

Supplementary Materials: Rate-aware Compression for NeRF-based Volumetric Video

Anonymous Authors

1 DECODING PIPELINE

Fig. 1 illustrates the decoding pipeline of our compression framework. Initially, the occupancy grid, rendering MLP, and implicit entropy model are decoded from the bitstream. Subsequently, voxels in the explicit representation are decoded from the bitstream in causal order. Using the already decoded voxels as context, the distribution of undecoded voxels is predicted through an implicit entropy model. Based on the parameters of the distribution, voxels are decoded from the bitstream.

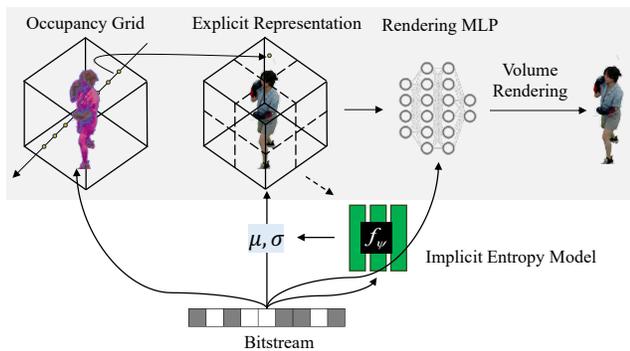


Figure 1: Decoding pipeline.

2 SUPPLEMENTARY EXPERIMENTAL RESULTS

Fig. 2 displays the subjective results with and without the use of adaptive quantization strategy on the *actor02* sequence of the HumanRF [2] dataset. It is observable that at similar bitrate, the reconstruction results utilizing adaptive quantization yield higher PSNR and provide more reliable visual details.

Fig. 3 illustrates the R-D curves from the ablation studies. It is evident that removing any module from the full model significantly diminishes R-D performance, thereby validating the effectiveness of our proposed modules.

Fig. 4 illustrates the results of applying our compression framework to the plane-based representation, TensorRF [1]. Notably, in comparison to the baseline, our method achieves compression rates several times higher.

3 ADDITIONAL EXPERIMENTS

The experiments section has already demonstrated that employing an implicit entropy model to estimate the bitrate, in conjunction with a joint loss function for end-to-end optimization, significantly enhances the rate-distortion performance compared to the baseline. To further illustrate the performance gains derived from utilizing the implicit entropy model, we replaced it within the framework with a simple method of rate estimation. Specifically, we assumed a



(a) adaptive quantization PSNR: 34.09 dB Size: 91.71 KB
(b) w/o adaptive quantization PSNR: 33.08 dB Size: 89.29 KB

Figure 2: Qualitative comparisons of with and w/o adaptive quantization.

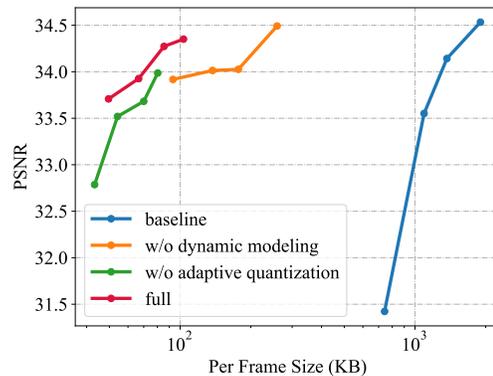


Figure 3: R-D curves of ablation studies.

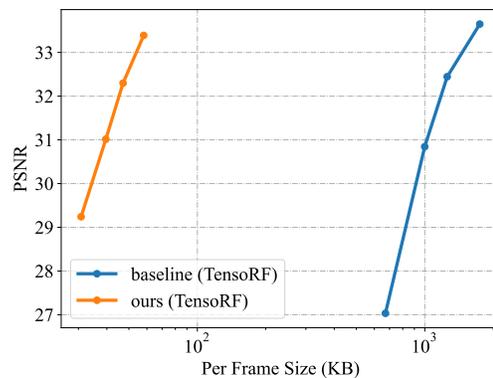


Figure 4: Results of applying our compression framework to TensorRF [1].

identical distribution across all voxels in the explicit representation and estimated the mean and variance of the Laplace distribution using two trainable parameters, μ and σ , to calculate the bitrate. Fig. 5 depicts the comparative results of substituting the framework’s implicit entropy model with two trainable parameters. It is evident that, although the simple estimation method also improves R-D performance relative to the baseline, the framework employing the implicit entropy model achieves superior R-D performance due to its more precise rate estimation.

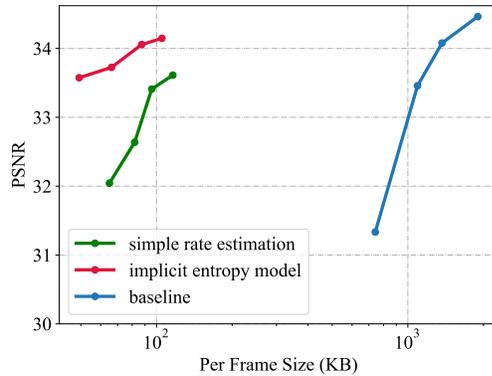


Figure 5: Comparison of R-D performance between using implicit entropy model and simple rate estimation method for rate estimation.

4 MORE VISUAL RESULTS

In this section, we will present more subjective results. Fig 6 displays qualitative comparison results on the HumanRF’s *actor05* sequence and the ReRF’s *sing* sequence.

Furthermore, in the supplementary materials folder, we have provided demo videos for the ReRF’s *kpop* sequence [3] and HumanRF’s *actor07* sequence [2], with the bitrate for each frame annotated in the top right corner of the videos.

REFERENCES

- [1] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. 2022. Tensorf: Tensorial radiance fields. In *European Conference on Computer Vision*. Springer, 333–350.
- [2] Mustafa Işık, Martin Rünz, Markos Georgopoulos, Taras Khakhulin, Jonathan Starck, Lourdes Agapito, and Matthias Nießner. 2023. Humanrf: High-fidelity neural radiance fields for humans in motion. *ACM Transactions on Graphics (TOG)* 42, 4 (2023), 1–12.
- [3] Liao Wang, Qiang Hu, Qihan He, Ziyu Wang, Jingyi Yu, Tinne Tuytelaars, Lan Xu, and Minye Wu. 2023. Neural Residual Radiance Fields for Streamably Free-Viewpoint Videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 76–87.

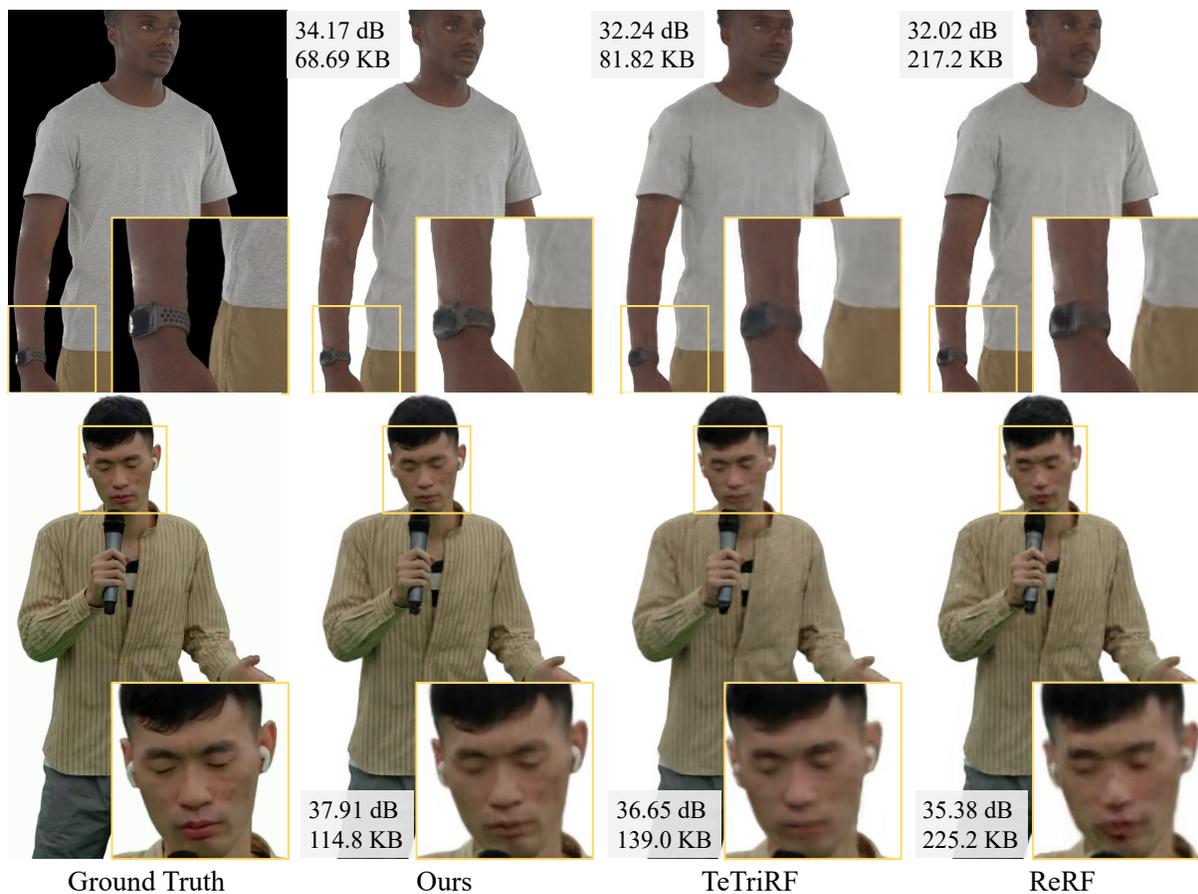


Figure 6: Qualitative comparisons of *actor05* sequence from HumanRF [2] and *sing* sequence from ReRF [3].