NeuSmoke: Efficient Smoke Reconstruction and View Synthesis with Neural Transportation Fields (Supplementary Materials)

ACM Reference Format:
. 2024. NeuSmoke: Efficient Smoke Reconstruction and View Synthesis with Neural Transportation Fields (Supplementary Materials). In *SIGGRAPH Asia 2024 Conference Papers (SA Conference Papers '24), December 3–6, 2024, Tokyo, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3680528.3687667

Implementation Details. For the first stage, we select TiNeuVox-B as the basic model which has $160^3 \times 6$ neural voxels. The channel dimension of hidden layers is 256. The dimension of time embeddings and voxel feature are set as 30. The frequency number of positional encoding for neural voxels is 2. 2048 rays are sampled randomly for a batch and it takes a totaly 20000 iterations during training. For the second stage, the encoder part consists of four blocks, each block down-samples the input feature as $\frac{1}{2}\times$ resolution. The decoder part upsamples features from the encoder with time embeddings and adds them with previous features of the encoder part. To enhance the representation ability of the whole model, we introduce spatial and channel attention before the last layer. It takes 800 epochs in total to train the second stage. The perceptual loss is weighted by 0.1. For both stages, we select Adam as the optimizer and train two models on one single GeForce RTX 3090 GPU.

Network architecture. Fig. 3 shows the details of adopted CNN networks in the second stage.

Evaluation metric of volume density estimation. In this task, volume density estimation is unsupervised learned from videos. We focus on the structure similarity between our results and reference. Then we design a novel metric to evaluate this performance. Firstly, we slice the volume along three axes respectively and obtain three sets of slice collections. Then we adopt the traditional SSIM to estimate the structure similarity of each slice and average them to get three values (ssim_1, ssim_2 and ssim_3). The final metric value $VSSIM$ (Density (SSIM)) is generated by:

$$VSSIM = \frac{ssim\_1 + ssim\_2 + ssim\_3}{3}. \tag{1}$$

Comparison on each scene. Tab. 1 shows comparison on novel view synthesis and volume density estimation in each scene.

Qualitative comparison on multiple time stamps. Tab. 1 shows a qualitative comparison between the multi-time-stamp scheme and the temporal interpolation scheme on novel view synthesis. Our method generates cleaner results.

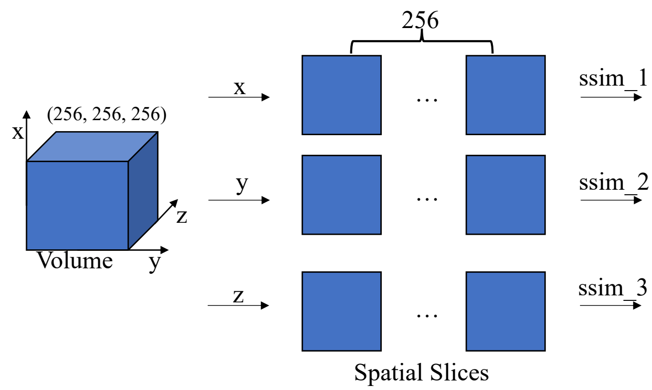

Fig. 1. Details of the metric of volume density estimation.

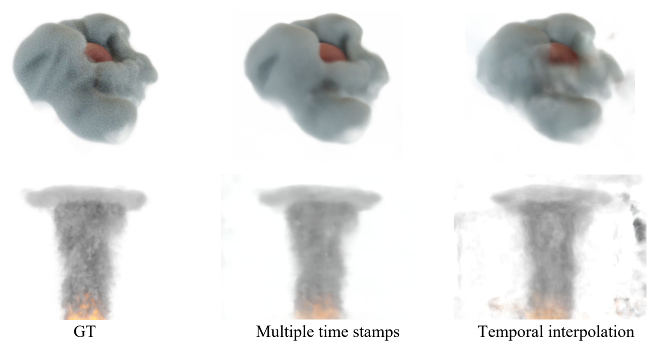

Fig. 2. Qualitative comparison between the multi-time-stamp scheme and the temporal interpolation scheme on novel view synthesis.

SA Conference Papers '24, December 3–6, 2024, Tokyo, Japan

ACM ISBN 979-8-4007-1131-2/24/12
https://doi.org/10.1145/3680528.3687667

Table 1. Comparison on novel view synthesis and volume density estimation. 'Ours': Our results of the first stage. 'Ours_r': Our results of the overall pipeline.

Scene	PSNR↑				SSIM↑				LPIPS↓				Density (SSIM)↑		
	PINF	TiNeuVox	Ours	Ours_r	PINF	TiNeuVox	Ours	Ours_r	PINF	TiNeuVox	Ours	Ours_r	PINF	TiNeuVox	Ours
Scalar	31.41	26.74	32.03	32.29	95.17	93.83	94.30	95.35	23.27	25.66	29.21	22.52	-	-	-
Sphere	28.85	22.93	31.27	30.79	85.06	73.37	88.09	86.99	7.03	10.48	6.81	5.25	-	-	-
Double	16.57	24.60	28.72	28.87	88.22	79.40	90.74	89.56	9.94	10.94	3.66	3.03	4.01	94.64	95.43
Yellow	19.01	27.99	32.50	32.47	91.29	93.02	96.06	95.96	8.30	3.04	2.48	2.06	18.92	97.72	98.23
Fire	17.45	28.61	32.97	33.29	90.91	92.41	95.83	95.36	5.30	3.27	4.57	2.95	27.40	95.45	97.51

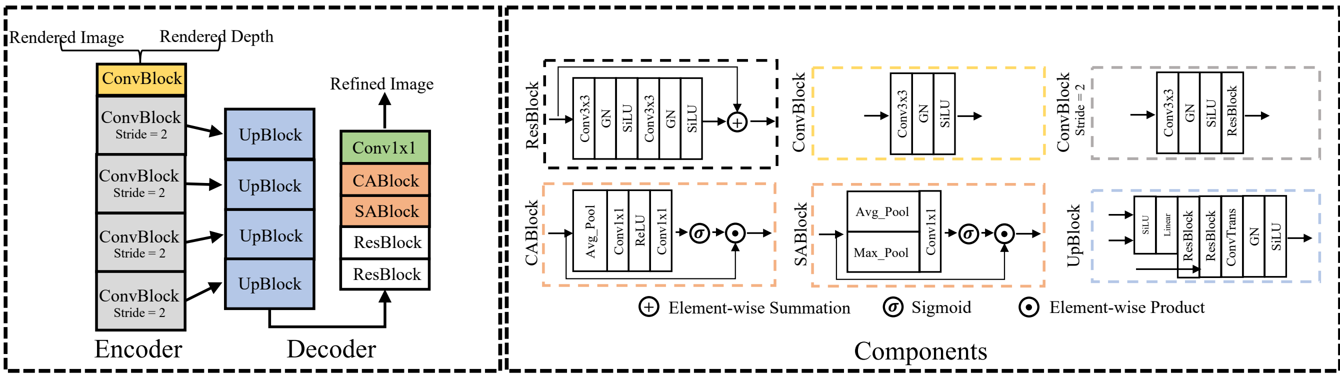

Fig. 3. Details of the encoder-decoder framework.