MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Single-pixel imaging of dynamic flows using Neural ODE regularization

Sholokhov, Aleksei; Rapp, Joshua; Nabi, Saleh; Brunton, Steven; Kutz, Nathan; Mansour, Hassan

TR2024-024 March 19, 2024

Abstract

Single-pixel imaging is an efficient image acquisition process where light from a target scene is passed through a spatial light modulator and then projected onto a single photodiode with a high temporal acquisition rate. The scene reconstruction is achieved us- ing computational methods that leverage prior assumptions on the scene structure. In this paper, we propose to model the structure of a dynamic spatio-temporal scene using a reduced-order model that is learned from training data examples. Specifically, by combining single-pixel imaging methods with a reduced-order model prior implemented as a neural ordinary differential equation, image sequence reconstruction can be accomplished with significantly reduced data requirements while maintaining performance levels on par with leading methods. We demonstrate superior reconstruction at low sampling rates for simulated trajectories governed by Burgers' equation and turbulent plumes emulating gas leaks.

IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2024

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

SINGLE-PIXEL IMAGING OF DYNAMIC FLOWS USING NEURAL ODE REGULARIZATION

Aleksei Sholokhov[1], Joshua Rapp[2], Saleh Nabi[3], Steven L. Brunton[1],J. Nathan Kutz[1], Hassan Mansour[2]

[1] University of Washington, Seattle, WA, USA
[2] Mitsubishi Electric Research Laboratories, Cambridge, MA, USA
[3] Schneider Electric, Andover, MA USA

ABSTRACT

Single-pixel imaging is an efficient image acquisition process where light from a target scene is passed through a spatial light modulator and then projected onto a single photodiode with a high temporal acquisition rate. The scene reconstruction is achieved using computational methods that leverage prior assumptions on the scene structure. In this paper, we propose to model the structure of a dynamic spatio-temporal scene using a reduced-order model that is learned from training data examples. Specifically, by combining single-pixel imaging methods with a reduced-order model prior implemented as a neural ordinary differential equation, image sequence reconstruction can be accomplished with significantly reduced data requirements while maintaining performance levels on par with leading methods. We demonstrate superior reconstruction at low sampling rates for simulated trajectories governed by Burgers' equation and turbulent plumes emulating gas leaks.

Index Terms— Single-pixel imaging, high-dimensional dynamical systems, reduced-order model, neural ODE

1. INTRODUCTION

Low-cost cameras with high pixel counts have enabled digital imaging to become ubiquitous. However, traditional pixel arrays are often too slow or too expensive for imaging applications that require high acquisition speed or detection at wavelengths outside of the visible spectrum. On the other hand, imaging systems that raster scan the illumination or detection pixel-by-pixel enable image formation from a cost-effective single-pixel detector, but the scanning process may likewise be too slow to capture dynamic scenes. For example, methane gas-leak monitoring requires mid-infrared (MIR) wavelengths sensors. Unfortunately, MIR array detectors come with a significant cost, and capturing scene dynamics requires high-frame rate acquisition [1]. In such scenarios, a candidate solution is single-pixel imaging (SPI), a technique that uses a single detector element and time-varying spatial light modulation to capture multiplexed measurements of light intensity [2].

SPI acquisition is often combined with prior information on the structure of the scene content along with computational reconstruction techniques to recover a full image despite using fewer modulation patterns than the target pixel resolution. Initial approaches to single pixel imaging often assumed sparsity of the scene in the Fourier or wavelet basis or in the spatial gradient domain (i.e., total variation) [2, 3]. Recent approaches have adopted deep learning (DL) architectures for SPI reconstruction and showed that using deep neural networks can dramatically reduce the sampling ratio and offer near-real-time performance [4, 5, 6, 7]. However, these methods still require a relatively large number of single-pixel samples per frame (SPF) to ensure satisfactory reconstruction of a full video sequence.

In this work, we aim to enable high quality reconstruction of videos from a small number of samples per frame by learning priors on dynamics that can act as a regularization across the entire sequence of frames. We call our method SPI-NODE because we combine SPI acquisition with a reconstruction algorithm that uses a neural ordinary differential equation (Neural ODE, or NODE) as regularization. We first pre-train a reduced-order model (ROM) to identify a low-dimensional latent space representation of a gas flow and calculate the system's evolution within the latent space using Neural ODEs [8, 9]. Next, we efficiently solve an ODE-regularized inverse problem using the adjoint-sensitivity method to obtain the necessary derivatives following the technique proposed in [8]. The reconstruction procedure minimizes the measurement mismatch between the samples recorded by the SPI detector and the synthesized measurements obtained by sampling the reconstructed trajectory using the compressive measurement operator. We demonstrate through numerical validation using simulated and real flows that our approach is capable of reconstructing dynamic flows at a sampling rate that is approximately two orders of magnitude lower than that achieved by competitive SPI algorithms.

This paper is organized as follows, in Section 2 we describe related work in the area of single pixel imaging and reduced-order modeling of dynamical flows. Section 3 introduces the signal acquisition model along with the proposed network architecture that will allow us to train suitable ROMs, and present the recovery algorithm with our proposed regularizer. In Section 4, we evaluate our approach using one synthetic dataset and one experimental dataset and demonstrate its superiority over state of the art methods. Finally, we conclude the paper in Section 5.

2. RELATED WORK

Our work lies at the intersection of single pixel imaging and reduced-order modeling using neural ODEs. SPI evolved from early observations of the benefits of using multiplexed illumination instead of raster scanning [10] and the possibility of combining multiplexing with a single pixel detector to form an image [11]. However, SPI development was largely motivated by compressive sensing theory [12, 13] and efficient reconstruction algorithms [2]. Recent works have focused on creating more sophisticated reconstruction techniques that are agnostic to the sampling patterns by improving the efficiency of image recovery using classical [3, 14] as well as learned priors [4, 5, 7] for image recovery. In the ensuing years, many im-

This work was conducted while Aleksei Sholokhov and Saleh Nabi were with MERL.

plementations of SPI have been explored with different hardware architectures, reconstruction algorithms, and imaging applications [15, 16, 17].

A Neural ODE model [8] is a differentiable ODE solver. It utilizes the adjoint sensitivity method [18] to back-propagate the gradients through the solution. The emergence of GPU-compatible implementations of differentiable ODE solvers such as `torchdiffeq` [8] and `Diffrax` [19] caused a broad adoption of Neural ODE models in many applications, including for system identification [20], classification [21] medical imaging [22], reduced-order modelling [9, 23], and modelling air flow [24] and fluids [25], to name a few. Neural ODEs have demonstrated superior ability to model highly non-linear dynamics compared to linear models especially when the dimensionality of the space over which the dynamics evolve is small.

3. PROBLEM FORMULATION

3.1. Acquisition System

Let $\boldsymbol{x}(t) \in \mathcal{X} \subseteq \mathbb{R}^n$, $t \in [0, T]$ denote temporal snapshots of a high-dimensional spatio-temporal system indexed in time by t, e.g. the desired video recording of a scene that we wish to observe. In many applications we cannot record $\boldsymbol{x}(t)$ at full resolution in real time due, for example, to a prohibitive cost of the required hardware or lack of high-dimensional sensors. Instead, we collect measurements from p detectors, where each detector provides a linear combination of the light emitted by $\boldsymbol{x}(t)$ at time t:

$$\boldsymbol{y}(t) = \boldsymbol{A}\boldsymbol{x}(t), \quad \boldsymbol{A} \in \mathbb{R}^{p \times n}, \quad t \in [0, T]. \tag{1}$$

In particular, we consider a single pixel camera setup where for every time instance t, a vector of p acquisitions $\boldsymbol{y} \in \mathbb{R}^p$ are obtained by a high sampling rate photo-detector using the projection matrix $\boldsymbol{A}$. The rows of the matrix $\boldsymbol{A}$ correspond to a binary mask pattern that can be optically encoded, for example, using a digital micromirror device (DMD) [26]. A DMD is a type of spatial light modulator commonly used for SPI consisting of an array of individually addressable mirrors with two possible angle settings. Each micromirror corresponds to a mask element, which can be turned "on" or "off" by switching the mirror angle. Fig. 1 illustrates a single pixel imaging setup where a gas plume is imaged using a DMD array and a single-pixel MIR photo-detector.

Multiple techniques attempt to reconstruct $\boldsymbol{x}(t)$ directly by inverting equation (1) under mild regularity priors; examples include differential ghost imaging [27], total variation regularization [15], and Fourier-domain regularized inversion [28]. Alternatively, one can recover $\boldsymbol{x}(t)$ by recovering its representation $\boldsymbol{z}(t) \in \mathcal{Z} \subseteq \mathbb{R}^m$ on a low-dimensional manifold ($m \ll n$). Specifically, given an invertible mapping $\psi(\boldsymbol{z}) : \mathcal{Z} \to \mathcal{X}$, we replace the problem of solving for a vector $\boldsymbol{x}$ in high-dimensional space using p linear observations with solving for a low-dimensional vector $\boldsymbol{z}$ from p non-linear observations:

$$\boldsymbol{y}(t) = \boldsymbol{A}\psi(\boldsymbol{z}(t)). \tag{2}$$

We call the space $\mathcal{X}$ an observable space, $\mathcal{Z}$ a latent space, and the mapping $\psi(\boldsymbol{z})$ a decoder. Often a suitable $\psi(\boldsymbol{z})$ is not known in advance, in which case an approximation $\psi_\theta(\boldsymbol{z})$ is trained using a dataset of full-resolution images $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$, where θ denotes the parameters of the model, e.g. the weights of an autoencoder network. However, if $\boldsymbol{y}(t)$ is insufficient for a unique reconstruction of $\boldsymbol{x}(t)$ then it will remain insufficient for reconstructing $\boldsymbol{z}(t)$. In order to reduce the sufficient number of measurements that can uniquely reconstruct $\boldsymbol{x}(t)$, and by proxy $\boldsymbol{z}(t)$, additional structural sparsity or priors may be introduced, such as joint training of the sampling matrices and the decoder [7]. In this work we take a different route and supplement the encoder $\psi(z)$ with prior knowledge of the dynamics of $\boldsymbol{z}(t)$ on the manifold over time.

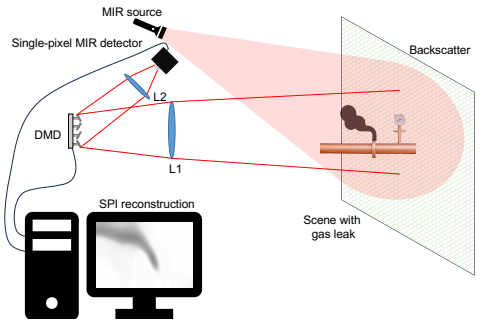


Fig. 1: Schematic of an SPI setup for gas leak monitoring. An MIR source flood-illuminates the region of interest. The light is absorbed by the gas plume but reflected from the backscatter surface. The reflected light is focused with lens L1 onto the DMD, which modulates the light by angling the micro-mirrors to reflect some of the light towards lens L2, which is focused onto the single-pixel MIR detector.

3.2. Reduced-Order Model with Non-Linear Latent Dynamics

Let $\boldsymbol{x}(t)$ be modelled as an autonomous dynamical system on a finite space $\mathcal{X} \subseteq \mathbb{R}^n$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}(t) = \boldsymbol{f}(\boldsymbol{x}(t)). \tag{3}$$

Considering the high dimensionality of $\boldsymbol{x}(t)$, it is often expensive to directly use the relationship (3) for predicting the behaviour of the system even when $\boldsymbol{f}(\boldsymbol{x})$ is known. However, many physical systems tend to evolve on a lower-dimensional manifold $\mathcal{Z}$. In that space, the dynamics evolve according to a (generally unknown) function $\boldsymbol{h}(\boldsymbol{z})$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{z}(t) = \boldsymbol{h}(\boldsymbol{z}(t)). \tag{4}$$

Thus, one can predict the dynamics of the system $\boldsymbol{x}$ at a future time T by projecting the initial condition $\boldsymbol{x}(0)$ into the latent space, performing an integration there, and mapping the resulting trajectory back to the observable space:

$$\begin{aligned} \boldsymbol{z}(0) &= \psi^{-1}(\boldsymbol{x}(0)) \\ \boldsymbol{z}(T) &= \boldsymbol{z}(0) + \int_0^T \boldsymbol{h}(\boldsymbol{z}(t))\mathrm{d}t \\ \boldsymbol{x}(T) &= \psi(\boldsymbol{z}(T)). \end{aligned} \tag{5}$$

When $m \ll n$ we refer to the triplet $(\psi, \psi^{-1}, \boldsymbol{h})$ as a Reduced-Order Model (ROM) of $\boldsymbol{f}$. It is often the case that for a given system $\boldsymbol{f}$ there exists no reduced-order representation $(\psi, \psi^{-1}, \boldsymbol{h})$ such that the relation (5) holds exactly. In this case, we seek an approximation ROM $(\psi_{\theta^*}, \phi_{\theta^*}, h_{\theta^*})$ that minimizes the difference between the data $\boldsymbol{x}(t)$ and the prediction $\hat{\boldsymbol{x}}(t)$ over a chosen class of models $(\psi_\theta, \phi_\theta, h_\theta)$ parameterized by θ. We model ψ, ψ^{-1} with a neural network autoencoder ψ_θ, ϕ_θ, and $\boldsymbol{h}$ is modeled by a fully connected network h_θ. The exact constitution of the networks ψ_θ, ϕ_θ, and h_θ are problem-dependent and discussed more closely in the related sections.

We define a *training loss* $\mathcal{L}$ as a sum of the reconstruction and prediction losses. The former ensures that ϕ_θ and ψ_θ are inverse mappings of each other, whereas the latter matches the model's predictions to the available data. Formally, for a given set of trajectories $\boldsymbol{X}_i$, $i \in \{1, \ldots, N\}$, where each trajectory $\boldsymbol{X}_i \in \mathbb{R}^{n \times K}$ is a set of K snapshots $\boldsymbol{X}_i(t_j) \in \mathbb{R}^n$ of the system, for K time-steps, t_j, $j \in \{0, \ldots, K-1\}$, the loss function $\mathcal{L}_\theta^{\text{data}}$ is defined as:

$$\mathcal{L}(\theta) = \sum_{i=1}^{N} \left[\sum_{j=0}^{K-1} \|\boldsymbol{X}_i(t_j) - \psi_\theta(\phi_\theta(\boldsymbol{X}_i(t_j)))\|^2 + \sum_{j=1}^{K-1} \left\| \psi_\theta \left(\phi_\theta(\boldsymbol{X}_i(t_0)) + \int_{t_0}^{t_j} h(\boldsymbol{z}(t))\mathrm{d}t \right) - \boldsymbol{X}_i(t_j) \right\|^2 \right]. \tag{6}$$

To obtain a ROM $(\psi_{\theta^*}, \phi_{\theta^*}, h_{\theta^*})$, we minimize the loss above with respects to the network parameters θ. We note that each trajectory $\boldsymbol{X}_i$ may be captured over its own time-frame and use a distinct, possibly non-uniform, step-size, in which case the loss function should be modified accordingly[1]. To simplify the notation, without loss of generality, in the rest of the paper we assume that all trajectories were recorded over the same time-frame with an equal and uniform step-size Δt.

3.3. Recovery Algorithm

We use the approximate ROM $(\psi_{\theta^*}, \phi_{\theta^*}, h_{\theta^*})$ above to regularize the latent-space dynamics of the single-pixel imaging reconstructions. Specifically, we obtain the reconstruction $\boldsymbol{x}^* = \psi(\boldsymbol{z}^*)$ by minimizing the following loss with respect to $\boldsymbol{Z} = [\boldsymbol{z}_0, \ldots, \boldsymbol{z}_{K-1}] \in \mathbb{R}^{m \times K}$:

$$\mathcal{L}_{\theta^*}^{\text{recon}}(z) = \sum_{j=0}^{K-1} \|y_j - A\psi_{\theta^*}(z_j)\|_2^2 + \lambda \left\| z_0 + \int_{t_0}^{t_j} h_{\theta^*}(z)\mathrm{d}z - z_j \right\|_2^2, \tag{7}$$

where the parameter λ controls the degree to which the compressing sensing algorithm relies on the latent dynamics h_{θ^*} during the signal reconstruction phase. We minimize the loss (7) using a gradient-based automatic differentiation.

4. EXPERIMENTS

4.1. Burgers' Equation

We first study the performance of our framework on flows governed by Burgers' equation with $[-\pi, \pi]$-periodic boundary conditions:

$$\begin{aligned} &u_t + uu_x = \nu u_{xx} \\ &u(-\pi, t) = u(\pi, t), \quad \forall t \in [0, T], \end{aligned} \tag{8}$$

where u_t, u_x, and u_{xx} represent partial derivatives in time, the first, and second spatial derivatives, respectively.

ROM Training. To obtain a training dataset for the ROM we replicate the experimental setup from [9, Section 3], generating 1024 trajectories on a discretized spatial domain $[-\pi, \pi]$ with 128 grid-points. To generate a diverse set of initial conditions we sum the first 10 harmonic terms with random coefficients:

$$u(x, 0) = \frac{1}{10} \sum_{k=1}^{10} a_k \cos(kx) + b_k \sin((k+1)x), \tag{9}$$

where $a_k, b_k \sim \mathcal{N}(0, 1)$. We solve Equation 8 for $t \in [0, 2]$ with $\Delta t = 0.1$ using a spectral solver [30] resulting in trajectories of length $K = 201$ time-steps each.

Architecture. We define ϕ_θ and ψ_θ to be fully-connected neural networks. Both functions had 3 hidden layers 512 neurons wide each, with ReLU activations after all but the output layers. We set the observable space dimension $n = 128$ and the latent space dimension $m = 16$. The network h_θ is a fully-connected network with 3 layers 256 neurons wide each with ReLU activation after each hidden layer except for the output layers.

SPI Recovery. For the recovery phase, we generate 128 trajectories with "bump" initial conditions – a smooth approximation of a bump with two opposing steeply-curved sigmoids:

$$u(x, 0) = \frac{1}{1 + \exp(-k(x - a))} - \frac{1}{1 + \exp(-k(x - b))}, \tag{10}$$

where $a < b$ are sampled uniformly in $[-\pi, \pi]$ and $k = 20$. We choose this shape to ensure that the training and sensing trajectories are sufficiently different to prevent memorization effects. Every row in the sensing matrix A_t is a 128 binary $\{0, 1\}$ vector with 64 non-zero components sampled uniformly.

Performance Evaluation. We compare our SPI-NODE regularization approach with four alternatives: an autoencoder-enhanced variation of differential ghost imaging [27](DGI+Conv. Decoder, [7]), Total Variation regularization (TVR, [15]), Fourier-domain regularized inversion (FDRI, [28]), and an approach which uses an autoencoder (AE, [31]). We treat the Burgers' trajectories as 2D-images in the spatial and temporal domains and measure reconstruction accuracy with three commonly-used metrics: the normalized mean-squared error (MSE), the peak signal-to-noise ratio (PSNR), and the structural similarity index measure (SSIM) [32]. The reconstruction performance is evaluated as a function of the samples-per-frame (SPF) rate, i.e., the number of SPI measurements divided by the number of grid points $n = 128$ in every frame that is recovered.

Results. An example Burgers' trajectory that was recovered from 2% samples per frame SPI measurements is shown in Fig. 2. The results in the figure demonstrate that benchmark methods, which reconstruct each snapshot separately, fail to recover the trajectory because of the extremely low SPF rate. On the other hand, SPI-NODE reconstructs the trajectory as a whole, and since the reduced-order model serves as a strong prior for the dynamics across time, the trajectory is recovered with much lower error. Quantitative SPI recovery results, presented in Fig. 3, show the median PSNR, MSE, and SSIM over different numbers of samples per frame taken. We observe that SPI-NODE regularization achieves $\approx 25\%$ higher PSNR and up to $\approx 33\%$ times higher SSIM relative to FDRI when the SPF rate is low. It also achieves $\approx 70\%$ better PSNR and $2\times$ higher SSIM relative to the autoencoder-based algorithm, which highlights the importance of modeling the temporal dynamics as in SPI-NODE.

4.2. Application to Real Data: Reconstruction of a Gas Plume

In this section we use our technique to reconstruct videos of gas leaks using observations from a simulated single-pixel camera from

[1]The implementation is affected only in evaluating the integral in (6). This part is handled by the `torchdiffeq` [29] library, which supports non-uniform time-frames within a batch.

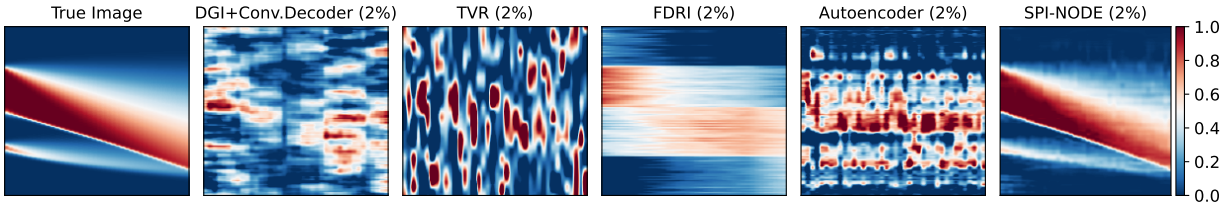


Fig. 2: Example reconstruction of a trajectory governed by Burgers' equation at SPF $= 2/128 \lesssim 2\%$. All algorithms except the proposed SPI-NODE method fail to recover the trajectory due to an insufficient number of samples per frame.

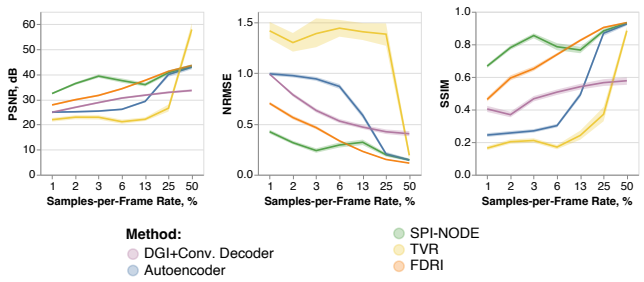
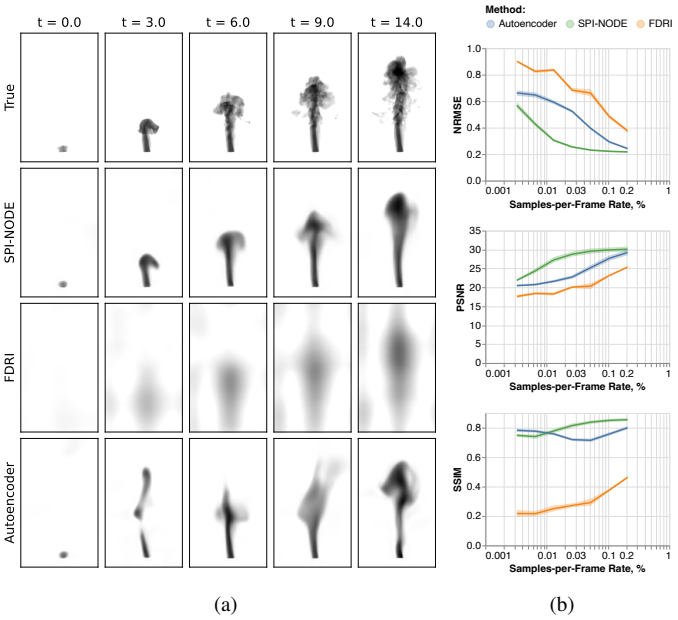


Fig. 3: Reconstruction performance for trajectories governed by Burgers' equation, shown as median PSNR, MSE, and SSIM with 95% confidence intervals. We see that SPI-NODE achieves $\sim 25\%$ higher PSNR and up to $\sim 33\%$ times higher SSIM relative to other algorithms when the SPF rate is low relatively to its strongest competitor (FDRI).



(a) (b)

Fig. 4: SPI reconstructions of gas plumes from the ScalarFlow dataset [33]. The left panel shows SPI reconstructions by three algorithm using 16 samples per frame (0.026% SPF rate). Right panel presents mean PNSR, NRMSE, and SSIM with 95% confidence intervals.

a real-world dataset: ScalarFlow [33]. The physical problem at hand is an ascending turbulent plume in an environment of larger density. The dataset consists of 3D reconstructions of density and velocity of 104 smoke plumes, each 150 frames long. Similar to the Burger's equation setting, the ROM consists of an encoder $\phi(u)$, a decoder $\psi(z)$, and latent dynamics operator $h(z)$. However in this setting, the pair $\phi(u)$ and $\psi(z)$ are represented by a convolutional autoencoder to capture spatial patterns, and the function $h(z)$ is represented by a fully-connected network. We assemble our dataset by taking the front views (2D) of the reconstructions and compressing their resolution to 320×192 pixels. We then split the trajectories into non-overlapping sets: train (92), dev (5), test (5), and reconstruction (2) respectively. We use the first three sets to train, fine-tune, and select a final ROM, respectively. After obtaining the ROM, we use the fourth set of data for SPI reconstruction. For our simulated SPI experiments, we sample 2 to 128 single-pixel observations per frame which corresponds to SPF rates ranging from $3 \times 10^{-3}\%$ to 0.2%.

The results are presented on Figure 4(a), the average MSE and SSIM are displayed on Figure 4(b). We see that SPI-NODE consistently outperforms FDRI by both metrics within the given ranges of SPF rates. For example, on Figure 4(a) we see that FDRI reconstructs a blurry image with 32 samples per frame available (SPF rate of 0.05%), whereas our method is able to reconstruct considerably more accurate image.

5. CONCLUSION

In this paper, we introduced a ROM-based regularization that can be used for single-pixel imaging of scenes with spatio-temporal dynamics governed by an ODE. By training a reduced-order model with a neural ODE, we learn an efficient representation of spatio-temporal flows that can act as a strong prior for SPI reconstruction. SPI-NODE is a good fit for settings when large amounts of clean and representative data are available for training, and when the application in question necessitates a low sample-per-frame ratio (e.g., $< 1\%$). In such settings, our method leverages the ROM to compensate for severly under-sampled measurements with a strong data-driven prior and yields superior recovery quality. Our approach to SPI regularizaton has promise for applications such as gas leak monitoring that are limited by the cost of mid-infrared pixel arrays and the slow speed of raster-scanned methods.

6. ACKNOWLEDGEMENTS

AS, SLB, and JNK acknowledge support from the National Science Foundation AI Institute in Dynamic Systems (grant number 2112085). JNK further acknowledges support from the Air Force Office of Scientific Research (FA9550-19-1-0011).
JR and HM are supported only by MERL.

7. REFERENCES

[1] G. M. Gibson, B. Sun, M. P. Edgar, D. B. Phillips, N. Hempler, G. T. Maker, G. P. A. Malcolm, and M. J. Padgett, "Real-time imaging of methane gas leaks using a single-pixel camera," *Optics Express*, vol. 25, no. 4, pp. 2998–3005, Feb. 2017.

[2] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[3] O. Katz, Y. Bromberg, and Y. Silberberg, "Compressive ghost imaging," *Applied Physics Letters*, vol. 95, no. 13, pp. 131110, 2009.

[4] M. Lyu, W. Wang, H. Wang, H. Wang, G. Li, N. Chen, and G. Situ, "Deep-learning-based ghost imaging," *Scientific Reports*, vol. 7, no. 1, pp. 17865, 2017.

[5] C. F. Higham, R. Murray-Smith, M. J. Padgett, and M. P. Edgar, "Deep learning for real-time single-pixel video," *Scientific Reports*, vol. 8, no. 1, pp. 2369, 2018.

[6] F. Wang, H. Wang, H. Wang, G. Li, and G. Situ, "Learning from simulation: An end-to-end deep-learning approach for computational ghost imaging," *Optics Express*, vol. 27, no. 18, pp. 25560–25572, 2019.

[7] F. Wang, C. Wang, C. Deng, S. Han, and G. Situ, "Single-pixel imaging using physics enhanced deep learning," *Photonics Research*, vol. 10, no. 1, pp. 104–110, 2022.

[8] T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, "Neural ordinary differential equations," in *Advances in Neural Information Processing Systems*, 2018, pp. 6571–6583.

[9] A. Sholokhov, Y. Liu, H. Mansour, and S. Nabi, "Physics-informed neural ODE (PINODE): embedding physics into models using collocation points," *Scientific Reports*, vol. 13, no. 1, pp. 10166, June 2023.

[10] Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur, "A theory of multiplexed illumination," in *IEEE International Conference on Computer Vision*, 2003, vol. 2, pp. 808–815.

[11] P. Sen, B. Chen, G. Garg, S. R. Marschner, M. Horowitz, M. Levoy, and H. P. A. Lensch, "Dual photography," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 745–755, 2005.

[12] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[13] E. J. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*. Madrid, Spain, 2006, vol. 3, pp. 1433–1452.

[14] J. Suo, L. Bian, F. Chen, and Q. Dai, "Signal-dependent noise removal for color videos using temporal and cross-channel priors," *Journal of Visual Communication and Image Representation*, vol. 36, pp. 130–141, 2016.

[15] L. Bian, J. Suo, Q. Dai, and F. Chen, "Experimental comparison of single-pixel imaging algorithms," *Journal of the Optical Society of America A*, vol. 35, no. 1, pp. 78–87, 2018.

[16] M. P. Edgar, G. M. Gibson, and M. J. Padgett, "Principles and prospects for single-pixel imaging," *Nature Photonics*, vol. 13, no. 1, pp. 13–20, Jan. 2019.

[17] G. M. Gibson, S. D. Johnson, and M. J. Padgett, "Single-pixel imaging 12 years on: A review," *Optics Express*, vol. 28, no. 19, pp. 28190–28208, 2020.

[18] L. S. Pontryagin, *Mathematical Theory of Optimal Processes*, Routledge, 2018.

[19] P. Kidger, *On Neural Differential Equations*, Ph.D. thesis, University of Oxford, 2021.

[20] A. Quaglino, M. Gallieri, J. Masci, and J. Koutník, "SNODE: Spectral discretization of neural ODEs for system identification," arXiv:1906.07038, 2019.

[21] D. Ruiz-Balet and E. Zuazua, "Neural ODE control for classification, approximation and transport," arXiv:2104.05278, 2021.

[22] J. Xu, E. Z. Chen, X. Chen, T. Chen, and S. Sun, "Multi-scale neural ODEs for 3D medical image registration," in *Medical Image Computing and Computer Assisted Intervention*, Strasbourg, France, 2021, Springer, pp. 213–223.

[23] A. J. Linot and M. D. Graham, "Data-driven reduced-order modeling of spatiotemporal chaos with neural ordinary differential equations," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 32, no. 7, 2022.

[24] J. M. Dhadphale, V. R. Unni, A. Saha, and R. Sujith, "Neural ODE to model and prognose thermoacoustic instability," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 32, no. 1, 2022.

[25] F. Zhou and L. Li, "Forecasting reservoir inflow via recurrent neural ODEs," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, vol. 35, pp. 15025–15032.

[26] T. B. Pittman, Y. Shih, D. Strekalov, and A. V. Sergienko, "Optical imaging by means of two-photon quantum entanglement," *Physical Review A*, vol. 52, no. 5, pp. R3429, 1995.

[27] F. Ferri, D. Magatti, L. Lugiato, and A. Gatti, "Differential ghost imaging," *Physical Review Letters*, vol. 104, no. 25, pp. 253603, 2010.

[28] A. Pastuszczak, R. Stojek, P. Wróbel, and R. Kotyński, "Differential real-time single-pixel imaging with Fourier domain regularization: applications to VIS-IR imaging and polarization imaging," *Optics Express*, vol. 29, no. 17, pp. 26685–26700, 2021.

[29] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, "Neural ordinary differential equations," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[30] L. N. Trefethen, *Spectral Methods in MATLAB*, SIAM, 2000.

[31] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed sensing using generative models," in *International Conference on Machine Learning*, 2017, pp. 537–546.

[32] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[33] N. T. Marie-Lena Eckert, Kiwon Um, "ScalarFlow: A large-scale volumetric data set of real-world scalar transport flows for computer animation and machine learning," *ACM Transactions on Graphics*, vol. 38(6):239, 2019.