APPENDIX

1 ADDITIONAL RESULTS

1.1 VS

Evaluation of image resolution. In Figure 1, we evaluate the capability of CoordNet, NeRV, and InSituNet in generating images with different resolutions using the vortex data set. Both CoordNet and InSituNet produce satisfactory results under 256 image resolution. However, taking a closer comparison, the image generated by InSituNet includes noise, and the features are not preserved well, for example, at the bottom region. Using the resolutions of 512 and $1,024$, CoordNet is the clear winner, while InSituNet does not produce acceptable results. This is because InSituNet only has hundreds of training images, and most GAN-based architectures do not have enough capacity to generate high-resolution images (e.g., 512 and $1,024$) [1], [2], [3], [4]. Besides qualitative analysis, Table 1 reports average PSNR and LPIPS values. Compared with NeRV and InSituNet, CoordNet achieves the best PSNR and LPIPS values under different image resolutions.

TABLE 1: Average PSNR (dB) and LPIPS for the VS task under different image resolutions using the vortex data set.

resolution	method	PSNR ↑	LPIPS ↓
256	NeRV	20.96	0.144
	InSituNet	19.38	0.162
	CoordNet	**22.84**	**0.083**
512	NeRV	20.89	0.201
	InSituNet	19.50	0.190
	CoordNet	**23.30**	**0.105**
1,024	NeRV	19.75	0.255
	InSituNet	20.36	0.193
	CoordNet	**23.74**	**0.129**

Additional results. Figure 2 displays the synthesized images under different view parameters using the Tangaroa (V) data sets. As these images show, CoordNet preserves the overall shapes and details under diversified view parameters.

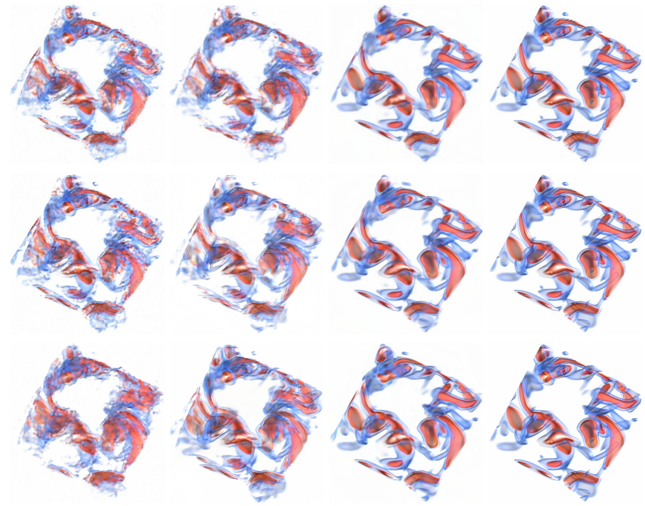
1.2 AOP

Figure 3 shows the volume rendering results with LAO. The difference image is displayed in the top-left corner for each approach. As the difference images indicate, CoordNet produces fewer differences than other methods.

1.3 TSR

Unsupervised time interpolation. Since CoordNet treats the coordinates as a continuous function; it can interpolate an arbitrary number of intermediate time steps, which is impossible with TSR-TVD. We produce six non-integer time steps between two neighboring time steps and compare their temporal coherence with LERP. The isosurface rendering results are shown in Figure 4. We can observe how the isosurfaces smoothly grow (refer to the red ellipses) and merge (refer to the blue ellipses) based on the rendering results generated by CoordNet, while LERP does not exhibit such smooth temporal variations.

Volume rendering results. Figure 5 shows the volume rendering results among TSR-TVD, CoordNet, and GT. For the combustion (MF) data set, TSR-TVD does not produce



(a) NeRV (b) InSituNet (c) CoordNet (d) GT

Fig. 1: Comparison of synthesized volume rendering images for the VS task under different image resolutions using the vortex data set. Top to bottom: 256, 512, and $1,024$ image resolutions.

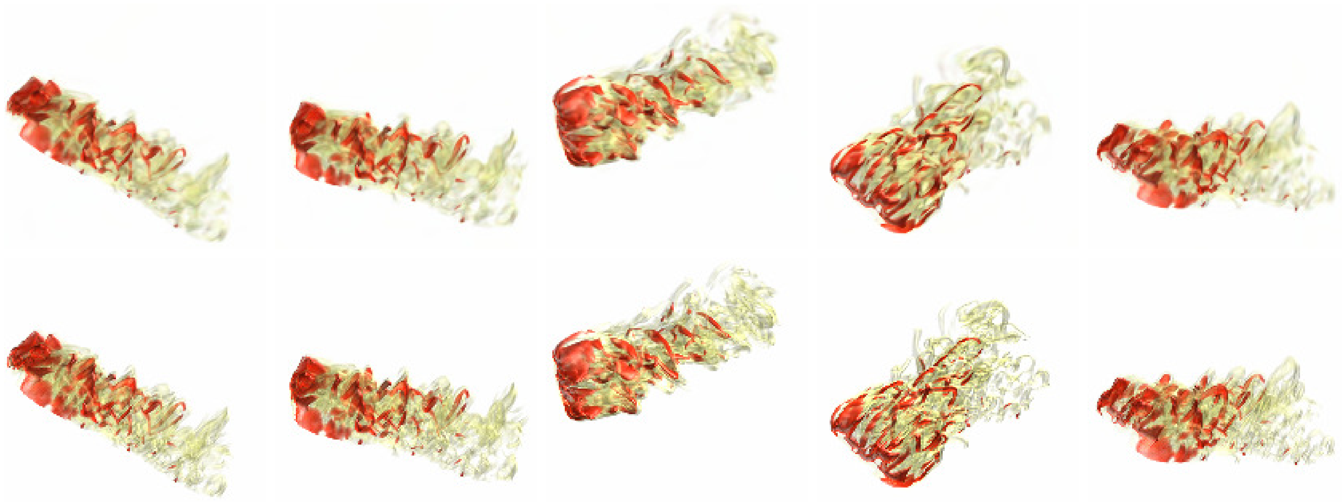
the green part at the bottom-left corner and the yellow part at the top-right corner well, while CoordNet preserves those details. For the half-cylinder (6400,V) and ionization (H2) data sets, both methods produce similar rendering results compared with GT. But taking a close comparison, the image produced by TSR-TVD contains more artifacts.

Slice of volume rendering results. Figure 6 shows a slice of volume rendering results for the TSR task. These results indicate the sharpness of the synthesized data generated by CoordNet.

Discussion. Compared with TSR-TVD, CoordNet achieves better visual quality (direct volume rendering and isosurface rendering) and better quantitative scores. Besides, CoordNet has the following advantages. (1) The interpolation process is unsupervised, which means CoordNet does not require to see the complete subsequence of early time steps for training. (2) Given two time steps, CoordNet can synthesize arbitrary numbers of time steps with coherent and high-quality results, while TSR-TVD needs to perform this recursively (i.e., the synthesized time steps are fed into TSR-TVD to produce new time steps), and the performance cannot be guaranteed due to error accumulation in the recursive process. (3) CoordNet can operate in non-uniform sampling cases, while TSR-TVD only assumes the time steps are selected uniformly.

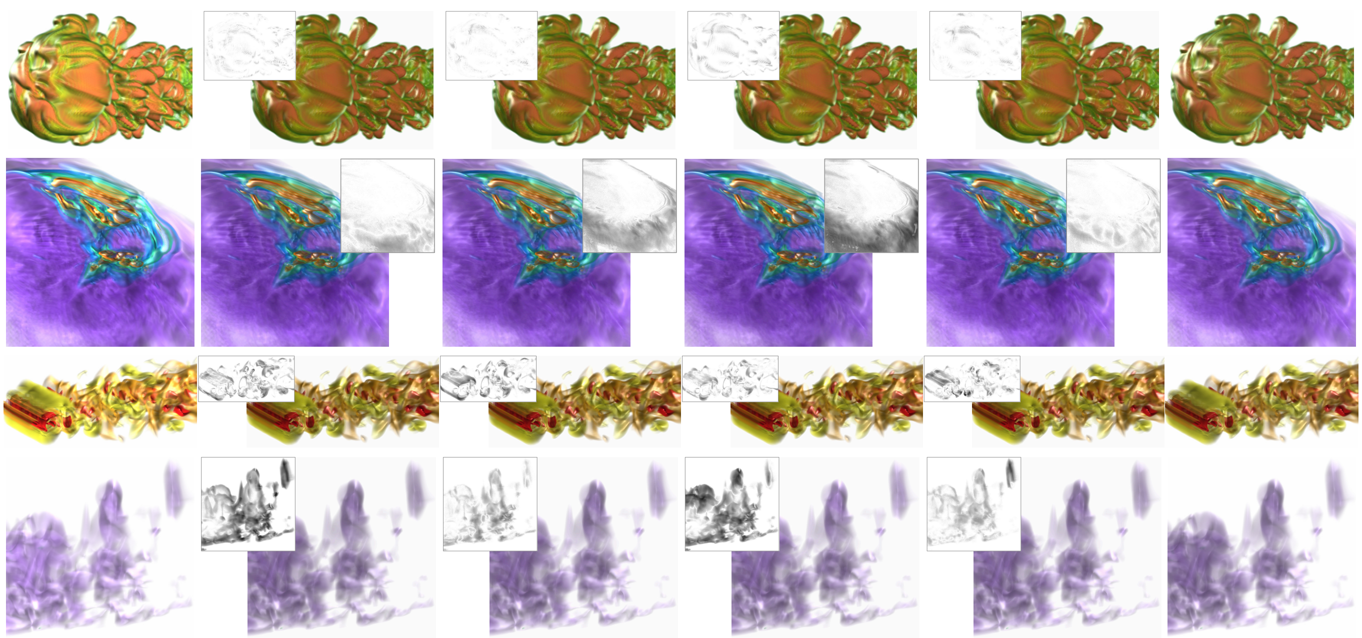
1.4 SSR

Unsupervised space interpolation. Because CoordNet processes the SSR task without supervision, it can produce higher-resolution volumes. That is, we can assume the original volumes (e.g., 128^3) are subsampled from higher-resolution volumes (e.g., 512^3), utilize these original volumes to train CoordNet, and inferCoordNet to synthesize higher-resolution ones. We use the vortex ($128 \times 128 \times 128$) and ionization (PD) ($600 \times 248 \times 248$) data sets to train CoordNet and produce volumes with higher-resolution (i.e., vortex with $512 \times 512 \times 512$ and ionization (PD) with



(a) $(9, 45)$ (b) $(27, 18)$ (c) $(45, 324)$ (d) $(90, 297)$ (e) $(135, 306)$

Fig. 2: Volume rendering results for the VS task under different view parameters (θ, ϕ) using the Tangaroa (V) data set. Top: CoordNet. Bottom: GT.



(a) GT wo. LAO (b) LAO (c) V2V (d) DVAO (e) CoordNet (f) GT w. LAO

Fig. 3: Zoom-in volume rendering with LAO results for the AOP task. Top to bottom: argon bubble, earthquake, half-cylinder (6400, VM), and Tangaroa (VM)
.

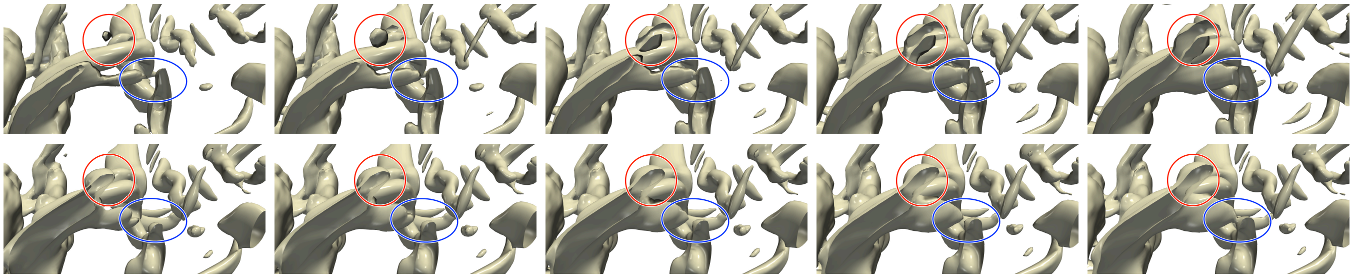
$2400 \times 992 \times 992$). We compare our results against BI. As displayed in Figure 7, CoordNet produces sharper results with fewer artifacts compared with BI (refer to the arrows in the images).

Additional results. Figure 8 displays the volume rendering results of the argon bubble, earthquake, and Tangaroa (VM) data sets. Compared with BI, CoordNet produces closer results in both shape and texture.

Slice of volume rendering results. Figure 9 shows a slice of volume rendering results for the SSR task. These results indicate that CoordNet preserves the sharpness and smoothness of the synthesized data.

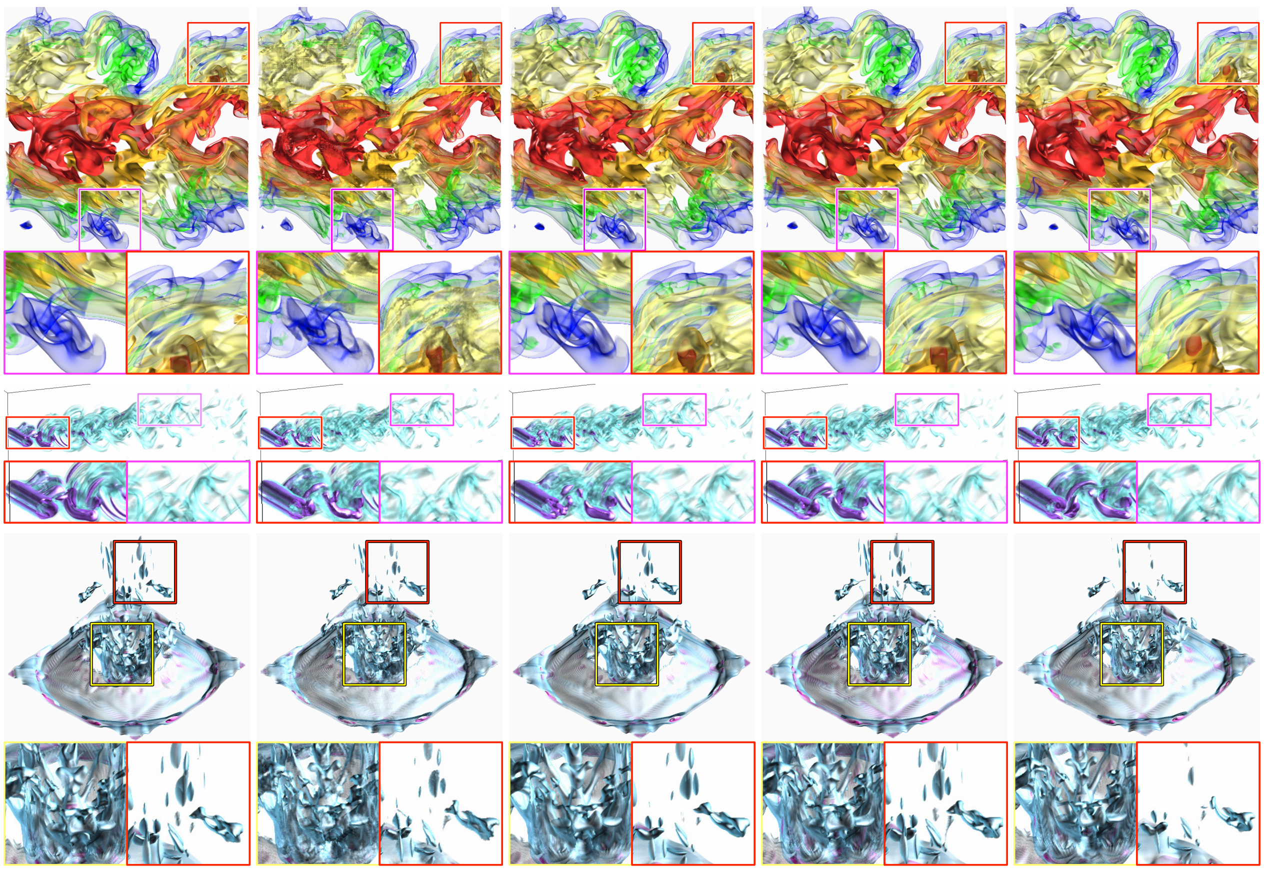
Isosurface rendering results. Figure 10 shows the isosurface rendering results among SSR-TVD, CoordNet, and GT. Both SSR-TVD and CoordNet produce close isosurface results of the combustion (HR) data set compared with GT, but SSR-TVD misses some isosurfaces at the top-right corner. For the ionization (PD) data set, SSR-TVD extracts the isosurfaces with artifacts and does not preserve the isosurface's shape in the feature region. For the vortex data set, CoordNet generates more similar isosurfaces compared to GT. For example, SSR-TVD cannot reconstruct the isosurfaces at the top-left corner.

Discussion. Compared with SSR-TVD, CoordNet achieves better visual quality and similar quantitative val-



(a) time step 95.167 (b) time step 95.333 (c) time step 95.500 (d) time step 95.667 (e) time step 95.833

Fig. 4: Zoom-in isosurface rendering results for the TSR task using the half-cylinder (640,V) data set. Top: LERP. Bottom: CoordNet. We generate 576 time steps from sparsely sampled 25 time steps. The chosen isovalue is -0.7.



(a) left observation (b) TSR-TVD (c) CoordNet (d) GT (e) right observation

Fig. 5: Volume rendering results for the TSR task with an interpolation interval of 3. Top to bottom: combustion (MF), half-cylinder (640,V), and ionization (H2). The displayed time steps are 95, 75, and 75, respectively. From top to bottom, left and right observations are 93 and 97, 73 and 77, 73 and 77.

ues. Still, CoordNet offers the following benefits. (1) The upscaling operation is completed in an unsupervised fashion. This means we do not need to store high- and low-resolution pairs for optimization. (2) CoordNet can upscale high resolution (e.g., $600 \times 248 \times 248$) to higher resolution (e.g., $2400 \times 992 \times 992$).

2 Hyperparameter Study

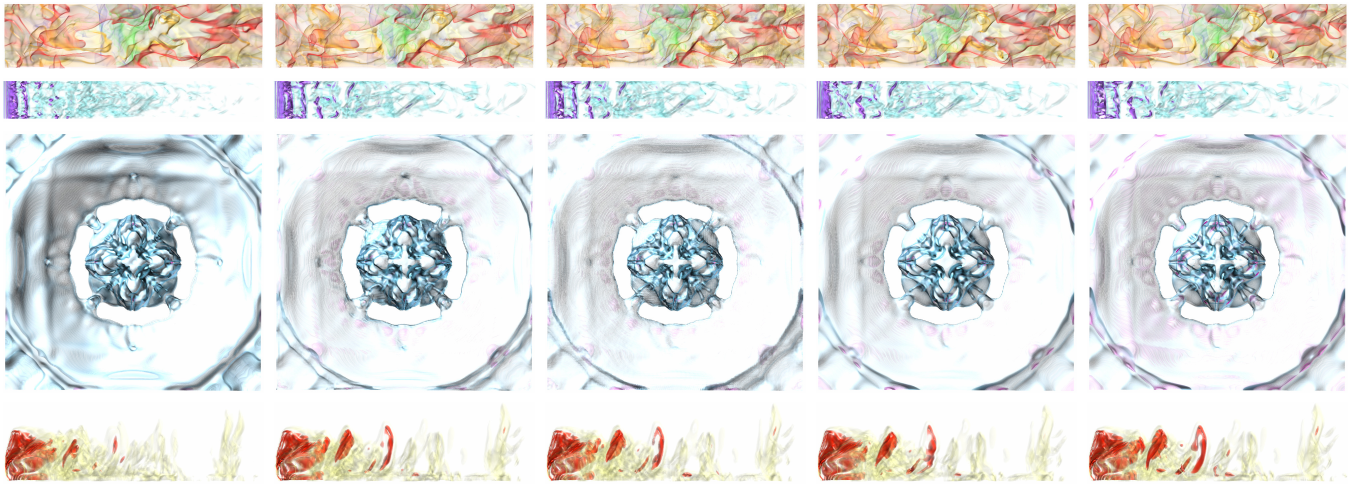
We further study the hyperparameters of CoordNet in the following aspects.

2.1 Sample Size (N)

To study the impact of sample size, we train CoordNet using different N for the TSR task. Table 2 reports the average PSNR, LPIPS, and training time under different sample sizes. The average PSNR and LPIPS are improved as we sample more voxels. However, the improvement becomes marginal as the sample size reaches 128K. In addition, as shown in Figure 13, the quality of rendering results benefits from the larger sample size. However, once N reaches 256K and 512K, the performance degrades since CoordNet begins to overfit the training data. Therefore, we suggest that the sample size should be 128K.

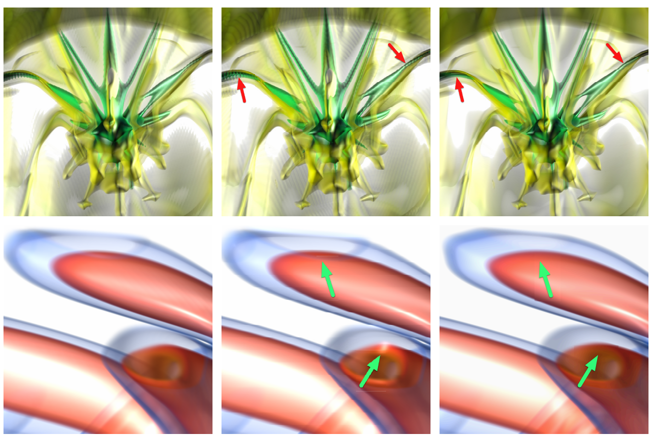
2.2 Number of Initial Neurons (m)

We optimize CoordNet using different numbers of m for the SSR task to determine an appropriate number of initial



(a) LERP (b) SloMo (c) TSR-TVD (d) CoordNet (e) GT

Fig. 6: Slice of volume rendering results for the TSR task with an interpolation interval of 3. Top to bottom: combustion (MF), half-cylinder (640,V), ionization (H2), and Tangaroa (V). The displayed time steps are 95, 75, 75, and 147, respectively.



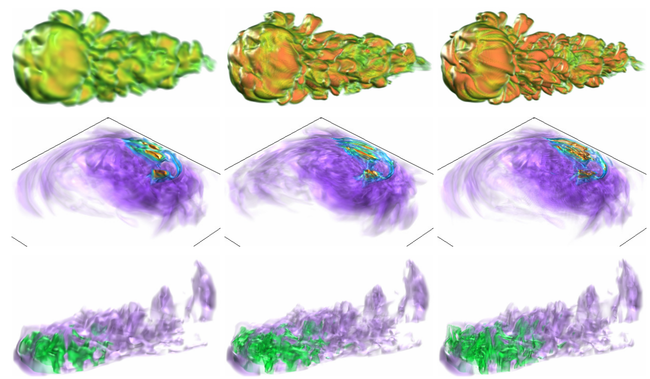
(a) input (b) BI (c) CoordNet

Fig. 7: Zoom-in volume rendering results for the SSR task. Top: ionization (PD). Bottom: vortex. For the ionization (PD) data set, we generate $2400 \times 992 \times 992$ volumes from $600 \times 248 \times 248$ ones. For the vortex data set, we generate $512 \times 512 \times 512$ volumes from $128 \times 128 \times 128$ ones.

neurons. Table 3 reports the average PSNR, LPIPS, training time, and model size under different numbers of m. In general, the average PSNR and LPIPS can be improved if a larger number of neurons is set. However, it takes longer to train, and more parameters need to be saved. Moreover, as shown in Figure 11, the quality of the rendering result is the best with 64 initial neurons. Beyond that, CoordNet could jump into overfitting, which decreases the performance. Therefore, we suggest that the number of initial neurons should be 64.

2.3 Choice of Network Depth (d)

To choose an appropriate network depth, we apply different d to train CoordNet for VS task under 512 image resolution. As displayed in Figure 12, we can observe that as d increases, the result can be improved. However, there is no significant difference between $d = 10$ and $d = 15$. In Table 4, we report the average PSNR, LPIPS, and model size under different d. The quantitative metrics are better as d gets larger. However, the increment is small when d changes from 10 to 15. Thus, we choose the network depth as 10 for CoordNet.



(a) BI (b) CoordNet (c) GT

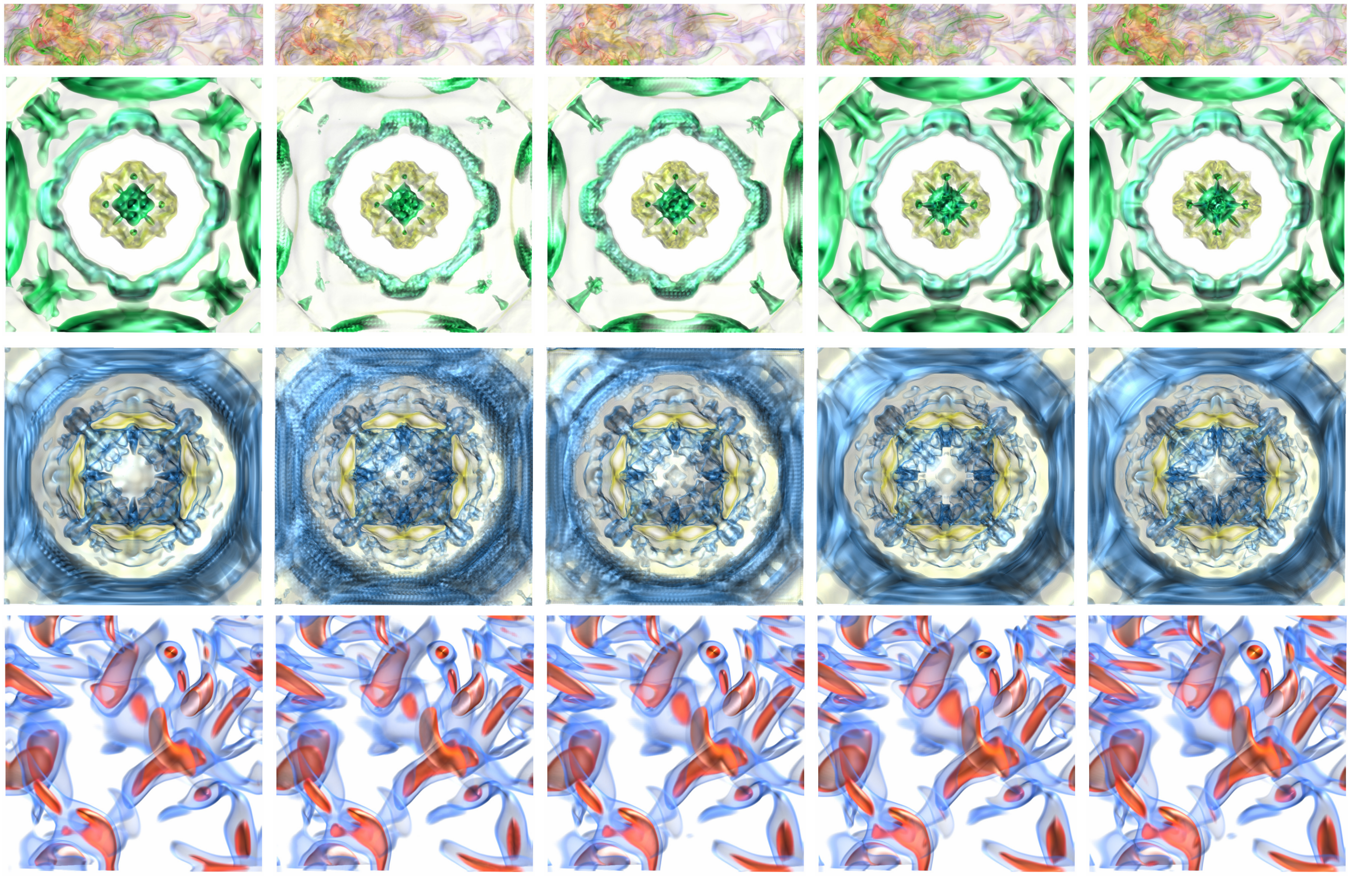
Fig. 8: Volume rendering results for the SSR task with an upscaling factor of 4×. Top to bottom: argon bubble, earthquake, and Tangaroa (VM).

TABLE 2: Average PSNR (dB), LPIPS values, and training time per epoch (in second) using the vortex data set under different numbers of sampled coordinates for the TSR task.

#coordinates	PSNR ↑	LPIPS ↓	train
32K	31.87	0.143	9.77
64K	35.56	0.103	20.41
128K	38.92	0.066	40.53
256K	39.68	0.058	90.55
512K	40.75	0.051	202.69

References

[1] M. Berger, J. Li, and J. A. Levine. A generative model for volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 25(4):1636–1650, 2019.



(a) BI (b) ESPCN (c) SSR-TVD (d) CoordNet (e) GT

Fig. 9: Slice of volume rendering results for the SSR task with an upscaling factor of $4\times$. Top to bottom: combustion (HR), ionization (PD), ionization (T), and vortex.

TABLE 3: Average PSNR (dB), LPIPS values, training time per epoch (in second), and model size (MB) using the half-cylinder (320, VM) data set under different numbers of initial neurons for the SSR task.

# neurons	PSNR ↑	LPIPS ↓	train	model
16	36.96	0.034	182.96	0.43
32	41.29	0.017	186.75	2.16
64	43.43	0.012	214.82	5.68
128	43.61	0.021	290.91	24.21

TABLE 4: Average PSNR (dB), LPIPS values, and model size (MB) using the combustion (CHI) data set under different network depths for the VS task. The training times across different depths are similar and therefore not reported here.

depth	PSNR ↑	LPIPS ↓	model
0	24.05	0.290	0.69
5	25.25	0.189	3.19
10	25.89	0.142	5.68
15	26.21	0.126	8.22

[2] W. He, J. Wang, H. Guo, K.-C. Wang, H.-W. Shen, M. Raj, Y. S. G. Nashed, and T. Peterka. InSituNet: Deep image synthesis for parameter space exploration of ensemble simulations. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):23–33, 2020.

[3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1125–1134, 2017.

[4] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 4681–4690, 2017.

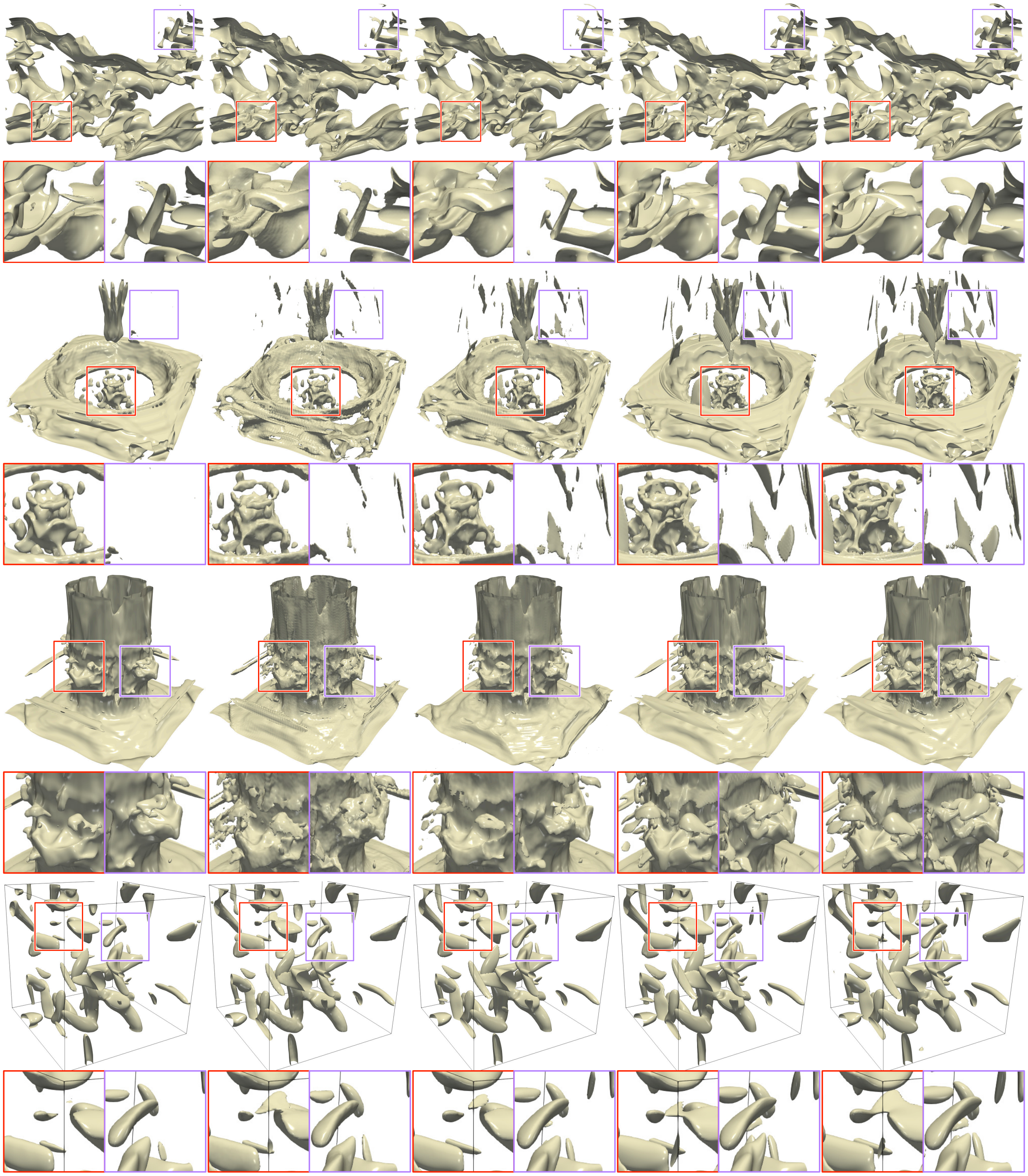
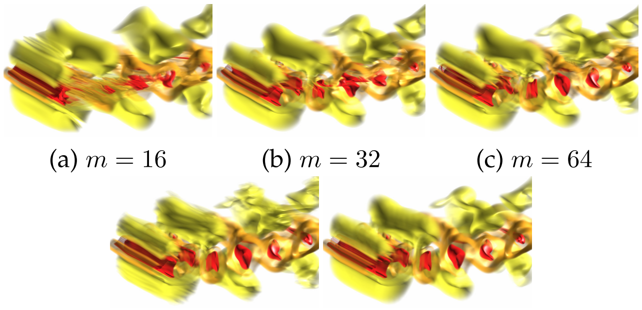


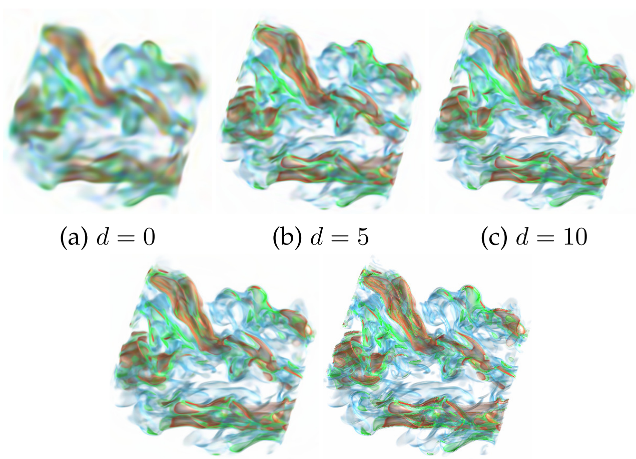
Fig. 10: Isosurface rendering results for the SSR task with an upscaling factor of $4\times$. Top to bottom: combustion (HR), ionization (PD), ionization (T), and vortex. The chosen isovalues are 0.4, -0.4, -0.3, and -0.1, respectively.



(a) $m = 16$ (b) $m = 32$ (c) $m = 64$

(d) $m = 128$ (e) GT

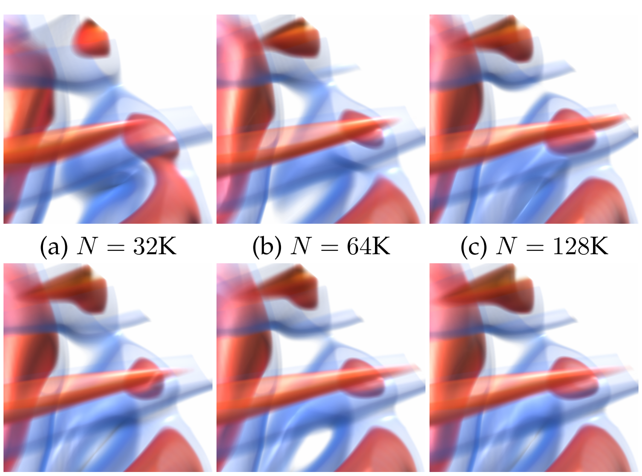
Fig. 11: Zoom-in volume rendering results for the SSR task under different numbers of initial neurons using the half-cylinder (320, VM) data set.



(a) $d = 0$ (b) $d = 5$ (c) $d = 10$

(d) $d = 15$ (e) GT

Fig. 12: Volume rendering results for the VS task under different network depths using the combustion (CHI) data set. The image resolution is 512.



(a) $N = 32$K (b) $N = 64$K (c) $N = 128$K

(d) $N = 256$K (d) $N = 512$K (f) GT

Fig. 13: Zoom-in volume rendering results for the TSR task under different sample sizes using the vortex data set.