Supplementary Material for Revitalizing Legacy Video Content: Deinterlacing with Bidirectional Information Propagation

Zhaowei Gao[*1, 2]
zhagao@student.ethz.ch
Mingyang Song[*1, 2]
misong@student.ethz.ch
Christopher Schroers[2]
christopher.schroers@disneyresearch.com
Yang Zhang[2]
yang.zhang@disneyresearch.com

[1] ETH Zurich
Zurich, Switzerland
[2] DisneyResearch|Studio
Zurich, Switzerland

1 Overview

The supplementary material provides some more specific information as follows:

- For reproducibility, we provide in-depth implementation details, including model parameters, training setting and evaluation metrics.
- For enhancing the understanding and appreciation of our proposed method's effectiveness, we have included a video demonstration in the Supplementary Materials.

2 More implementation details

2.1 Model parameters

We use a pretrained SpyNet[4] for optical flow estimation. Due to the pyramid structure of SpyNet, we obtain flow at 4 different scales. We have designed two networks with different amounts of parameters, namely *Ours-S* contains 0.5M and *Ours-L* contains 9M parameters. The number of FRBs was set to 7. The number of S-NAF blocks for each FRB was set to 3 and 6 for Ours-S and Ours-L, respectively. The hyperparameter of the *dim* in Table. 1 was designed for the feature input channel in each Flow-guided deformable alignment (FGDA) block as follows, where *dim* = [20, 20, 20, 20, 20, 20, 20] and *dim* = [20, 40, 80, 160, 80, 40, 20] are applied for Ours-S and Ours-L model. The DCN kernel size was set to 3 and the number of deformable groups was set to 4.

* These authors contributed equally to this work

Layer	Conv^O	Conv^M
1.	conv(dim*2+2, dim, 3)	
2.	LeakyReLU(0.1)	
3.	conv(dim, dim, 3)	
4.	LeakyReLU(0.1)	
5.	conv(dim, dim, 3)	
6.	LeakyReLU(0.1)	
7.	conv(dim,288,3)	conv(dim,144,3)

Table 1: The architecture of $\text{Conv}^{O,M}$ is designed to compute the offset and mask required by the deformable convolution network in the FGDA Block.

2.2 Training Setting

We adopt AdamW [3] optimizer, and the learning rate decays from 1×10^{-4} to $1 \times 10{-}7$ with Cosine Annealing [2] scheduler. The training process consists of 600K iterations. The batch size was set to 8 and the patch size was 128×128. Our models were end-to-end trained via a L_1 loss function. All the experiments were performed on one Nvidia GeForce RTX 3090.

2.3 Evaluation Metrics

To conduct a comprehensive evaluation, we compare our approach to previous methods in terms of restoration accuracy and inference efficiency. In order to fairly compare with existing methods, we followed the evaluation method in [7]. To assess the fidelity, we employ Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [6] as evaluation metrics. The reported scores were derived by calculating the average scores across the entire test set. As for efficiency evaluation, we calculate the runtime based on an image crop with 256 $\times$256 resolution on various models with similar amounts of parameters.

3 Video Visualization

In the Supplementary Material, we provide video demonstrations that showcase the deinterlacing results obtained with our method compared to those using VFIT[5] and Liu's[1] techniques. These comparisons are crucial for illustrating the practical improvements our approach offers in the temporal domain. Notably, both the VFit and Liu methods exhibit temporal flickering artifacts, which can degrade the viewer's experience by causing inconsistencies in video playback. Such artifacts are imperceptible in the results obtained from our model, providing a clear visual testament to the enhanced stability and quality of our method. For a detailed comparison, please refer to the video examples in the Supplementary Material.

References

[1] Yuqing Liu, Xinfeng Zhang, Shanshe Wang, Siwei Ma, and Wen Gao. Spatial-temporal correlation learning for real-time video deinterlacing. In *2021 IEEE International Con-*

ference on Multimedia and Expo (ICME), pages 1–6. IEEE, 2021.

[2] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

[3] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[4] Anurag Ranjan and Michael J Black. Optical flow estimation using a spatial pyramid network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4161–4170, 2017.

[5] Zhihao Shi, Xiangyu Xu, Xiaohong Liu, Jun Chen, and Ming-Hsuan Yang. Video frame interpolation transformer. In *CVPR*, 2022.

[6] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13 (4):600–612, 2004. doi: 10.1109/TIP.2003.819861.

[7] Yin-Chen Yeh, Jilyan Dy, Tai-Ming Huang, Yung-Yao Chen, and Kai-Lung Hua. Vdnet: video deinterlacing network based on coarse adaptive module and deformable recurrent residual network. *Neural Computing and Applications*, 34(15):12861–12874, 2022.