VDL-Surrogate: A View-Dependent Latent-based Model for Parameter Space Exploration of Ensemble Simulations

Neng Shi, Jiayi Xu, Haoyu Li, Hanqi Guo, *Member, IEEE,* Jonathan Woodring, and Han-Wei Shen, *Member, IEEE*

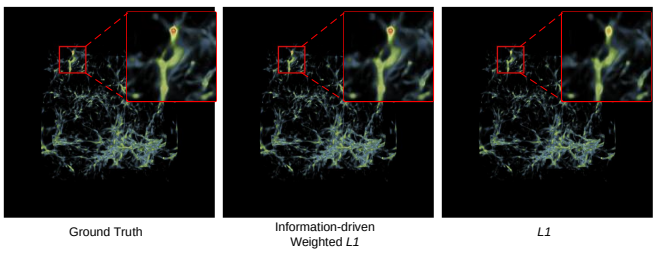

Fig. 1. Evaluation of different loss functions for the RAE training on the Nyx dataset.

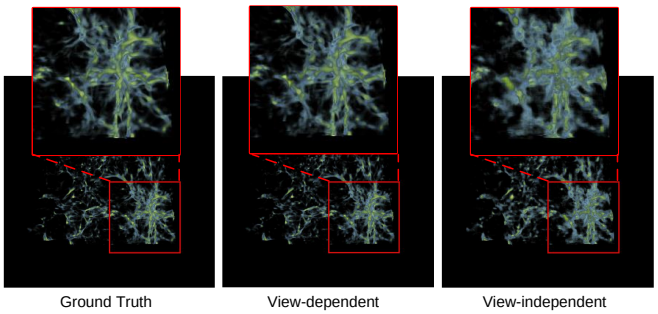

Fig. 2. Comparison of the images generated with VDL-Surrogate and the baseline View-Independent Surrogate with the ground truth images.

1 Model Evaluation by Additional Ablation Study

In this section, we show the results of ablation studies and evaluate our model at data and image levels. There are three ablation studies we perform: RAE loss functions (Section 1.1), the use of view-dependent or view-independent approaches (Section 1.2), and the use of latent representations or raw data (Section 1.3). Note that we provide the quantitative evaluation results on training and testing datasets. The similar quantitative evaluation results for the training and testing dataset indicate that all the methods are not overfitting the training dataset and ensure fair comparisons between VDL-Surrogate and alternative methods.

1.1 RAE Loss Functions

In Section 4 of the main text, we explain why we use an information-driven weighted $\mathscr{L}_1$ loss to guide the RAE training. The data values may be imbalanced, causing an RAE trained with the $\mathscr{L}_1$ loss function to ignore these few but important samples. In the experiment, we trained two RAEs, one with the weighted $\mathscr{L}_1$ loss and one with the regular $\mathscr{L}_1$ loss, on the Nyx dataset. The reconstruction results are provided in Figure 1. Red regions are with high dark matter density. After a 200-timestep simulation, these regions are rare but important for cosmologists' analysis. The RAE trained with the weighted $\mathscr{L}_1$ loss (middle) succeeded in reconstructing these high-density values. However, the result from the RAE trained with the $\mathscr{L}_1$ loss does not faithfully display the dark matter convergence. Although the error does not influence global accuracy much, it can mislead cosmologists.

- *Neng Shi, Jiayi Xu, Haoyu Li, and Han-Wei Shen are with the Department of Computer Science and Engineering, The Ohio State University, Columbus, OH, 43210, USA.*
  E-mail: {shi.1337, xu.2205, li.8460, shen.94}@osu.edu
- *Hanqi Guo is with the Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL 60439, USA.*
  E-mail: hguo@anl.gov
- *Jonathan Woodring is with the Applied Computer Science Group (CCS-7), Los Alamos National Laboratory, Los Alamos, NM 87544.*
  Email: woodring@lanl.gov

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

Table 1. Data-level quantitative comparison of the output predicted with VDL-Surrogate for three directions and the baseline VI-Surrogate for the Nyx dataset.

		VDL*X*	VDL*Y*	VDL*Z*	VI
test	PSNR (dB)	**36.38**	**34.36**	**35.59**	29.55
	MD	**0.2235**	**0.1869**	**0.2429**	0.4616
train	PSNR (dB)	**35.92**	**33.93**	**35.21**	29.44
	MD	**0.2237**	**0.1877**	**0.2466**	0.4649

1.2 View-Dependent Versus View-Independent Strategy

This section compares our viewpoint-dependent approach with three viewpoint-independent methods with or without using latent representations. One viewpoint-independent method is to train a deep neural network surrogate model to learn the mapping between input simulation parameters to simulation outputs. We call the surrogate model **V**iew-**I**ndependent **Surrogate** (VI-Surrogate). For a fair comparison, the GPU memory usage for VI-Surrogate was set to 48GB, the sum of memory used to train all VDL-Surrogate sub-networks, and the training time was similar for VDL-Surrogate and VI-Surrogate. The data-level quantitative results of VDL-Surrogate and VI-Surrogate for the Nyx dataset are reported in Table 1, where we can see that VI-Surrogate gives a lower PSNR and a higher MD than VDL-Surrogate. Table 2 provides image-level quantitative results of VDL-Surrogate and VI-Surrogate, which indicates that VI-Surrogate has lower SSIMs than VDL-Surrogate. Figure 2 compares the images produced by VDL-Surrogate and VI-Surrogate with the ground truth. We find that the image generated by VI-Surrogate is blocky. VI-Surrogate fails to gen-

Table 2. Quantitative comparison of images generated with VDL-Surrogate and VI-Surrogate for the Nyx dataset.

			VDLSurro	VISurro
test	TF1	SSIM	**0.9291**	0.7973
		EMD	0.0068	**0.0050**
	TF2	SSIM	**0.9634**	0.8922
		EMD	**0.0012**	0.0044
train	TF1	SSIM	**0.9197**	0.7983
		EMD	0.0086	**0.0054**
	TF2	SSIM	**0.9565**	0.8907
		EMD	**0.0017**	0.0043

erate high-quality results because it tries to predict high-resolution simulation outputs directly. Although the GPU memory constraint was 48GB, we still could not afford to use a powerful VI-Surrogate ($k_v = 3$ compared with $k_v = 64$ in the VDL-Surrogate). Because MPAS-Ocean has a higher simulation output resolution than Nyx, training a VI-Surrogate with $k_v = 1$ is not affordable with 48GB GPU memory.

The second solution is to divide the target volume into slightly overlapping blocks and train one predictive model for each block to address the constraint of having limited GPU memory. We divide the volume into eight blocks, and the block dimension is $9/16$ of the volume dimension. We call the surrogate model **Block-Surrogate**. For a fair comparison, the GPU memory usage for every Block-Surrogate subnetwork was set to 16GB, the memory used to train one VDL-Surrogate sub-network. Because Block-Surrogate has eight sub-networks, the training time for one subnetwork was set to $3/8$ of the VDL-Surrogate training time. Table 3 and Table 4 provide the data-level quantitative results of VDL-Surrogate and Block-Surrogate for the Nyx and MPAS-Ocean datasets, respectively. The image-level quantitative results are shown in Table 5. The tables illustrate that quantitatively, Block-Surrogate works well for the $512 \times 512 \times 512$ Nyx data but not for the $1536 \times 768 \times 768$ MPAS-Ocean data. Moreover, the block-wise strategy does not guarantee continuity between neighboring blocks. Boundary effects still exist even if the blocks overlap, as shown in Figure 3.

Another solution to resolve the GPU memory issue is to use compact latent representations to supervise the predictor training. In the model training stage, first, we encode the simulation output to a latent representation. Second, a predictor is trained to learn the mapping from input simulation parameters to the corresponding encoded latent output. In the model inference stage, we feed the new simulation input parameter into the trained predictor to predict the output latent representation, then decode the latent representation to data space by the trained decoder. Finally, the predicted data can be visualized.

The difference between the view-independent and our view-dependent methods is that the view-independent autoencoder takes a 3D volume as input instead of ray samples. However, it is hard to feed the entire volume into GPUs both in the model training and inference stages because the GPU memory cannot hold the intermediate feature maps. A solution is to divide the volume into blocks and feed a block into the autoencoder each time. Therefore, we name the autoencoder **B**lock **A**uto**E**ncoder (BAE). In the BAE training stage, we randomly cropped a volume as the input data to update the network parameters after each iteration. In the BAE inference stage, we evenly cropped the input volume, and the latent representations of all the blocks are tiled together to form the latent representation for the input volume. Like the VDL-Predictor in the VDL-Surrogate, another predictor should be trained to learn the mapping from simulation parameters to BAE-encoded-latent representations to form a complete surrogate model together with BAE. Here, however, we assume the predictor can be perfectly accurate and compare the quality of the output reconstructed by BAE and predicted by VDL-Surrogate.

In the experiment, for a fair comparison, the GPU memory constraint was 48GB, the sum of memory used for training the three RAEs; the network size of BAE was set to the sum of all the three RAEs; the BAE-generated latent representation size was three times the RAE-generated latent representation size, and the training time of BAE and RAE was similar. We list the data-level quantitative results for Nyx and MPAS-Ocean datasets in Table 6 and Table 7, respectively. Compared with VDL-Surrogate for three viewing directions, the simulation output reconstructed by BAE has lower PSNR and higher MD to the ground truth. We also provide the image-level quantitative results for Nyx and MPAS-Ocean datasets in Table 8. Compared with VDL-Surrogate, the final visualization images generated by BAE have lower SSIMs and higher EMDs to the ground truth. Moreover, as shown in Figure 4, there are noticeable boundary effects on reconstructed simulation outputs because of the discontinuity between blocks.

Table 3. Data-level quantitative comparison of the output predicted with VDL-Surrogate for three directions and reconstructed with Block-Surrogate (BSurro) for the Nyx dataset.

		VDLX	VDLY	VDLZ	BSurro
test	PSNR (dB)	36.38	34.36	35.59	**37.94**
	MD	**0.2235**	**0.1869**	**0.2429**	0.2882
train	PSNR (dB)	35.92	33.93	35.21	**37.82**
	MD	**0.2237**	**0.1877**	**0.2466**	0.2961

Table 4. Data-level quantitative comparison of the output predicted with VDL-Surrogate for three directions and reconstructed with Block-Surrogate (BSurro) for the MPAS-Ocean dataset.

		VDLΘ	VDLΦ	VDLD	BSurro
test	PSNR (dB)	**42.14**	**43.20**	**41.04**	36.46
	MD	**0.5044**	**0.5008**	0.6288	0.6081
train	PSNR (dB)	**44.37**	**45.30**	**42.45**	36.77
	MD	**0.5067**	**0.5058**	0.6345	0.6060

Table 5. Quantitative comparison of images generated with VDL-Surrogate and Block-Surrogate (BSurro).

				VDLSurro	BSurro
test	Nyx	TF1	SSIM	0.9291	**0.9415**
			EMD	0.0068	**0.0016**
		TF2	SSIM	0.9634	**0.9829**
			EMD	0.0012	**0.0008**
	MPAS-Ocean	TF1	SSIM	**0.9888**	0.9773
			EMD	**0.0021**	0.0050
		TF2	SSIM	**0.9126**	0.8837
			EMD	**0.0012**	0.0016
train	Nyx	TF1	SSIM	0.9197	**0.9426**
			EMD	0.0086	**0.0017**
		TF2	SSIM	0.9565	**0.9834**
			EMD	0.0017	**0.0009**
	MPAS-Ocean	TF1	SSIM	**0.9895**	0.9777
			EMD	**0.0020**	0.0049
		TF2	SSIM	**0.9234**	0.8891
			EMD	**0.0011**	0.0016

Table 6. Data-level quantitative comparison of the output predicted with VDL-Surrogate for three directions and reconstructed with Block AutoEncoder (BAE) for the Nyx dataset.

		VDLX	VDLY	VDLZ	BAE
test	PSNR (dB)	**36.38**	**34.36**	**35.59**	30.94
	MD	**0.2235**	**0.1869**	**0.2429**	0.3524
train	PSNR (dB)	**35.92**	**33.93**	**35.21**	30.77
	MD	**0.2237**	**0.1877**	**0.2466**	0.3616

Table 7. Data-level quantitative comparison of the output predicted with VDL-Surrogate for three directions and reconstructed with Block AutoEncoder (BAE) for the MPAS-Ocean dataset.

		VDLΘ	VDLΦ	VDLD	BAE
test	PSNR (dB)	**42.14**	**43.20**	**41.04**	26.27
	MD	**0.5044**	**0.5008**	**0.6288**	0.7517
train	PSNR (dB)	**44.37**	**45.30**	**42.45**	26.29
	MD	**0.5067**	**0.5058**	**0.6345**	0.7511

Table 8. Quantitative comparison of images generated with VDL-Surrogate and Block AutoEncoder (BAE).

				VDLSurro	BAE
test	Nyx	TF1	SSIM	**0.9291**	0.8114
			EMD	**0.0068**	0.0153
		TF2	SSIM	**0.9634**	0.7628
			EMD	**0.0012**	0.0122
	MPAS-Ocean	TF1	SSIM	**0.9888**	0.9412
			EMD	**0.0021**	0.0103
		TF2	SSIM	**0.9126**	0.7487
			EMD	**0.0012**	0.0063
train	Nyx	TF1	SSIM	**0.9197**	0.8087
			EMD	**0.0086**	0.0161
		TF2	SSIM	**0.9565**	0.7659
			EMD	**0.0017**	0.0123
	MPAS-Ocean	TF1	SSIM	**0.9895**	0.9411
			EMD	**0.0020**	0.0102
		TF2	SSIM	**0.9234**	0.7488
			EMD	**0.0011**	0.0063

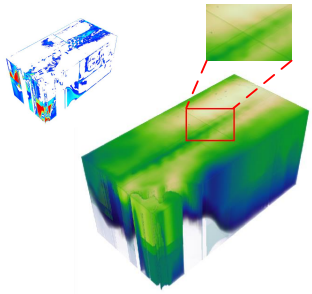

Fig. 3. An image generated from the volume reconstructed by the baseline Block-Surrogate together with difference image. The volume block-wise strategy introduces a boundary effect.

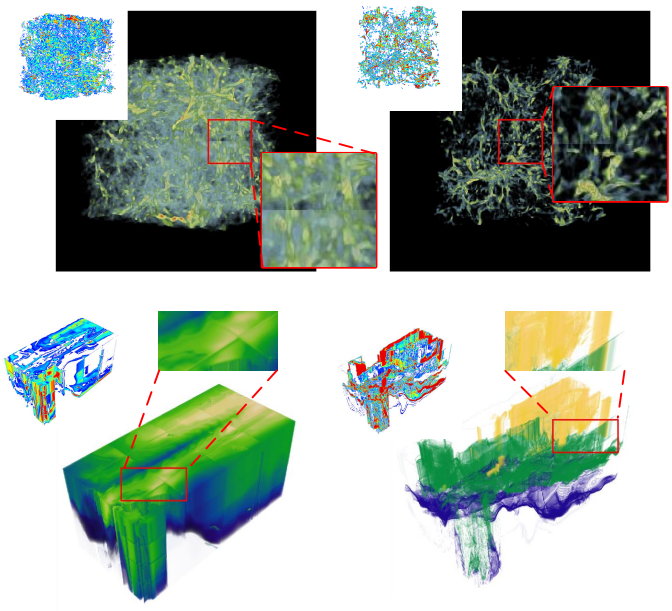

Fig. 4. Images generated from the volume reconstructed by the baseline Block AutoEncoder together with difference images. The reconstruction is low-quality, and the volume block-wise strategy introduces a boundary effect.

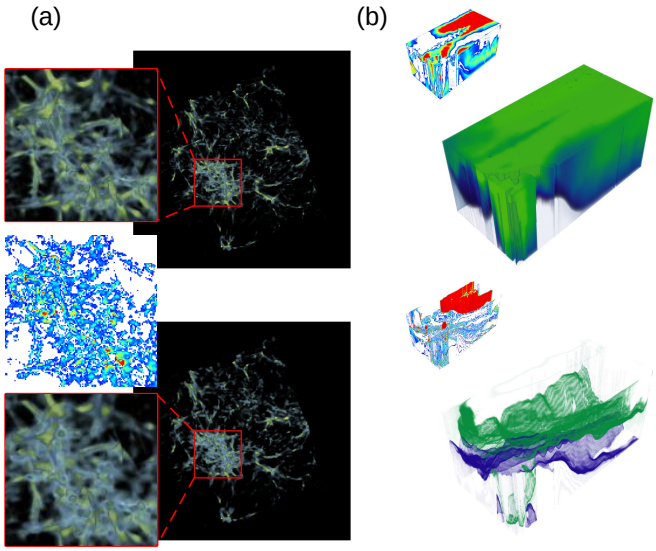

Fig. 5. Qualitative evaluation of the baseline View-Dependent Raw Data Predictor (VDR-Predictor). (a) Ground truth (top) and image generated by VDR-Predictor (bottom) for the Nyx dataset together with difference images. (b) Two images generated by VDR-Predictors and rendered with different transfer functions for the MPAS-Ocean dataset together with difference images.

1.3 Latent Versus Raw Representation

Next, we conducted experiments to study the necessity of using latent representations. Our baseline method is the same as our view-dependent latent-based approach, except that for a selected viewpoint, we only train a network to learn the mapping from simulation parameters to the viewpoint-dependent data without relying on latent representations, and is thus named **V**iew-**D**ependent **R**aw Data **Predictor** (VDR-Predictor). For a fair comparison, the GPU memory constraint for VDR-Predictor training was set to 16GB, and VDR-Predictor training time was equal to RAE training time plus VDL-Predictor training time.

We provide the data-level quantitative results for the Nyx and MPAS-Ocean datasets in Table 9 and Table 10, respectively. Our VDL-Surrogate produces results with higher PSNR and lower MD. Moreover, as shown in Figure 5, the images generated by VDR-Predictor are not qualitatively high-quality enough. Specifically, in the first column, we can see more light-colored transparent regions in the zoom-in area of the image generated with VDR-Predictors than the ground truth image. In the second column, images for the MPAS-Ocean dataset miss high-temperature regions.

2 Failure Case

The failure case for the Nyx dataset is illustrated in Figure 6. We scan all the predicted visualization images and select the one with the lowest SSIM, 0.8268, as shown in the middle of the first row. While VDL-Surrogate is able to predict the high-density region (as shown in the second row, where we use a transfer function focusing on the high-density region, the error is less evident than the one in the first row), it makes mistakes in the low-density region. The main reason is the use of information-driven weighed $\mathscr{L}_1$ loss, which lets VDL-Surrogate pay more attention to high-density regions. Then given limited network capacity, sometimes we see sacrifices in the low-density regions. Because scientists are more interested in high-density regions as one important post-hoc task is to find dark matter halos for the Nyx simulation [1], we think this sacrifice is worthwhile. In the future, we will attempt to reduce errors in the low-density region by exploring ways to increase the network capacity. For example, we can increase the GPU memory constraint and use a network with more parameters.

Table 9. Data-level quantitative comparison between VDL-Surrogate and View-Dependent Raw Data Predictor (VDR) for the Nyx dataset.

		VDLX	VDLY	VDLZ	VDRX	VDRY	VDRZ
test	PSNR (dB)	**36.19**	**34.14**	**35.36**	30.74	30.91	30.57
	MD	**0.2354**	**0.2138**	**0.2500**	0.3062	0.2878	0.3946
train	PSNR (dB)	**35.74**	**33.72**	**34.99**	30.62	30.81	30.47
	MD	**0.2345**	**0.2119**	**0.2541**	0.3085	0.2952	0.3984

Table 10. Data-level quantitative comparison between VDL-Surrogate and View-Dependent Raw Data Predictor (VDR) for the MPAS-Ocean dataset.

		VDLΘ	VDLΦ	VDLD	VDRΘ	VDRΦ	VDRD
test	PSNR (dB)	**41.92**	**43.12**	**41.34**	23.61	23.66	23.59
	MD	**0.4718**	**0.4759**	**0.6277**	0.9611	0.8302	0.9763
train	PSNR (dB)	**44.02**	**45.17**	**42.88**	23.61	23.66	23.60
	MD	**0.4791**	**0.4809**	**0.6325**	0.9614	0.8299	0.9770

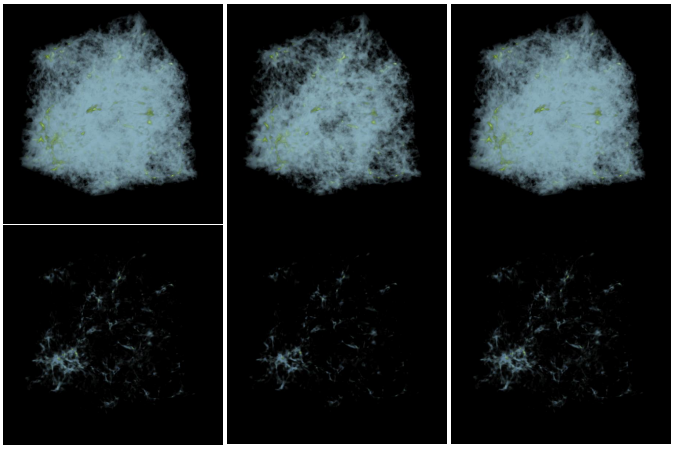

Fig. 6. A failure case for the Nyx dataset. The first row uses a transfer function to visualize the high-density and low-density regions together, and we see an obvious error at the low-density region for VDL-Surrogate. The second row uses a transfer function to look at the high-density region, and the loss for VDL-Surrogate is much less evident than the one in the first row.

REFERENCES

[1] B. Friesen, A. Almgren, Z. Lukić, G. Weber, D. Morozov, V. Beckner, and M. Day. In Situ and In-Transit Analysis of Cosmological Simulations. *Computational Astrophysics and Cosmology*, 3(1):1–18, 2016.