LOCALIZED RESPONSE BASIS FOR DATA-EFFICIENT TRANSVERSE BEAM DISTRIBUTION RECONSTRUCTION USING A MULTIMODE FIBER RELAY

Q. Xu*[1,2], G. Trad[2], S. Burger[2], A. Gilardi[2], M. Bergamaschi[2], F. Roncarolo[2],
H. D. Zhang[1], A. Hill[1], C. P. Welsch[1]
[1] University of Liverpool, Liverpool, UK; Cockcroft Institute, Daresbury, UK
[2] CERN, Geneva, Switzerland

Abstract

Transverse beam imaging in radiation areas can be supported by relaying scintillation light through a multimode fiber (MMF) to a camera placed in a shielded area. However, the MMF scrambles the input, so a trained model is required to recover the beam distribution. This work studies a data efficient calibration method in which measured input and MMF output basis pairs are used as building blocks to synthesize training data for the reconstruction model. After an initial digital micromirror device based validation, the method was assessed using real beam data from CERN CLEAR, where data synthesized from a raster scan basis were used to train a Convolutional Autoencoder. The best model using this strategy achieved 7.37% mean normalized root mean square error (RMSE) across four transverse beam parameters, compared with 6.02% for a random scan reference model using roughly twice as many fully paired random scan samples. These results suggest that basis-based synthesis training, when combined with suitable beam image priors, can reduce reliance on large random scan MMF calibration datasets by replacing part of the calibration with a controlled scan of fixed size.

INTRODUCTION

In high radiation accelerator areas, transverse beam imaging can be limited by radiation damage to cameras and associated electronics [1]. One possible mitigation is to relay light from a scintillating screen through a large-core multimode fiber (MMF) to a camera placed in a low radiation area [2, 3]. However, modal mixing in the MMF scrambles the spatial intensity distribution, so the fiber output is not directly interpretable. An inverse reconstruction model is therefore required. Previous work has shown that Convolutional Autoencoder (CAE) models can reconstruct transverse beam distributions from MMF output images [4]. A practical challenge remains the amount and coverage of paired calibration data required for training. In particular, acquiring large paired random scan datasets can be time consuming and may not be convenient for recalibration.

This work investigates whether a measured MMF input-output basis can be used as a compact calibration source for synthetic training data. The approach is motivated by the approximate linearity of the measured intensity mapping under fixed alignment and acquisition conditions [5, 6]. A set of measured input patterns and corresponding MMF output images is used as the basis, and new paired training samples are generated by linearly combining this basis with different coefficient maps. The method is first tested using a controlled digital micromirror device (DMD) source under LED illumination, where a position basis is used to validate the synthesis procedure. It is then assessed using data from the CERN CLEAR facility, where a raster scan basis is used for calibration and independent beam images are used for evaluation.

DMD-SOURCE VALIDATION

MMF Intensity Basis

Light transport through a multimode fiber is often described by a transmission matrix (TM), which maps an input optical field to an output optical field. In optics laboratory systems, such a matrix can be calibrated using an orthogonal basis, such as a Hadamard basis [7]. For accelerator beam imaging context, the input is instead an intensity distribution formed on a screen, so arbitrary optical patterns are not the most suitable calibration target. We therefore use a simple position basis, where the transverse distribution is represented by square blocks on a finite grid. This basis resembles a raster scan and is more directly connected to beam based calibration. After background subtraction and within the linear camera signal range, the MMF relay is treated as a fixed intensity mapping. The synthetic input and output images are written as

$$\begin{aligned} I_{\text{in}}^{\text{syn}}(x, y) &= \sum_{i,j} a_{ij} B_{ij}^{\text{in}}(x, y), \\ I_{\text{out}}^{\text{syn}}(u, v) &= \sum_{i,j} a_{ij} B_{ij}^{\text{out}}(u, v). \end{aligned} \tag{1}$$

Here B_{ij}^{in} is the input square block at grid position (i, j), B_{ij}^{out} is the MMF output image measured for that block, and a_{ij} is the shared intensity coefficient. The same coefficient map is used to synthesize both input and output images. An $n_x \times n_y$ position basis requires $n_x n_y$ measured basis pairs.

DMD Setup and Data Synthesis

To validate the position basis method under controlled conditions, an optical setup based on a DMD was used before the beamline assessment. The setup follows the same MMF relay concept shown in Fig. 1, with the scintillating screen replaced by a DMD. A 680 nm peak emission LED

---

* qiyuan.xu@liverpool.ac.uk

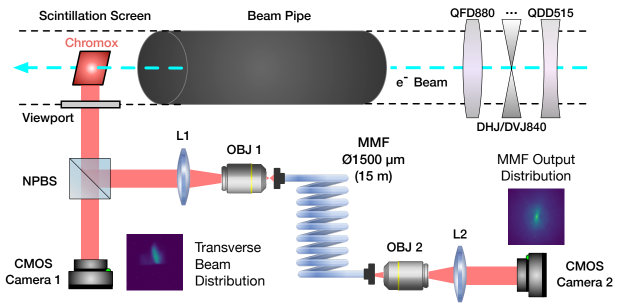

Figure 1: Schematic of the MMF relay concept for transverse beam distribution imaging. The beamline implementation is shown; for the DMD validation, the same optical concept is used with the scintillating screen replaced by a DMD source.

(LED680L) illuminated the DMD and provided the input intensity patterns. This wavelength was chosen to be close to the Chromox emission peak around 694 nm, while also providing incoherent illumination. The light was split by a non polarising beam splitter (BS031). One arm was recorded by a reference camera (acA1920-40gm), while the other was coupled into a 1 m long MMF (FP1500ERT). The fiber output was imaged by a second CMOS camera to record the corresponding MMF image. Data were collected in two stages. First, a 32×32 position basis was measured by displaying one square block at a time on the DMD, scanned row by row. The block intensity was set high enough to improve the signal-to-noise ratio while remaining below saturation. This gave 1024 paired basis samples. Second, evaluation data were collected by displaying measured CLEAR transverse beam distributions on the DMD and recording the corresponding reference and MMF images.

The synthetic training data were generated from the measured position basis. A Stochastic Gaussian Mixture (SGM) model, which samples mixtures of 2D Gaussian components with varied centroids, widths, and amplitudes [8], was used to generate diverse beam-like transverse distributions. Each generated distribution was downsampled to the same 32×32 grid as the position basis, giving the coefficient map a_{ij}. The coefficient map was then applied directly to the measured basis pairs by index matching, giving paired synthetic input and MMF output images for CAE training. Using synthetic training data generated from the 1024 measured basis pairs, the trained CAE achieved a 1.5% mean normalized root mean square error (RMSE) across four transverse beam parameters on the DMD test set. This result motivated the following assessment with real beam data using the same basis synthesis concept.

CLEAR BEAMLINE ASSESSMENT

Beamline Raster Scan Acquisition

For the real beam study, the ideal position basis cannot be produced exactly, because the beam cannot be set to arbitrary square blocks with perfectly fixed size and intensity. The practical approximation was to use a focused beam spot on the Chromox screen as the scanning unit. The raster-scan spots were approximately Gaussian, with a fitted Gaussian width $\sigma_{\text{basis}} = 0.90 \pm 0.06$ mm on the screen. This spot size defines the effective spatial scale of the raster scan. The beam parameters used in the experiment are listed in Table 1. After tuning to this focused spot condition, a raster scan was performed line by line using upstream corrector magnets to vary the beam position on the screen. A second dataset was then collected using a random scan, where both the beam position and transverse shape were varied to provide independent evaluation data.

Table 1: CLEAR Beam Parameters Used for Data Collection

Beam configuration	**Specifications**
Particle type	e^-
Beam energy	192–199 MeV
Repetition rate	10 Hz
Bunches per pulse	1–25
Total pulse charge	~500 pC (15 bunches)

The optical relay followed the same concept as described above and shown in Fig. 1. The main difference was that the MMF length was increased to 15 m, routing the light from the CLEAR beamline to the upstairs shielded camera area. In total, 1503 raster scan samples were used to form the basis. A separate random scan dataset of 4078 samples was collected for evaluation and reference comparisons. Before training, 636 samples were held out as a common test set for all CLEAR results, while the remaining 3442 samples were reserved for beam image priors and for the directly trained random scan reference model.

Transverse Beam Distribution Reconstruction

For the CLEAR data, the synthesis pipeline was adapted because the raster scan basis was measured with real beam spots rather than ideal square blocks. Each coefficient source was first reduced to a 23×23 coefficient map. This was chosen as a practical compromise between the measured beam spot size and avoiding intensity saturation in the synthesized images. Each raster scan sample was assigned a centroid position on the screen. For each coefficient in the map, the nearest raster scan sample was selected according to the distance between the coefficient cell centre and the measured centroid position. A random jitter of maximum half grid cell was applied to the coefficient map before matching. This increased the diversity of the synthesized samples and allowed more raster scan samples to contribute during training.

The coefficient maps were obtained from two sources. The first used the 3442 non-test random scan beam images as beam image priors. The second used SGM distributions. A mixed dataset combining both sources was also tested. For comparison, a CAE was trained directly on the raster scan samples, and a linear TM baseline was built from the same raster scan basis. A random scan reference CAE was also trained using the non-test random scan pairs.

For the best model trained with synthesized data from beam image priors, representative CLEAR test samples are shown in Fig. 2. The model is more reliable for single-peak

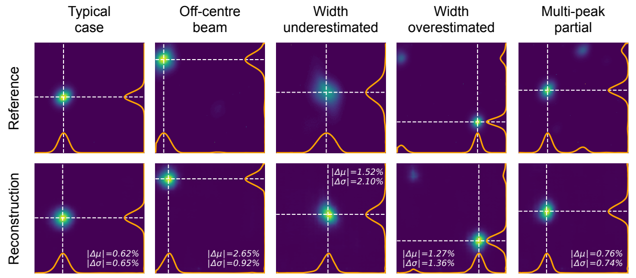

Figure 2: Representative CLEAR test set reconstructions using the CAE trained with beam image prior synthesis. Top row: reference images; bottom row: reconstructed images. Orange curves show transverse intensity profiles, dashed white lines mark centroid positions, and labels give normalized centroid and width errors.

beams with moderate transverse size and centroids near the image centre, while larger errors are observed for beams with centroids close to the image boundary, very small or broad beams, and cases with multi-peak structure.

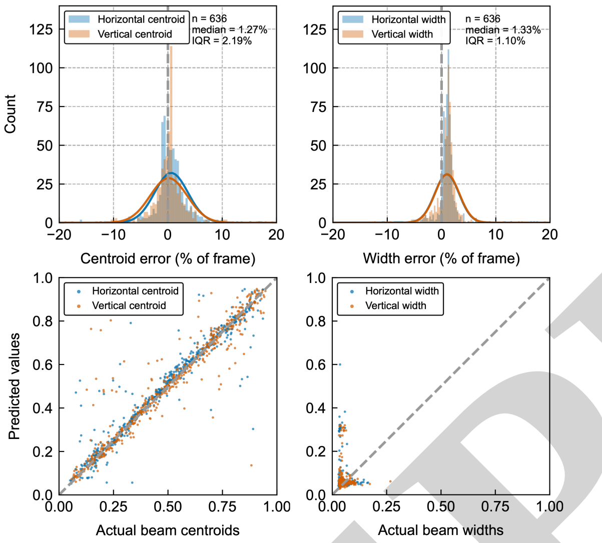

Figure 3: CLEAR test set statistics over 636 images. Top row: normalized centroid and width error histograms with Gaussian fits. Bottom row: predicted versus reference centroid and width values, with dashed diagonal lines indicating ideal agreement.

The test set statistics for the model trained with beam image prior synthesis are shown in Fig. 3. The width errors are slightly shifted toward positive values, indicating a tendency to predict larger beam sizes. This is likely caused by the finite size and Gaussian spread of the raster scan beam spots: when synthetic samples are formed by combining these components, the resulting distributions can become slightly broadened. This remains a limitation of using the finite raster-scan beam spot as the basis element.

The comparison over different training strategies is shown in Fig. 4. A pseudo TM formed from the raster scan basis by least squares fitting is included as a baseline. Among the basis methods, beam image prior synthesis gives the best result and approaches the random scan reference, although the reference model remains better. The SGM based case performs worse, likely because its coefficient maps do not fully match the beam distribution and need further tuning.

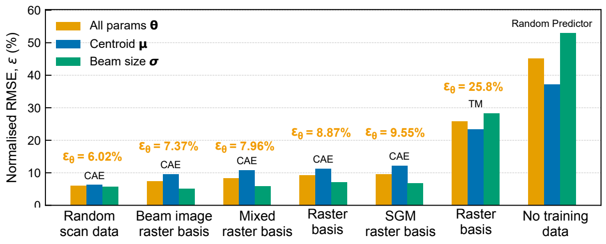

Figure 4: Mean normalized RMSE on the CLEAR test set for different training strategies. $\theta = (\mu_x, \mu_y, \sigma_x, \sigma_y)$ denotes all four transverse beam parameters.

CONCLUSION

This study first used a DMD setup to validate the position basis synthesis method for the MMF relay, and then extended the idea to real beam data from the CLEAR beamline. The results suggest that a raster scan basis can provide a practical route toward reducing the need for large random scan datasets in MMF beam imaging. The model trained with beam image prior synthesis achieved a mean normalized RMSE of 7.37% across four transverse beam parameters, compared with 6.02% for the random scan reference model, which used roughly twice as many fully paired random-scan samples. The method also changes the calibration process from an open ended random scan procedure to a controlled scan whose size is set by the basis resolution and scan range. The remaining limitations are that the synthetic coefficient maps must match the target beam distribution, and that the finite Gaussian width of the raster scan beam spot limits the achievable spatial resolution. Future work will investigate improved SGM tuning with reduced dependence on measured beam priors, and optical basis calibration using an external light source as an alternative to beam-steered scans.

ACKNOWLEDGEMENTS

This work was supported by the Science and Technology Facilities Council (STFC) through the LIV.INNO Centre for Doctoral Training under grant agreement ST/W006766/1. The authors would like to thank the CLEAR operation team and the CERN Beam Instrumentation Group for their support in the preparation and execution of the beamline measurements at the CLEAR facility. Model training was performed on Barkla, part of the High Performance Computing facilities at the University of Liverpool, UK.

REFERENCES

[1] F. Roncarolo *et al.*, “Review of CERN beam instrumentation for fixed target experiments”, in *Proc. IPAC’23*, Venice, Italy, May 2023, pp. 4625–4628. doi:10.18429/JACoW-IPAC2023-THPL080

[2] G. Trad and S. Burger, “Artificial Intelligence-Assisted Beam Distribution Imaging Using a Single Multimode Fiber at CERN”, in *Proc. IPAC’22*, Bangkok, Thailand, Jun. 2022, pp. 339–342. doi:10.18429/JACoW-IPAC2022-MOPOPT041

[3] Q. Xu *et al.*, “Estimation of beam transverse parameters through a multimode fiber using deep learning”, in *Proc. IBIC’24*, Beijing, China, Sep. 2024, pp. 170–173. doi:10.18429/JACoW-IBIC2024-TUP54

[4] Q. Xu *et al.*, “Evaluating machine learning models for multimode-fiber-based transverse beam profile reconstruction”, in *Proc. IBIC’25*, Liverpool, UK, Sep. 2025, pp. 373–376. doi:10.18429/JACoW-IBIC2025-TUPCO11

[5] S. Li, C. Saunders, D. J. Lum, J. Murray-Bruce, V. K. Goyal, T. Čižmár, and D. B. Phillips, “Compressively sampling the optical transmission matrix of a multimode fibre”, *Light Sci. Appl.*, vol. 10, Art. no. 88, 2021. doi:10.1038/s41377-021-00514-9

[6] B. Rahmani, D. Loterie, G. Konstantinou, D. Psaltis, and C. Moser, “Multimode optical fiber transmission with a deep learning network”, *Light Sci. Appl.*, vol. 7, Art. no. 69, 2018. doi:10.1038/s41377-018-0074-1

[7] S. M. Popoff, G. Lerosey, R. Carminati, M. Fink, A. C. Boccara, and S. Gigan, “Measuring the transmission matrix in optics: an approach to the study and control of light propagation in disordered media”, *Phys. Rev. Lett.*, vol. 104, no. 10, p. 100601, 2010. doi:10.1103/PhysRevLett.104.100601

[8] Q. Xu, H. D. Zhang, G. Trad, A. Hill, F. Roncarolo, and C. P. Welsch, “Development of radiation-tolerant beam imaging via multimode fiber and synthetic data-driven machine learning”, *Phys. Rev. Accel. Beams*, vol. 29, no. 3, p. 032801, 2026. doi:10.1103/wddb-m37l