Neural Fields for Single-Shot Image Reconstruction in Fiber Bundle Imaging Systems

Amir Reza Vazifeh, Jason W. Fleischer
Department of Electrical and Computer Engineering, Princeton University

PRINCETON UNIVERSITY

FiO LS
Frontiers in Optics + Laser Science

GitHub Project Page

1. Introduction

- **Background:** Fiber bundle imaging is widely used in endoscopy, OCT, and cellular imaging. The discrete fiber arrangement, however, limits lateral resolution and introduces sampling artifacts, motivating computational image restoration.
- **Single-frame Reconstruction:** These methods either train a supervised model on paired distorted-to-clean images [1], or use the known fiber layout for interpolation [2] or inverse reconstruction [3]. The former generalizes poorly beyond its training distribution or to unseen fiber layouts; the latter requires precise geometric calibration.
- **Multi-frame Reconstruction:** Using multiple frames can reveal information lost in individual frames but requires alignment and fusion across views. For frame alignment, some methods impose explicit motion priors based on known probe shifts or fiber-core spacing [4, 5]. Others estimate inter-frame transformations directly from data [6, 7].
- **Neural Fields:** Neural radiance fields are an emerging approach for modeling signal values as functions of spatial or temporal coordinates, parameterized by MLPs [8]. Here, we first introduce a single-shot method that reconstructs a high-resolution image from one blurred, fiber-masked measurement using a coordinate network with the known fiber layout and PSF. We then introduce a multi-frame method that jointly aligns misaligned image bursts and reconstructs a clean scene using two coordinate MLPs trained end-to-end, without ground truth, fiber layout, or motion priors.

2. Single-Shot Methodology

- **Inverse Problem:** Recover a clean, high-resolution image G from a single blurred, fiber-masked measurement M.
- **Forward Model:** A coordinate network f_θ — an MLP with learnable parameters θ — maps each 2D position (x, y) to scene intensity G ($G \in \mathbb{R}$ for grayscale, $G = [R, G, B] \in \mathbb{R}^3$ for RGB). The estimated measurement $\widehat{M}$ is obtained by convolving G with the known PSF and multiplying by the known fiber mask F:

$$\widehat{M}(x,y) = (G * PSF) \times F(x,y) \qquad G(x,y) = f_\theta\big(\gamma(x,y)\big)$$

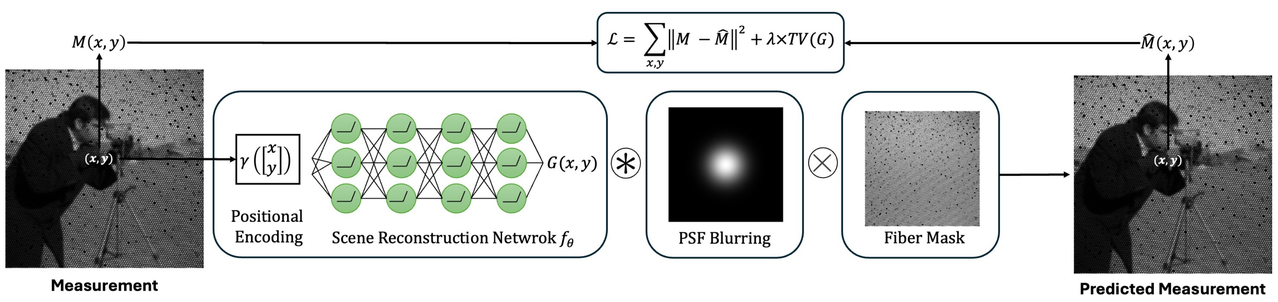
- **Positional Encoding:** MLPs are biased toward low-frequency output and struggle with fine detail. Following NeRF [8], coordinates are mapped to a higher-dimensional space using sinusoids at logarithmically spaced frequencies:

$$\gamma(x,y) = [x, \sin(2^0\pi x), \cos(2^0\pi x), \ldots \sin(2^L\pi x), \cos(2^L\pi x), y, \sin(2^0\pi y), \cos(2^0\pi y), \ldots \sin(2^L\pi y), \cos(2^L\pi y)]$$

- **Optimization:** At test time, f_θ is optimized to minimize the reconstruction loss between predicted measurement $\widehat{M}$ and real measurement M, with a total-variation (TV) regularization on G, where λ controls the smoothness of G:

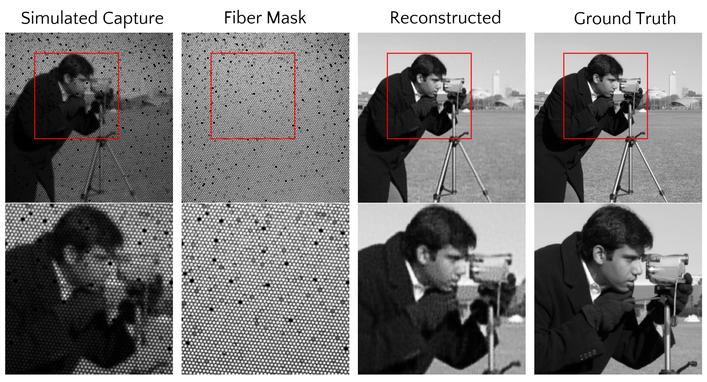
$$loss = \sum_{x,y} \left\| M(x,y) - \widehat{M}(x,y) \right\|^2 + \lambda \times TV(G)$$

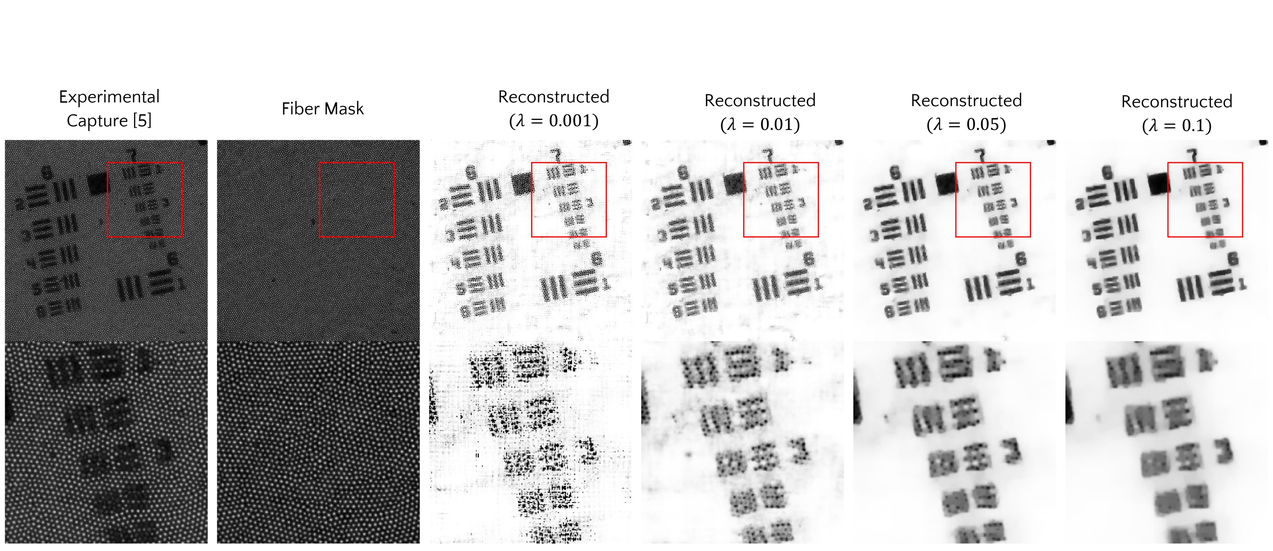
- **Inference:** After training, the high-resolution scene can be reconstructed by feeding coordinates (x, y) into f_θ.



3. Single-Shot Results

- **Simulation Validation:** The Cameraman image is Gaussian-blurred with $\sigma = 1.5$ and known fiber mask is applied . Reconstructed over 25,000 iterations with $\lambda = 0.001$, reaching (PSNR/SSIM) of (29.03 dB/0.81), a (+18.65dB/+0.62) gain over the raw measurement.
- **Experimental Validation:** Resolution-chart image from [5], captured with a Fujikura fiber bundle (FIGH-30-650S). Fiber mask obtained via a white-reference calibration; true PSF unavailable, approximated as Gaussian ($\sigma = 0.5$). $\lambda = 0.05$ gives the sharpest artifact-free reconstruction.





4. Multi-Frame Methodology [7]

- **Joint Frame Alignment and Fusion:** The fiber bundle artifacts are removed from a burst of misaligned frames using two components: a motion estimation network that represents inter-frame transformations, and a scene representation network. The reconstructed RGB value $\hat{I}$ at coordinate (x, y) in frame t is given by:

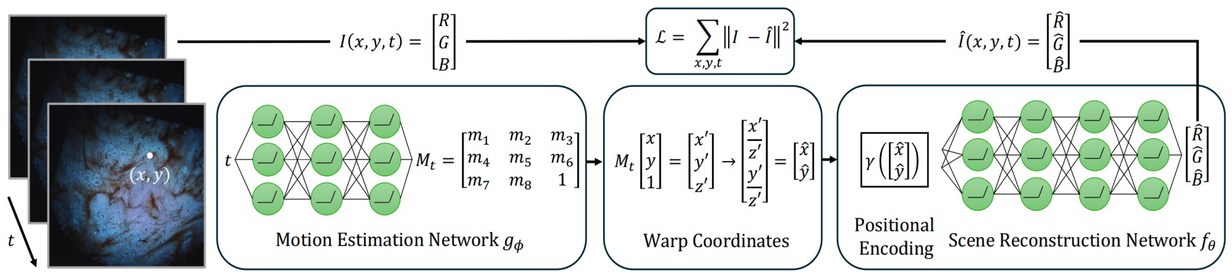
$$\hat{I}(x,y,t) = \left[\hat{R}, \hat{G}, \hat{B}\right] = f_\theta\left(\gamma(T_{g_\phi}(x,y,t)\right)$$

- **Frame Alignment (Motion Estimation):** T_{g_ϕ} represents a spatial transformation (e.g., homography or optical flow) parameterized by neural network g_ϕ.

$$\text{Homography} \rightarrow M_t = g_\phi(t) = \begin{bmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ m_7 & m_8 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = M_t \begin{bmatrix} x \\ y \\ z \end{bmatrix} \rightarrow (\hat{x}, \hat{y}) = \left(\frac{x'}{z'}, \frac{y'}{z'}\right)$$

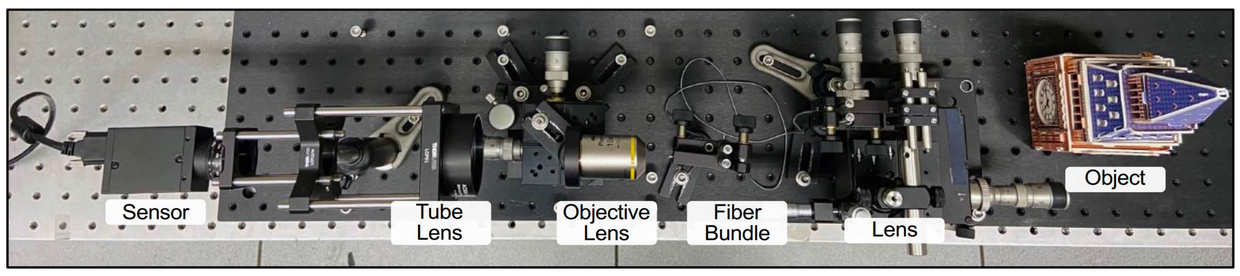
$$\text{Optical Flow} \rightarrow (\Delta x_t, \Delta y_t) = g_\phi(x,y,t) \rightarrow (\hat{x}, \hat{y}) = (x + \Delta x_t, y + \Delta y_t)$$

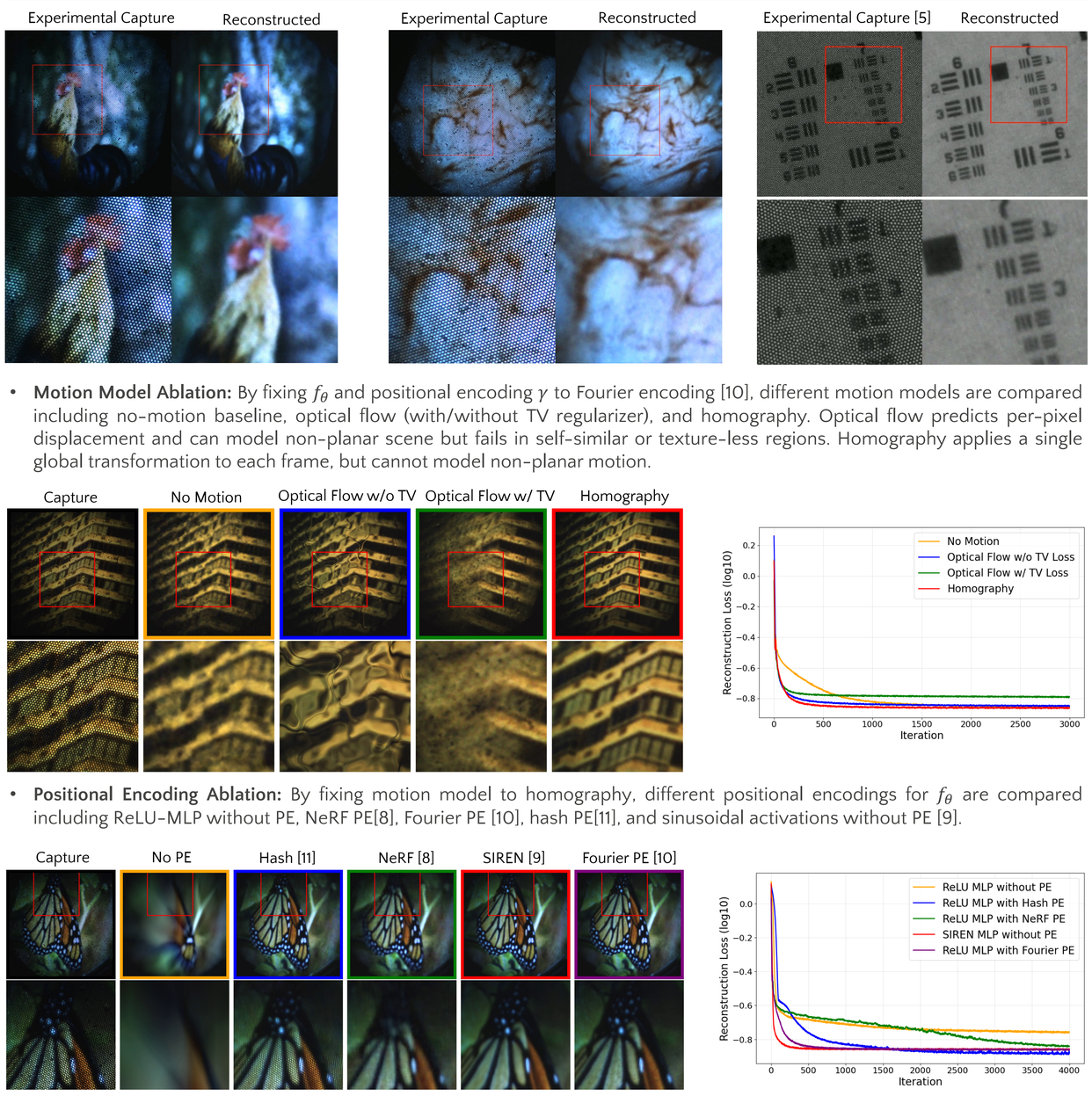
- **Optimization:** The warped coordinates $(\hat{x}, \hat{y}) = T_{g_\phi}(x, y, t)$ are fed into the scene representation network f_θ, which maps them to RGB values. To mitigate spectral bias — MLPs' bias toward low-frequency output— the warped coordinates $(\hat{x}, \hat{y})$ are first passed through a positional encoding γ. Networks g_ϕ and f_θ are trained jointly to minimize the MSE loss between I and $\hat{I}$.
- **Inference:** After training, the high-resolution video can be reconstructed by feeding coordinates (x, y, t) into g_ϕ and f_θ.



5. Multi-Frame Results

- **Experimental Validation:** A fiber bundle imaging system (38°FOV lens, SCHOTT fiber bundle — 7.6 μm core, ~18,000 fibers) captured 20-frame videos of screen-displayed images under applied random motion. We reconstruct high-resolution videos from these bursts without any additional data — fiber layout, ground truth, or motion prior. We also tested on the resolution chart burst from [5] to assess robustness across fiber configurations.





- **Motion Model Ablation:** By fixing f_θ and positional encoding γ to Fourier encoding [10], different motion models are compared including no-motion baseline, optical flow (with/without TV regularizer), and homography. Optical flow predicts per-pixel displacement and can model non-planar scene but fails in self-similar or texture-less regions. Homography applies a single global transformation to each frame, but cannot model non-planar motion.

- **Positional Encoding Ablation:** By fixing motion model to homography, different positional encodings for f_θ are compared including ReLU-MLP without PE, NeRF PE[8], Fourier PE [10], hash PE[11], and sinusoidal activations without PE [9].

6. References

[1] J. Shao et al., "Fiber bundle image restoration using deep learning," Opt. Lett. 44, 1080-1083 (2019).
[2] P. Wang et al., "Fiber pattern removal and image reconstruction method for snapshot mosaic hyperspectral endoscopic images," Biomed. Opt. Express 9, 780-790 (2018).
[3] J. Shao et al., "Resolution enhancement for fiber bundle imaging using maximum a posteriori estimation," Opt. Lett. 43, 1906-1909 (2018)
[4] C. Renteria et al., "Depixelation and enhancement of fiber bundle images by bundle rotation," Appl. Opt. 59, 536-544 (2020).
[5] M. Hughes, "Real-timing processing of fiber bundle endomicroscopy images in Python using PyFibreBundle," Appl. Opt. 62, 9041-9050 (2023).
[6] J. Shao et al., "Fiber bundle imaging resolution enhancement using deep learning," Opt. Express 27, 15880-15890 (2019).
[7] A. Vazifeh et al., "Seeing through fibers: unsupervised image reconstruction in fiber bundle imaging systems," Opt. Express 34, 7429-7444 (2026).
[8] B. Mildenhall et al., "NeRF: representing scenes as neural radiance fields for view synthesis," Commun. ACM 65, 99-106 (2022).
[9] V. Sitzmann et al., "Implicit neural representations with periodic activation functions," in Advances in Neural Information Processing Systems (2020).
[10] M. Tancik et al., "Fourier features let networks learn high frequency functions in low dimensional domains," in Advances in Neural Information Processing Systems (2020), pp. 7537-7547.
[11] T. Müller et al., "Instant neural graphics primitives with a multiresolution hash encoding," ACM Trans. Graph. 41, 1-15 (2022).

7. Acknowledgements

- The authors thank the members of the Princeton Computational Imaging Lab, Congli Wang and Prof. Felix Heide, for valuable discussions and contributions to the multi-frame method [7].