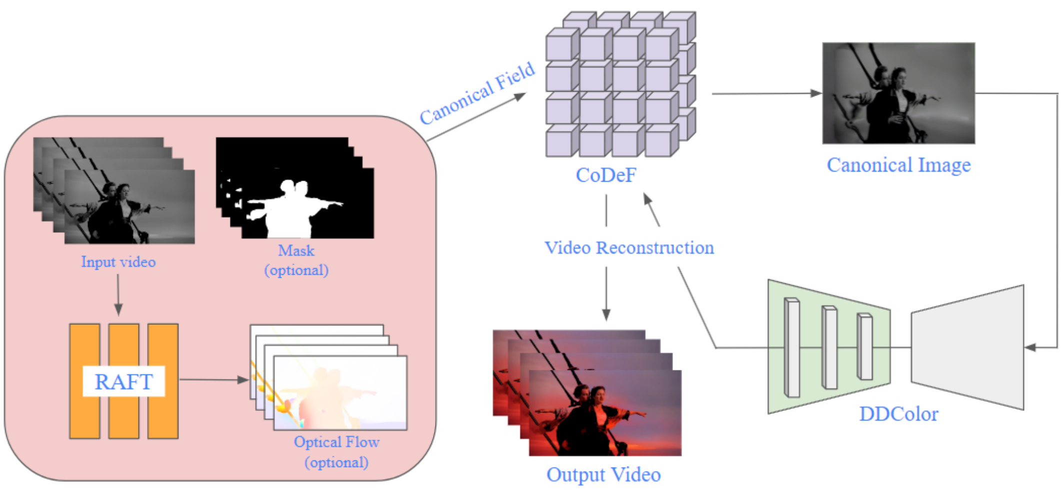

Figure 4: Pipeline of our method

Based on Figure 4, we obtain the canonical image of the input video via CoDeF, and apply the image colorization model, DDcolor, to the canonical image. Finally, we reconstruct the canonical images back into the video according to the features extracted by CoDeF.

Additionally, an advantage of our pipeline is the flexibility to substitute the DDColor component with any colorization model of your choice, or even to draw manually.

4 Experiments

4.1 Comparisons

We compare different colorization methods, such as Deoldify[4] and DDColor which colors videos frame by frame.

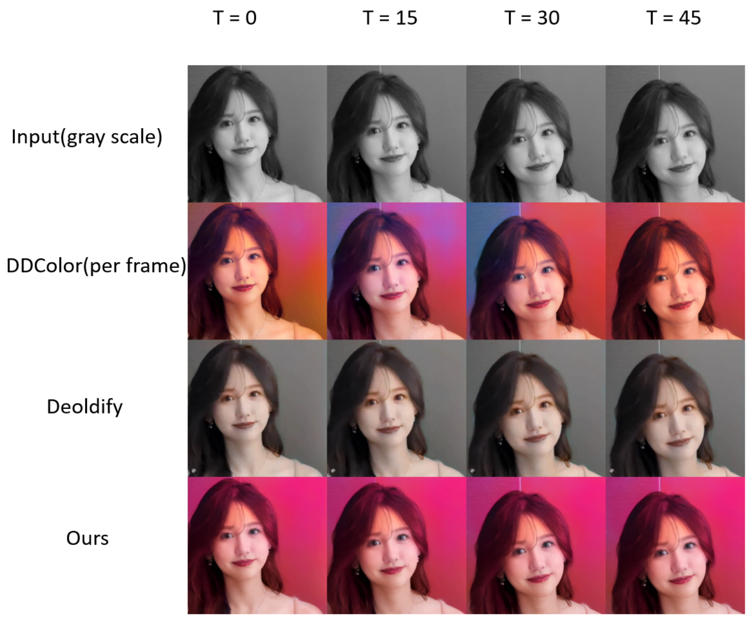

Figure 5: Qualitative comparison (frames 0, 15, 30, and 45 selected for visualization)

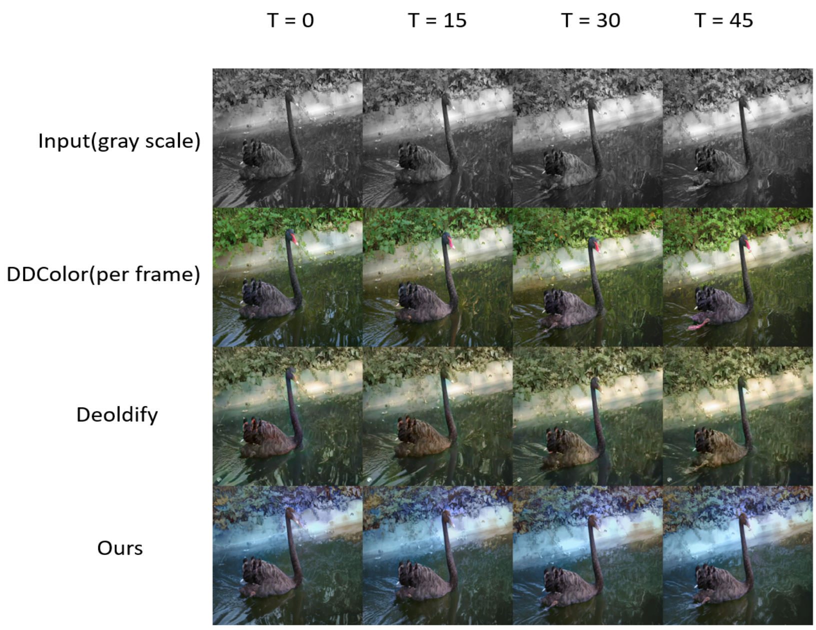

Figure 6: Qualitative comparison (frames 0, 15, 30, and 45 selected for visualization)

4.2 Edit yourself

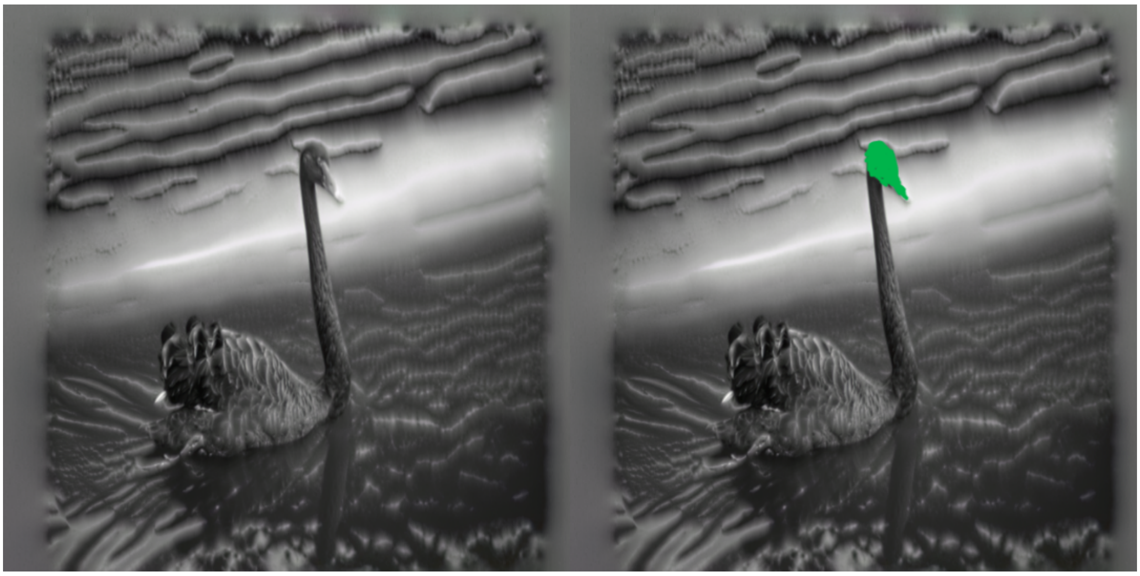

Figure 7: Left : Original canonical image; Right : We color the swan with green head on the canonical image

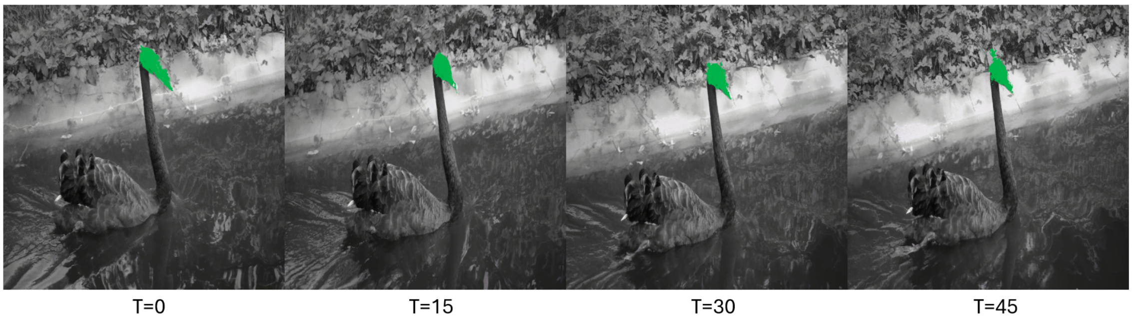

Figure 8: Restored video after editing (frames 0, 15, 30, and 45 selected for visualization).

5 Results

Video Result with comparisons : Link
User editing video result (we color the swan with a green head, and the operation is performed on the canonical image) : Link2

6 Failure cases

During our experiments, we found that CoDeF often fails with fast motion videos, resulting in artifacts on the canonical image. Due to the failure of the canonical image, the image colorization model may produce unsatisfactory results or even fail altogether.

Figure 9: Failed canonical image for the break dance video

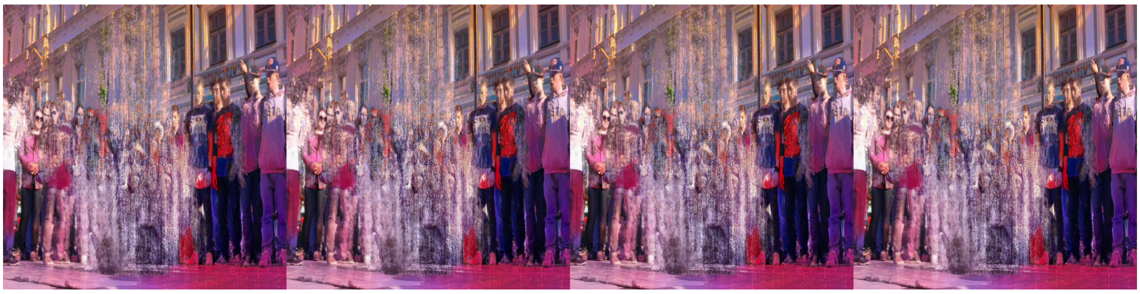

Figure 10: Restored video from the failed canonical image of the break dance sequence (frames 0, 15, 30, and 45 selected for visualization).

7 Learned from project

After this project, we have found that 'transferring video style' is possible. Additionally, while exploring the structure of CoDeF, we discovered that a major advantage of this model is that it enables us to apply image processing algorithms to video, thereby expanding the variety of image processing techniques available. Although some samples did not meet our criteria during the training process, we are glad to have identified the limitations of this model.

8 Future Work

During the training session, we discovered that CoDeF indeed can't handle video with fast-moving objects or scenes. To deal with this problem, we can divide video into multiple canonical images and 3D hashing maps which can significant decreases the carried amount of information of each mapping. Although it may increases time of training session, it is still worth doing for the goal of quality improvement.

9 Contributions

王昱力: model running, report
蔡師睿: model running, result comparison, report
林書愷: model running, video editing, report

References

[1] Z. Teed and J. Deng, "Raft: Recurrent all-pairs field transforms for optical flow," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, Springer, 2020, pp. 402–419.

[2] H. Ouyang, Q. Wang, Y. Xiao, *et al.*, "Codef: Content deformation fields for temporally consistent video processing," *arXiv preprint arXiv:2308.07926*, 2023.

[3] X. Kang, T. Yang, W. Ouyang, P. Ren, L. Li, and X. Xie, "Ddcolor: Towards photo-realistic image colorization via dual decoders," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 328–338.

[4] J. Antic, *A deep learning based project for colorizing and restoring old images (and video!)* `https://github.com/jantic/DeOldify`, 2019.