Studying the Impact of Latent Representations in Implicit Neural Networks for Scientific Continuous Field Reconstruction

Wei Xu[1], Derek Freeman DeSantis[2], Xihaier Luo[1], Avish Parmar[1], Klaus Tan[3],
Balu Nadiga[2], Yihui Ren[1], Shinjae Yoo[1]

[1]Brookhaven National Laboratory
[2]Los Alamos National Laboratory
[3]Not Affiliated

xuw@bnl.gov, ddesantis@lanl.gov, xluo@bnl.gov, aparmar@bnl.gov, halfanteater87@gmail.com, balu@lanl.gov,
yren@bnl.gov, sjyoo@bnl.gov

Abstract

Learning a continuous and reliable representation of physical fields from sparse sampling is challenging and it affects diverse scientific disciplines. In a recent work, we present a novel model called MMGN (Multiplicative and Modulated Gabor Network) with implicit neural networks. In this work, we design additional studies leveraging explainability methods to complement the previous experiments and further enhance the understanding of latent representations generated by the model. The adopted methods are general enough to be leveraged for any latent space inspection. Preliminary results demonstrate the contextual information incorporated in the latent representations and their impact on the model performance. As a work in progress, we will continue to verify our findings and develop novel explainability approaches.

Introduction

Field reconstruction is the process involving the recreation of a physical field denoted as u, depicting the spatial and temporal distribution of a specific parameter (e.g., temperature, velocity, or displacement). Its significance spans diverse scientific disciplines such as geophysics, climate science, and fluid mechanics, due to several key functions. First, it facilitates the extrapolation of values at unmeasured locations by leveraging available data (Kutz et al. 2016). It also contributes to the identification of patterns, trends, and variations within the investigated parameters (Shu, Li, and Farimani 2023; Wang et al. 2022). Lastly, field reconstruction plays a crucial role in optimizing sensor placement and refining data collection strategies, thereby enhancing the effectiveness of experimental designs (Fukami et al. 2021; Krause, Singh, and Guestrin 2008).

Recent progress in machine learning has sparked a growing interest in addressing field reconstruction challenges through implicit neural representations (Xie et al. 2022; Mildenhall et al. 2021). These representations harness multilayer perceptrons (MLPs) to represent images, videos, and 3D objects. Specifically, coordinate-based MLPs operate by taking low-dimensional coordinates, such as spatial position (x) or time index (t), as input and predicting the value of a learned signal (u) for each coordinate point. The coordinate-based MLP effectively learns an implicit, continuous, and differentiable function, mapping input coordinates to the target signal $u = F_\theta(x, t)$.

Driven by the quest for an enhanced representation method, we introduce a context-aware indexing mechanism. Specifically, we integrate existing contextual information into the temporal coordinate (t). Given the dynamic changes in the quantity and locations of available measurements over time t, we propose a framework where an encoder extracts a latent representation from real-time measurements. This latent representation is then utilized to steer the model towards the desired time instance. Paired with an implicit neural representation-based decoder, our proposed method excels in the seamless reconstruction of continuous fields (Luo et al. 2024).

Meanwhile, as neural networks become more complicated, it is a grand challenge to increase the overall interpretability and trust especially when adopted by scientific users. For our field reconstruction model, the learned latent representation is an essential component and incorporates information about the measurements beyond temporal indexing. In this work, we present the application of several dimension reduction approaches and use them as explainability methods for latent representation understanding. Concretely, we leverage explainability tools to demonstrate that the latent representation has sufficiently encoded the real-time measurements. We take the use case of continuous field reconstruction using implicit neural networks for climate science and report the findings as a work in progress.

Related Work

Explainable Artificial Intelligence (XAI) technique has flourished in the past decade. For the latent space understanding, since the latent vectors are difficult to interpret by humans directly, many methods are proposed to tackle the dimensionality challenges. Embedding techniques e.g., t-Distributed Stochastic Neighbor Embedding (t-SNE) (van der Maaten and Hinton 2008) are used to project high-dimensional data into 2D space and study whether the internal correlations of the data is preserved in the latent space. Clustering is another common method to apply to the latent vectors and demonstrate the relationship. Moreover, both linear and nonlinear methods are used to estimate the sim-

ilarity between latent representations (Mehrer et al. 2020). Developing visual analytics frameworks is another branch of research to facilitate user-directed analysis exploring the latent space interactively (Shen, Wang, and Shen 2020). We find that existing works for dimensionality and decomposition studies can be beneficiary for general-purpose latent understanding, which are less commonly adopted beyond t-SNE visualization and ablation studies.

Methodology

The MMGN model

Traditional implicit neural networks incorporate the time index along with spatial coordinates. Here, we present a neural network reconstruction method called MMGN (Multiplicative and Modulated Gabor Network) (Luo et al. 2024), characterized by an encoder-decoder architecture. For the encoder, we propose to utilize measurements, represented as $u_t^1, u_t^2, u_t^3,...$ at time t, for implicit model guidance instead of relying on t for explicit pointing. Therefore, we construct an encoder $E(\cdot)$ to convert the observed measurements into a latent code z_t. Subsequently, we employ a decoder $D(\cdot)$ that utilizes both the encoded information and spatial coordinates x to reconstruct the underlying physical field.

$$\text{Baseline}: u(x, t) = F(x, t)$$
$$\text{Proposed}: u(x, t) = D(x, z_t) = D(x, E(u_t^1, u_t^2, \dots))$$
$$(1)$$

• **Encoder**: Autoencoders (AE) and their probabilistic variant, variational autoencoders (VAE) (Kingma and Welling 2014), are frequently employed for representation learning, leveraging their inherent latent variable structures. However, conventional AE and VAE encounter challenges when dealing with the variability of sensor locations and their fluctuations over time. While their graph counterparts can manage spatial variability, and modified versions address dynamic graphs, they become computationally demanding for extensive graphs and encounter difficulties with long-range dependencies. In contrast, auto-decoder, as exemplified in Park et al. (2019), demonstrates reduced underfitting and heightened flexibility. It accommodates observation grids of varying forms, including irregular or those existing on a manifold, without necessitating a specialized encoder architecture, provided the decoder shares the same property. Therefore, the auto-decoder is chosen as the backbone for the encoder due to its ability to accommodate free-formed observations, considering that the number and positions of available observations vary over time.

• **Decoder**: The decoder inputs comprise two components: spatial coordinates (x) and latent codes (z). Initially, processing x through fully-connected feed-forward layers results in a coordinate-based MLP. Although such an MLP based on coordinates can provide a continuous representation, it encounters challenges in learning high-frequency signals, a phenomenon referred to as spectral bias. Recent studies suggest that addressing this issue involves incorporating positional encoding with Fourier features (Tancik et al. 2020) or introducing periodic non-linearities in the first hidden layer (Sitzmann et al. 2020). In our approach, we employ Gabor filters $g(\cdot)$ over Fourier bases to transform the coordinates. Following the transformation of coordinates x, we introduce a modulation step where the transformed coordinates $g(x)$ undergo modulation through a multiplicative layer, facilitating the integration of $g(x)$ and z. Specifically, the decoder incorporates a multiplicative filter network (MFN) (Fathony et al. 2020) as its backbone network. This choice is motivated by the recursive mechanisms embedded in MFN, enabling improved fusion of spatial coordinates (x) and latent codes (z).

• **Trained models**: The MMGN model used in this work is trained with 5% sampling rate of the original dataset generated by the CESM2 climate model (Danabasoglu et al. 2020) that simulates Earth's climate states. We take monthly averaged global surface temperature data for model testing. That said, the network can be trained using a sampling rate at 5% of the training data to recover a continuous representation of it without substantial errors. Moreover, by altering the hyperparameter k defining the size of latent codes z_k, we can train different MMGN models. In our previous study, we train 10 models with latent sizes, ranging from 1 to 512 by doubling the latent size at a time, and collect the corresponding learned latent codes for all T time steps. We found that increasing the latent size can reduce the reconstruction error and the improvement becomes more subtle when the latent size is moderate.

Explainability approaches

We adopt several explainability methods to conduct additional studies that further understand the impact of latent codes on the model performance and illustrate the contextual information latent codes integrate.

• **Embedding and Clustering** Latent representations are spatiotemporal vectors in high-dimensional space. To compare the representations in different latent spaces, we first adopt embedding techniques to project data into a 2D plane. For the demonstration, we use t-SNE to generate the distribution of T latent vectors of each latent space and use K-means clustering to partition them into groups. Then, we visualize the distributions and quantify the statistics of clusters for all the latent spaces so that the variations can be observed. This helps to explain the impact of latent size on the acquired contextual information by the encoder $E(\cdot)$.

• **Correlation Analysis** Latent codes can also be represented by a $T \times k$ matrix where k-dimensional latent vectors are stacked in raster order. Comparing the latent spaces becomes the comparison of the corresponding matrices so that the rows represent samples and the columns represent features. In that sense, we can use correlation analysis to generate patterns of different latent spaces and compare their correlations. We leverage both Principal Component Analysis (PCA) (Hotelling 1936) and Canonical Correlation Analysis (CCA) (Hä 2007) in this study.

PCA identifies the direction i.e., the principal component, formed by a linear combination of original features that captures the maximal variance of the data. The explained variance ratio (i.e., normalized eigenvalue) is a measurement
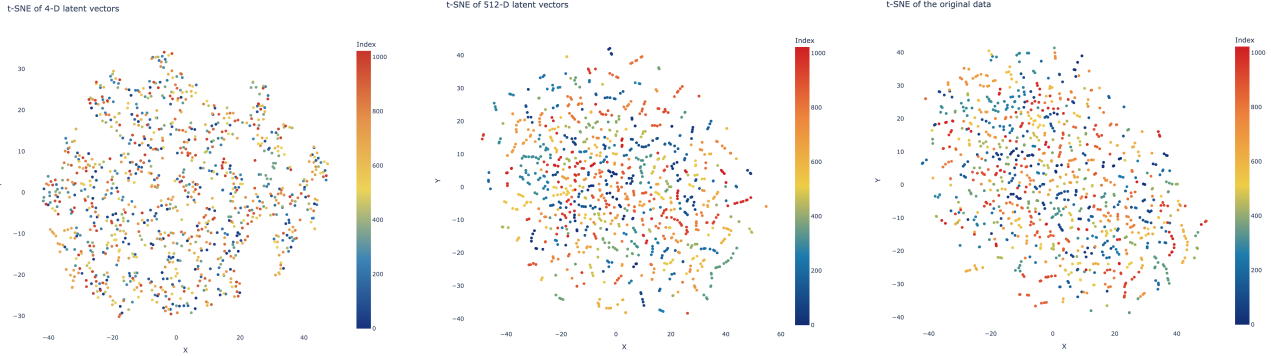
Figure 1: The t-SNE distribution changes in three latent spaces: (left) 4-D, (middle) 512-D, and (right) the original data; the color represents the temporal indexing.

used to explain the proportion of variance of the matrix explained by the corresponding principal component. Thus, we generate k principal components with their corresponding explained variance ratios for the matrix with k-dimensional latents and compare the trends. Meanwhile, we also apply PCA to the original data and generate 512 explained variance ratios. By visualizing the trends of ratios, we could evaluate the coherence of the rank of information preserved in latent spaces and the original data.

CCA provides a more direct way to identify and measure the associations between two sets of features, here two matrices. It determines pairs of canonical variates u from the first set and v from the second set so that they are maximally correlated with each other. The difference between PCA and CCA roots in that while PCA finds the linear combination of features that maximizes variance in the data, CCA looks for a linear combination of features in both data that have the strongest correlation between each other.

• **Tensor Factorizations** Tensor factorizations provide a method for effectively explaining surrogate model complexity. Tensor factorizations are techniques for decomposing multidimensional arrays (tensors) into simpler, constituent components. Among various tensor factorization techniques, Tucker Decomposition plays a pivotal role in analyzing and interpreting multiway data. This method decomposes a tensor into a core tensor multiplied by factor matrices along each dimension of the tensor. The factor matrices describe the dominant modes along each variable, while the core tensor describes their interactions or mixing. Both the data used, and output derived from field reconstruction, intrinsically arises as a tensor. The Tucker decomposition is a natural tool to reveal the dominant modes and interactions within these data sets. See (Kolda and Bader 2009) for more details.

We leverage the Tucker Decomposition to assess how well MMGN represents the dominant modes (i.e., $\boldsymbol{x}$ and t) and their mixing. By taking the latent size at 256 for this study, the Tucker decomposition of both the training set and the model output are computed. The factor matrices and the mixing tensor are then compared to see how well MMGN

does at reproducing the modes and mixing complexity of the Earth system data.

To compare the modes, we compute the Pearson correlation coefficient between the modes for each variable. A high matching across modes in a variable indicates that MMGN has learned the data generating process for that particular variable. To compare the cores, we compute the Tucker decomposition for cores shaped $[r, r, r]$ for each $r = 1, 2, \ldots 200$. The values of the core are then normalized, and the entropy of the core is computed. The entropy measures the complexity of the mixing at these different multiranks. One should expect that the entropy of the MMGN model to be less than the entropy of the training data for each r, with exact matching indicating it has captured the mixing process between the variables.

• **Ablation Study** Finally, an ablation study is designed to complement the previous effort inspecting the contribution of each dimension of the latent vector. By ablating one dimension at a time during the inference, we observed a reduced performance but the mean squared error (MSE) keeps mitigating when a bigger latent size is adopted. In this study, we want to decompose the overall error into the temporal and spatial counterparts and check if any resonance exists.

Results

• **Embedding and Clustering** The t-SNE visualizations are shown in Figure 1 to illustrate how the distributions of latents in different latent spaces vary. Only the 4-D, 512-D, and the original spaces are shown for parsimony. Our goal is to understand the impact of the latent size on the learned representations. We observe that the lower dimensional space has more diverse but compact clusters and it gets more evenly spreading out when the dimension increases. We also find more string-shaped local clusters having the same color in the higher dimensional space. Since we use the temporal indexing t as colors, these clusters are formed by the temporally adjacent latent vectors. The same pattern and distribution are observed in the original data. We conclude that the higher latent space better captures the global distribution of the original data, and to some extent, also maintains lo-
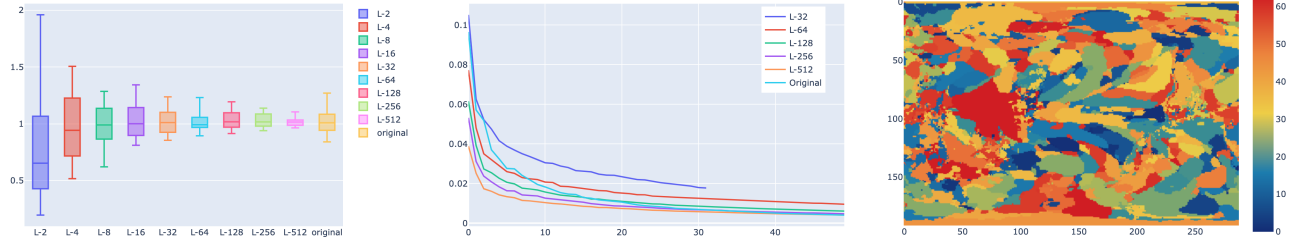
Figure 2: Left: The standard deviations of clusters diminish across latent spaces and they are compared with the ones of the original data (the rightmost boxplot). Middle: The trendings of explained variance ratios of the principal components for all latent spaces (illustrating only the top 5 sizes) and compare them with the trending of the original data. Right: Ablation result indicating spatial linkage of temporal indexing.
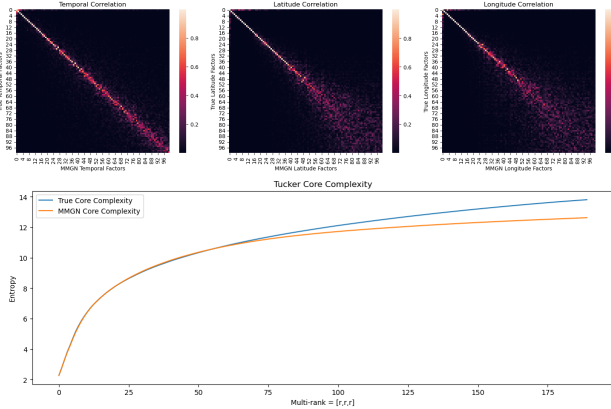

Figure 3: Top: Correlation plots for the three variables. Bottom: Complexity of Tucker core for different multi-ranks.

cal coherence (string-shaped small clusters). Note here the use of t-SNE projections of 2-D latents instead of their original coordinates is meant for a fair comparison with other latent spaces. Then, the clustering results are generated and we collect the standard deviations $\sigma_1, ..., \sigma_n$ of the vectors in the corresponding n clusters for all the latent spaces. As a preliminary result, the increased σ pattern also confirms that the distribution shifts from compact and diverse groups to a wider space spanning as shown in Figure 2(left).

• **Correlation Analysis** We aim to use correlation analysis to understand the impact of latent size in terms of rank complexity and potential information loss. First, the PCA results are shown in Figure 2(middle) where the explained variance ratios of higher latent spaces are compared with the one of the original data. The slope of the curve suggests the rank of data complexity. For example, a steep slope means with the first few principal components the most variances can be preserved. Compared to the slope of the original data, we observe a similar slope starting from a latent size at 64 and beyond. Furthermore, we also generate the CCA canonical variate pairs between two cases: 1) two latent spaces and 2) a latent space and the original data. Then we use Pearson correlation coefficient to gauge their relations. We find

that all the paired canonical variates are fully correlated for both cases. We conclude that any latent code is a reduced representation presenting consistent information about the original data.

• **Tensor Factorizations** Similar to the purpose of adopting correlation analysis, the results of the tensor based analysis are shown in Figure 3. For the factor matrices, we see that the MMGN model has strong agreement with the true factors for the first 20 temporal and longitudinal components, and first 40 latitudinal components. These top modes are responsible for the majority of the reconstruction of the tensor (roughly $\sim 4\%$ relative error), so agreement here indicates MMGN has learned the dominant spatial temporal patterns of the dataset. The divergence of higher order modes suggests that MMGN fails to capture the higher frequency spatial components of the data (see Figure 4). The fact that there is stronger agreement in the latitude modes indicates the model had an easier time capturing this component.

From Figure 3 we see that the core of the MMGN model has complexity that matches the true data up to multi-rank [75, 75, 75]. At this high of a rank, the tensor decomposition for both the true data and the MMGN output have a relative reconstruction error of $\sim 1\%$. The agreement of model complexity up to this large multi-rank indicates that MMGN accurately captures the mixing processes of the underlying physical phenomenon.

• **Ablation Study** We collect the sets of temporal and spatial errors for the model of a latent size at 64. A preliminary result for the spatial linkage of the latent dimension is shown in Figure 2(right). We use colors to indicate the latent dimension index (ranging between 0 and 63) so that once ablated it causes the most performance reduction at a spatial location. It is interesting to observe that individual latent dimensions contribute to adjacent regions for low-fidelity areas. This plot also shows the MMGN model can capture long-range dependencies. We will further verify this finding in future work.

Conclusion

The present studies leveraged dimensionality reduction approaches to explain the effect of the encoder latent space dimension has on the representation. These tools have im-
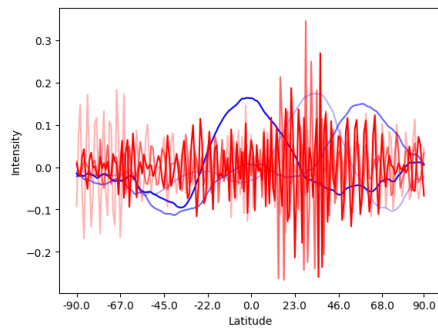
Figure 4: Latitude modes $1, 2, 3$ in blue versus latitude modes $101, 102, 103$ in red. Higher order modes capture higher frequency information compared to lower order modes.

proved understanding of the learned latent representations. As a work in progress, we will keep verifying the current findings and work on potential extensions, developing feature importance methods attributing the input features for the prediction.

Acknowledgments

References

2007. *Canonical Correlation Analysis*, 321–330. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-540-72244-1.

Danabasoglu, G.; Lamarque, J.-F.; Bacmeister, J.; Bailey, D.; DuVivier, A.; Edwards, J.; Emmons, L.; Fasullo, J.; Garcia, R.; Gettelman, A.; et al. 2020. The community earth system model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12(2): e2019MS001916.

Fathony, R.; Sahu, A. K.; Willmott, D.; and Kolter, J. Z. 2020. Multiplicative filter networks. In *International Conference on Learning Representations*.

Fukami, K.; Maulik, R.; Ramachandra, N.; Fukagata, K.; and Taira, K. 2021. Global field reconstruction from sparse sensors with Voronoi tessellation-assisted deep learning. *Nature Machine Intelligence*, 3(11): 945–951.

Hotelling, H. 1936. Relations Between Two Sets of Variates. *Biometrika*, 28(3/4): 321–377.

Kingma, D. P.; and Welling, M. 2014. Auto-Encoding Variational Bayes. In *ICLR*.

Kolda, T. G.; and Bader, B. W. 2009. Tensor decompositions and applications. *SIAM review*, 51(3): 455–500.

Krause, A.; Singh, A.; and Guestrin, C. 2008. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9(2).

Kutz, J. N.; Brunton, S. L.; Brunton, B. W.; and Proctor, J. L. 2016. *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM.

Luo, X.; Xu, W.; Nadiga, B.; Ren, Y.; and Yoo, S. 2024. Continuous Field Reconstruction from Sparse Observations with Implicit Neural Networks. In *The Twelfth International Conference on Learning Representations*.

Mehrer, J.; Spoerer, C.; Kriegeskorte, N.; and Kietzmann, T. 2020. Individual differences among deep neural network models. *Nature Communications*, 11.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106.

Park, J. J.; Florence, P.; Straub, J.; Newcombe, R.; and Lovegrove, S. 2019. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 165–174.

Shen, J.; Wang, R.; and Shen, H.-W. 2020. Visual exploration of latent space for traditional Chinese music. *Visual Informatics*, 4(2): 99–108. PacificVis 2020 Workshop on Visualization Meets AI.

Shu, D.; Li, Z.; and Farimani, A. B. 2023. A physics-informed diffusion model for high-fidelity flow field reconstruction. *Journal of Computational Physics*, 478: 111972.

Sitzmann, V.; Martel, J.; Bergman, A.; Lindell, D.; and Wetzstein, G. 2020. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33: 7462–7473.

Tancik, M.; Srinivasan, P.; Mildenhall, B.; Fridovich-Keil, S.; Raghavan, N.; Singhal, U.; Ramamoorthi, R.; Barron, J.; and Ng, R. 2020. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems*, 33: 7537–7547.

van der Maaten, L.; and Hinton, G. 2008. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*, 9(nov): 2579–2605. Pagination: 27.

Wang, S.; Zhou, T.; Lu, Y.; and Di, H. 2022. Detail-preserving transformer for light field image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 2522–2530.

Xie, Y.; Takikawa, T.; Saito, S.; Litany, O.; Yan, S.; Khan, N.; Tombari, F.; Tompkin, J.; Sitzmann, V.; and Sridhar, S. 2022. Neural fields in visual computing and beyond. In *Computer Graphics Forum*, volume 41, 641–676. Wiley Online Library.