

Supplementary Materials: Hierarchical Progressive Coding Framework for Volumetric Video

Anonymous Authors



Figure 1: The rendered outcomes of our HPC framework on different datasets.

1 MORE QUALITATIVE RESULTS

The enhanced rendering outcomes from our HPC framework applied to the ReRF[2] and DNA-Rendering[1] datasets further demonstrate the effectiveness of our method across diverse datasets. Additionally, this showcases our method’s capability to support variable bitrate streaming within a single model.

2 PROGRESSIVE TRAINING STRATEGY

A more detailed exposition of the progressive training strategy is presented in Algorithm.(1).

3 MORE DETAILED CONFIGURATIONS

In all comparative experiments, we set the bounding boxes for the entire DNA dataset to $[[[-0.7, -0.2, -0.7], [0.7, 1.55, 0.7]]]$. For the ReRF dataset, the bounding boxes were set as follows: the kpop sequence to $[[[-0.6, -0.5, -1.25], [0.6, 0.7, 0.1]]]$, the box sequence to $[[[-0.55, -0.5, -1.18], [0.55, 0.55, 0]]]$, and the sing sequence to $[[[-0.85, -0.5, -1.15], [0.35, 0.3, 0.1]]]$.

To ensure fairness, all manual post-rendering adjustments to the background based on masks were prohibited, and the resolution of the generated images was set to match that of the input.

Algorithm 1 Progressive Training Strategy

```

1: Initialize residual feature grids  $R_t^1, R_t^2, \dots, R_t^L$ 
2: Initialize learning parameters  $\lambda_1, \lambda_2$ , set maximum iterations  $maxiter$ , and active iteration step  $actiter$ 
3: Get  $\hat{F}_{t-1}^1, \hat{F}_{t-1}^2, \dots, \hat{F}_{t-1}^L$  from the reference frame buffer
4: Fix higher resolution grids  $R_t^2, R_t^3, \dots, R_t^L$ 
5: Start training on  $R_t^1$  with initial low-resolution
6: for  $it \leftarrow 1$  to  $maxiter$  do
7:   if  $it == l \times actiter$  then
8:     Start gradient backpropagation for  $R_t^{l+1}$ 
9:     Begin entropy estimation and simulate quantization on  $R_t^{l+1}$ 
10:   end if
11:   if  $it == maxiter - l \times actiter$  then
12:     Stop training on  $R_t^{l-l}$ 
13:     if current frame is keyframe then
14:       Set  $\lambda_1, \lambda_2 \leftarrow 0$  for refining the training focus
15:     end if
16:   end if
17:   Update learning rates and regularization parameters dynamically
18:   Combine activated  $R_t^l$  with corresponding  $\hat{F}_{t-1}^l$  to get  $\hat{F}_t^l$ 
19:   Calculate the corresponding  $L^l$  and perform gradient backpropagation
20: end for
21: Put well-trained  $\hat{F}_t$  into the reference frame buffer

```

REFERENCES

[1] Wei Cheng, Ruixiang Chen, Siming Fan, Wanqi Yin, Keyu Chen, Zhongang Cai, Jingbo Wang, Yang Gao, Zhengming Yu, Zhengyu Lin, et al. 2023. Dna-rendering: A diverse neural actor repository for high-fidelity human-centric rendering. In

Proceedings of the IEEE/CVF International Conference on Computer Vision. 19982–19993.
[2] 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 (CVPR)*. 76–87.