Neural Residual Flow Fields for Efficient Video Representations: Supplementary Materials

Daniel Rho[1 [0000−0002−8568−9489]], Junwoo Cho[1], Jong Hwan Ko[1,2⋆], and Eunbyung Park[1,2⋆]

[1] Department of Artificial Intelligence, Sungkyunkwan University
[2] Department of Electrical and Computer Engineering, Sungkyunkwan University
{daniel231,jwcho000,jhko,epark}@skku.edu

1 Performance Comparison using MPI SINTEL Videos

1.1 Experiment Setup

In this section, we present all the measured performance on all 23 MPI SINTEL [1] videos of five methods: H.264 [2], color-based baseline (SIREN [3]), our approach with a single reference, multiple references, and lastly, multiple references with separate optical flow and residual models. We used every frame in each video at its original resolution (436 x 1024). We set the size of the group of pictures (GOP) to five for all methods.

In this work, we used H.264 for key frame compression in order to compare the compression performance of NRFF with H.264. Given a key frame (I-frame), H.264 encodes optimal block-wise flows and residuals, and NRFF uses a neural network to compress pixel-wise flows and residuals. Since key frames are encoded in the same way as in H.264, we can compare how well each method compresses flows and residuals. This also enables performance comparison between a single reference and multiple reference NRFFs. When comparing the compression performance of those five methods, keep in mind that we did not apply any network compression methods to neural field-based methods.

1.2 Results

Fig. 2 shows the results of five methods on MPI SINTEL videos. We measured the performance using PSNR and SSIM. The x-axis of each graph is bits per pixel (bpp). On most videos, optical flow and the residual-based approach (NRFF) perform better than or at least similar to the color-based approach (SIREN), and on more than half of the videos, it outperforms by a large margin. NRFF shows lower video quality in *ambush_2* or *ambush_6*, which are characterized by the abrupt appearance or rapid movement of an object that spans a substantial amount of visual area. Regarding the number of reference frames, a single reference NRFF showed inferior performance compared to the multiple reference version even with five times longer training time. Splitting the network improved

⋆ Corresponding authors.

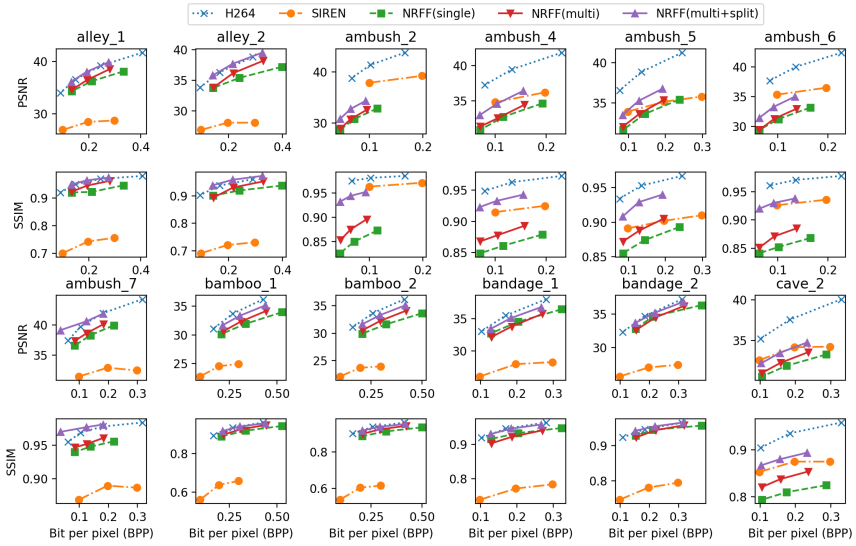

Fig. 1: Performance Comparison using MPI SINTEL Videos (1/2)

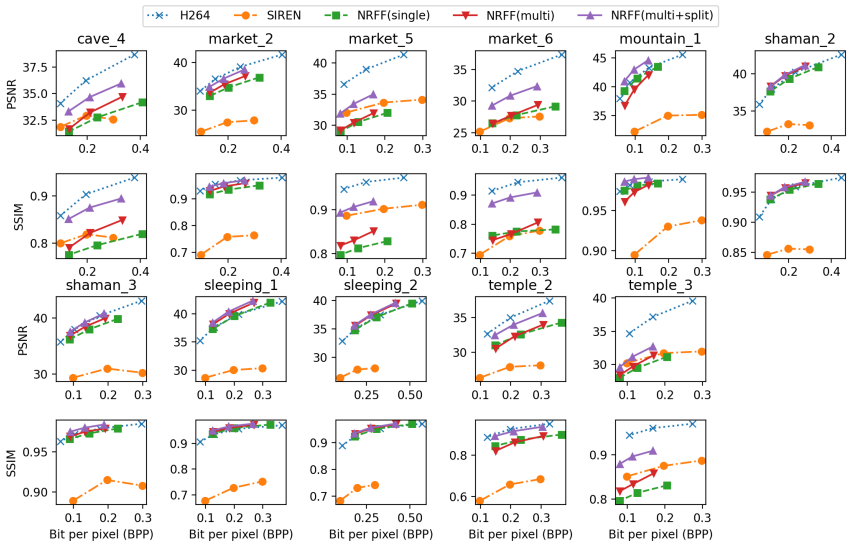

Fig. 2: Performance Comparison using MPI SINTEL Videos (2/2)

video quality in most cases and, surprisingly, never degraded video quality. The assumption that optical flows and residuals have different dynamics appears to be supported by these results.

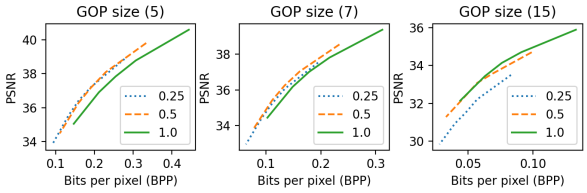

Fig. 3: The rate distortion curves with different network and GOP sizes in Alley_1. The size of group of pictures (the total number of frames for each group of pictures) are shown inside the parenthesis. 0.25, 0.5, 1.0 are ratios between network size and the keyframe size.

2 The ratio of network size and keyframe size

As shown in Fig. 3, we found that the optimal ratio of network size and keyframe size is proportional to the size of the group of pictures. For example, the ratio of 0.25, which means the network size is one quarter of the key frame size, was optimal in a small GOP, while the larger GOP requires a much higher ratio.

3 Batch Size

Due to the fact that our proposed method does not restrict the range of optical flows, the training process for some videos may be unstable. We solved this issue by increasing the batch size to be more than one.

References

1. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.: A naturalistic open source movie for optical flow evaluation. In: Proceedings of the European Conference on Computer Vision (ECCV). (2012)
2. Wiegand, T., Sullivan, G., Bjontegaard, G., Luthra, A.: Overview of the h.264/avc video coding standard. IEEE Transactions on Circuits and Systems for Video Technology **13** (2003) 560–576
3. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H., eds.: Advances in Neural Information Processing Systems. Volume 33., Curran Associates, Inc. (2020) 7462–7473