Efficient Methods for Modeling Shift-Varying Operators

Nebiyou Yismaw
University of California Riverside
nyism001@ucr.edu

Chinmay Hegde
New York University
chinmay.h@nyu.edu

M. Salman Asif
University of California Riverside
sasif@ucr.edu

***Abstract*—Many imaging systems require accurate characterization of their forward operators for reliable reconstruction. While shift-invariant systems admit efficient convolutional representations, many practical imaging systems are shift-varying and cannot be captured by classical convolution models. We propose the *Shift-Varying Neural Operator*, an efficient and expressive architecture for learning spatially varying linear operators directly from measurements. Our method builds on existing factorizations of spatially varying convolutions and expresses them in a learnable architecture. Each layer implements a spatially adaptive transformation constructed from a low-rank factorization of modulated convolutional bases. Our experiments show that our proposed method accurately recovers spatially varying point spread functions (PSFs) and learns interpretable operators. Furthermore, we show that the learned forward operator can be integrated into existing iterative inverse problem solvers.**

I. Introduction

Several imaging systems such as microscopy [1], atmospheric imaging [2], multi-aperture optics [3], and other non-stationary systems [4]–[6] can be formulated as general linear mappings between finite-dimensional spaces. Let $\mathbf{x} \in \mathbb{R}^{N\times N}$ denote the unknown scene and $\mathbf{y} \in \mathbb{R}^{N\times N}$ the measured image. The forward model takes the form of $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{A} \in \mathbb{R}^{N^2\times N^2}$ is an imaging operator determined by the system point spread function (PSF). In the continuous domain, the forward model can be expressed as

$$y(u) = \int h(u,v)\, x(v)\, dv, \tag{1}$$

where $h(u,v)$ describes how a point source at location v contributes to the measurement at location u. A special case arises when the PSF is *shift-invariant*, i.e., $h(u,v) = h(u-v)$. In this case, the forward model reduces to convolution and the operator $\mathbf{A}$ becomes a structured convolution matrix. This structure enables significant computational advantages as convolution operators can be computed using fast Fourier transforms (FFTs) in $O(N^2 \log N)$ time.

However, many real-world systems do *not* satisfy shift invariance. Spatially varying aberrations [1], [3], [7] lead to PSFs that change significantly across the field of view. Consequently, the associated operators lack the convenient diagonalization and structure of the convolutional case. In the worst case, an $N \times N$ image with a spatially varying kernel requires $O(N^4)$ storage and computational cost. This makes naive implementations infeasible for high-resolution sensors.

To address this challenge, a broad line of work has explored how to approximate the fully space-variant kernel $h(r,s)$ using structured, computationally efficient decompositions. A comprehensive and unified treatment of these approximations was provided in [7]. The spatially varying PSFs are expressed using a separable expansion

$$h(u,v) \approx \sum_r m_r(u)\, w_r(v),$$

where $\{m_r\}$ denotes a set of basis kernels defining the PSF model and $\{w_r(v)\}$ are spatially varying weights. After substituting this model into (1), the integral reduces to a product convolution approximation of the form

$$y(u) \approx \sum_r \big(m_r * (w_r\, x)\big)(u).$$

Another related approach [8] distinguishes two factorizations of the discrete operator, *column-wise* and *row-wise* decompositions. In the *column-wise* decomposition, each local PSF is expressed as a mixture of basis kernels, leading to an operator of the form

$$\mathbf{A}\mathbf{x} = \sum_r \big(\mathbf{m}_r * (\mathbf{w}_r \odot \mathbf{x})\big),$$

where the spatial weight field $\mathbf{w}_r$ modulates the input first, and the result is then passed through a convolution

This paper is supported in part by the DARPA DIAL program under project HR00112490486 and NSF award 2046293.

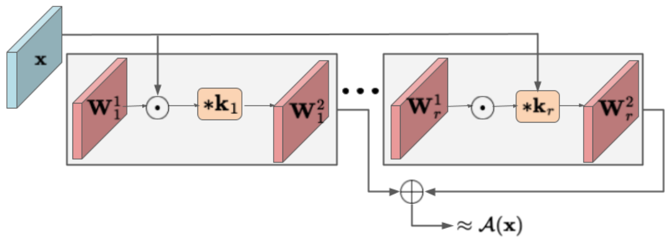

Fig. 1. Shift-varying operator expressed as a low-rank sum of modulated convolutions. Each kernel is applied to a spatially weighted input, and the outputs are summed to form the final result.

with kernel $\mathbf{m}_r$. In the *row-wise* decomposition, the roles are reversed and the operator becomes

$$\mathbf{A}\mathbf{x} = \sum_r \mathbf{w}_r \odot \left(\mathbf{m}_r * \mathbf{x}\right),$$

where each kernel $\mathbf{m}_r$ produces a filtered image that is combined using spatially varying weights.

A complementary perspective comes from the analysis of spatially varying models in [9]. The study identifies two basic constructions, *gathering* and *scattering*. In the gathering formulation, a fixed convolution is applied first and its outputs are combined with spatially varying weights. This follows the same ordering as the *row-wise* model above and is commonly used for image filtering applications. In the scattering formulation, the spatial weights act first and the weighted input is then passed through a fixed convolution. This matches the *column-wise* ordering and is appropriate for optical simulation and image formation models.

In this work, we introduce the **Shift-Varying Neural Operator**, a deep architecture for efficiently representing general spatially varying linear operators. Each layer implements a spatially adaptive linear transformation realized through a learned low-rank factorization of modulated convolutional bases. This construction provides the expressive capacity needed to capture locally varying PSFs while avoiding the large parameter count of fully unstructured operators. The resulting parameterization is computationally efficient and compatible with iterative inverse problem frameworks.

We demonstrate that our approach can accurately recover spatially varying PSFs and learn operator structure directly from data in a supervised manner. Our method provides a flexible and powerful alternative to existing structured approximations for shift-varying systems.

II. Proposed Method

To learn shift-varying imaging operators in a scalable manner, we adopt a structured parameterization based on spatially modulated convolutions. This viewpoint builds on earlier decomposition strategies to design a learnable neural framework. Instead of pre-specifying bases or computing global factorizations, the model learns weight maps and kernels directly from data while preserving the interpretability and efficiency of low-rank decomposition structure.

Shift-varying operator parameterization. We represent a shift-varying linear operator using spatial modulation and convolution. For an input image $\mathbf{x}$, a single operator layer is defined as

$$\mathcal{A}(\mathbf{x}) = \sum_{r=1}^{R} \Big(\big(\mathbf{x} \odot \mathbf{W}_r^{(1)}\big) * \mathbf{k}_r \Big) \odot \mathbf{W}_r^{(2)}, \qquad (2)$$

where $\mathbf{W}_r^{(1)}$ and $\mathbf{W}_r^{(2)}$ are spatial weight maps applied before and after convolution, $*$ denotes convolution, and $\mathbf{k}_r$ is the kernel associated with index r. This representation allows each kernel to process a spatially modulated version of the input and then apply a second spatial modulation to the filtered response. The resulting linear operator is expressive enough to approximate general space-variant imaging models.

This structure relates closely to several classical formulations of space-variant blur. Modulation before convolution corresponds to the column-wise PSF expansion in [7], while modulation after convolution reflects the row-wise formulation in [8]. Similarly, our method also combines both scattering and gathering interpretations from [9] into the learning framework.

Efficient representation of weight maps. A full-resolution weight map $\mathbf{W}_r^{(j)}$ for an $N \times N$ image would require N^2 parameters per index r, which is not scalable. To reduce dimensionality, we parameterize each weight map at a lower resolution,

$$\widehat{\mathbf{W}}_r^{(j)} \in \mathbb{R}^{(N/d)\times(N/d)}, \qquad j = 1, 2,$$

and obtain the full-resolution map through bilinear interpolation, $\mathbf{W}_r^{(j)} = \mathcal{U}\Big(\widehat{\mathbf{W}}_r^{(j)}\Big)$, where $\mathcal{U}$ denotes 2D upsampling to size $N \times N$. This reduces the number of learnable weight parameters by a factor of d^2. For a kernel of size $K \times K$, the parameter count of a single layer is $R\left(\frac{N^2}{d^2} + K^2\right)$, which scales well to large images.

Shift-Varying Neural Operator. To increase modeling capacity, we compose several operators of the form in equation (2). Let $\mathcal{A}^{(1)}, \ldots, \mathcal{A}^{(L)}$ denote these layers. The overall mapping is

$$\mathcal{A}^L(\mathbf{x}) = \mathcal{A}^{(L)} \circ \mathcal{A}^{(L-1)} \circ \cdots \circ \mathcal{A}^{(1)}(\mathbf{x}), \qquad (3)$$

Fig. 2. The first column shows the spatial mask and the motion blur kernel (bottom-left corner). The last two columns show the measurement and the ground-truth images.

TABLE I
SUMMARY OF LEARNED SHIFT-VARYING OPERATORS.

Forward model	# Parameters	Measurement Loss
Shift-varying motion blur	0.1M	4.7×10^{-5}
Ring convolution	0.59M	4.0×10^{-4}
Wide-field microscopy	2.1M	4.5×10^{-3}

which creates a deep sequence of spatially adaptive linear transformations. Each layer is implemented using an input and output modulated convolutional structure. All parameters, including the kernels and the modulation weights are learned directly from data.

III. Experiments and Results

We conduct a set of experiments using fully simulated data derived from natural images in the DIV2K dataset [10]. We evaluate the ability of the Shift-Varying Neural Operator to learn three spatially varying forward models: (1) a simulated spatially varying motion blur operator, (2) a ring-convolution microscopy model [1], and (3) a wide-field microscopy PSF [3]. We train our model on randomly extracted 256×256 patches from DIV2K, and all models are implemented in PyTorch. Each measurement operator is learned in a supervised manner by minimizing the measurement loss.

A. Shift-Varying Motion Blur

We begin with a simulation-based experiment designed to evaluate whether the proposed model can recover the structure of a spatially varying linear system. In this setting, we construct a synthetic forward operator that applies motion blur selectively across the image. The operator is defined by a fixed motion blur kernel together with a spatial mask that specifies where the blur is applied. Regions outside the mask are passed through unchanged. Figure 2 shows an example input image, the resulting measurement, and the forward-operator parameters. This system can be represented exactly as a rank-$R = 2$ model under the formulation in (2).

Learning the operator. We train the Shift-Varying Neural Operator using 128 measurement pairs generated from DIV2K [10]. The network is trained to minimize the ℓ_2 loss generated by the true shift-varying motion blur operator. The trained operator achieved an MSE validation loss of 4.7×10^{-5}, as shown in Table I.

Reconstruction. We then evaluate whether the learned measurement operator can support model-based image recovery. For this experiment, we use the learned forward and adjoint operators within a plug-and-play reconstruction method based on DPIR. Figure 3 shows that reconstructions obtained with the learned operator are visually similar to those produced using the true shift-varying operator: textures, contrast, and the spatially varying blur pattern are all reliably recovered. These results indicate that the learned model provides sufficiently accurate gradients for iterative reconstruction and can be reliably used within a model-based pipeline.

B. Ring Convolution

Ring convolution describes a class of spatially varying imaging models in which each location is blurred by a radially symmetric PSF whose shape depends on the distance from the optical center [1]. In these systems, the PSF changes radially with distance from the optical center but is invariant with respect to the angular coordinate. Such models arise naturally in microscopes with rotationally symmetric aberrations.

The ring-convolution model in [1] is implemented by transforming the image to polar coordinates, applying radius-dependent 1D convolutions along the radial dimension, and mapping the result back to Cartesian space. We generate training pairs using this forward model and supervise our model to learn the corresponding shift-varying blur. The learned operator contains approximately 5.9×10^5 parameters and achieves a measurement loss of 4.0×10^{-4} on the validation set. The trained model generates measurements in a single forward pass, providing a computationally efficient alternative to the original ring-convolution implementation, which requires per-radius PSF evaluation and explicit coordinate transforms.

C. Wide-Field Microscopy

We conduct an additional study to learn a spatially varying point spread function from a microlens array. For this setup, we construct a synthetic forward model by selecting the calibrated PSFs from a single microlens in

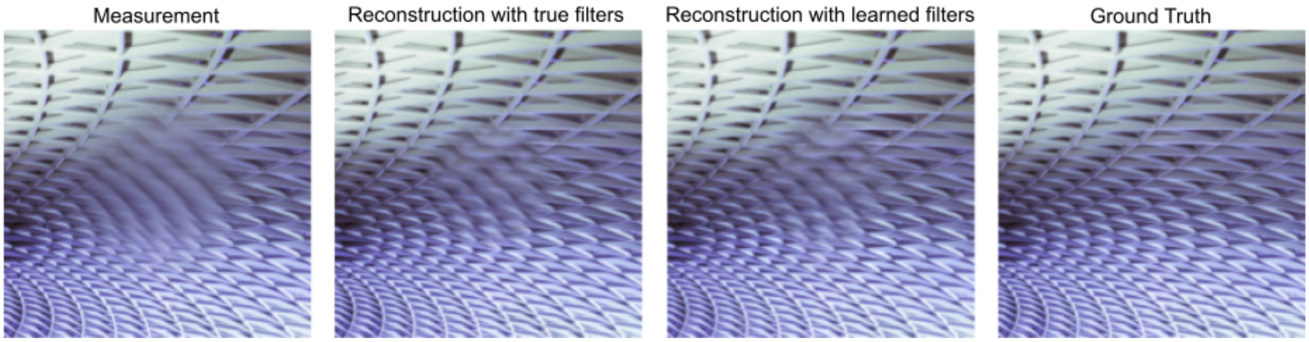

Fig. 3. Reconstruction results show that the learned operator produces outputs that closely match the output obtained using the true filters.

the multi-aperture system of [3]. Following the procedure in the paper, we use sampled PSFs on a uniform grid across the field of view and perform a truncated SVD to obtain a compact low-rank basis that captures the smooth spatial variation of the aberrations. We use this system to obtain simulated measurements. Using these measurements, we train our shift-varying operator with 2.1M parameters to approximate the wide-field forward process directly from input images. The learned operator achieves a final measurement error of 4.5×10^{-3}.

IV. Discussion and Limitations

While the proposed Shift-Varying Neural Operator is effective across a wide range of spatially varying imaging models, its performance depends on the intrinsic structure of the forward operator. In our wide-field microscopy experiment, the ground-truth measurements are generated using an SVD-based approximation of the space-variant PSF field. At low truncation ranks, the model is able to learn the forward map accurately. However, as we increase the truncation rank and the underlying system becomes more complex, we observe a degradation in performance and the measurements from the learned operator begin to deviate from the true measurements. This suggests that modeling highly complex shift-varying systems may require enhancing the proposed operator architecture.

V. Conclusion

We presented a data-driven framework for learning shift-varying measurement operators and demonstrated its effectiveness on multiple spatially varying imaging models. By training on simulated pairs generated from physically grounded forward operators, our approach accurately recovers both local kernel structure and global field-dependent variations. The learned operators can also be integrated into iterative reconstruction. This enables high-quality image recovery using the learned filters. Experiments on spatially varying models show that the proposed method provides a compact, scalable, and consistent representation of spatially varying systems.

VI. Acknowledgment

The authors would like to acknowledge the support of the DARPA DIAL program under project HR00112490486 and NSF award 2046293.

References

[1] A. Kohli, A. N. Angelopoulos, D. McAllister, E. Whang, S. You, K. Yanny, F. M. Gasparoli, B.-J. Chang, R. Fiolka, and L. Waller, “Ring deconvolution microscopy: exploiting symmetry for efficient spatially varying aberration correction,” *Nature Methods*, pp. 1–10, 2025.

[2] S. H. Chan, N. Chimitt, *et al.*, “Computational imaging through atmospheric turbulence,” *Foundations and Trends® in Computer Graphics and Vision*, vol. 15, no. 4, pp. 253–508, 2023.

[3] Q. Yang, R. Guo, G. Hu, Y. Xue, Y. Li, and L. Tian, “Wide-field, high-resolution reconstruction in computational multi-aperture miniscope using a fourier neural network,” *Optica*, vol. 11, no. 6, pp. 860–871, 2024.

[4] M. Sorel and J. Flusser, “Space-variant restoration of images degraded by camera motion blur,” *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 105–116, 2008.

[5] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, “Non-stationary correction of optical aberrations,” in *2011 International Conference on Computer Vision*, pp. 659–666, IEEE, 2011.

[6] D. Xie, W. Dong, J. Zheng, and C. Tian, “Spatially-variant image deconvolution for photoacoustic tomography,” *Optics Express*, vol. 31, no. 13, pp. 21641–21657, 2023.

[7] L. Denis, E. Thiébaut, F. Soulez, J.-M. Becker, and R. Mourya, “Fast approximations of shift-variant blur,” *International Journal of Computer Vision*, vol. 115, pp. 253–278, 2015.

[8] F. Sroubek, J. Kamenicky, and Y. M. Lu, “Decomposition of space-variant blur in image deconvolution,” *IEEE signal processing letters*, vol. 23, no. 3, pp. 346–350, 2016.

[9] N. Chimitt, X. Zhang, Y. Chi, and S. H. Chan, “Scattering and gathering for spatially varying blurs,” *IEEE Transactions on Signal Processing*, vol. 72, pp. 1507–1517, 2024.

[10] E. Agustsson and R. Timofte, “Ntire 2017 challenge on single image super-resolution: Dataset and study,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.