ISBN: 978-3-95450-262-2 ISSN: 2673-5350 doi: 10.18429/JACoW-IBIC2025-TUPCO11

EVALUATING MACHINE LEARNING MODELS FOR MULTIMODE-FIBER-BASED TRANSVERSE BEAM PROFILE RECONSTRUCTION

Q. Xu*,[1,2,3], A. Hill[1,2], H. D. Zhang[1,2], F. Roncarolo[3], G. Trad[3], S. Burger[3], W. Farabolini[3], P. Korysko[4], D. Metin[3], C. P. Welsch[1,2]
[1] University of Liverpool, Liverpool, UK
[2] Cockcroft Institute, Daresbury, UK
[3] CERN, Geneva, Switzerland
[4] University of Oxford, Oxford, UK

Abstract

Transverse beam profile monitoring is essential for safe and efficient accelerator operation. In high-radiation environments such as beam dumps, cameras degrade rapidly. To address this, a single multimode fiber (MMF) transmission system was previously tested to transport scintillation light from a screen to a remote camera. Because multiple guided modes are excited and coupled during propagation, the fiber output does not preserve the image and requires reconstruction. This contribution evaluates seven machine-learning reconstruction models for recovering the original transverse beam distribution from MMF output. Using data from the MMF-relayed Chromox screen campaign at CERN's CLEAR facility, the study compares models in terms of reconstruction error, convergence speed, and run-to-run stability, with particular attention to the use of incoherent light. The results indicate robust options for radiation-tolerant, MMF-based transverse diagnostics.

INTRODUCTION

CERN is investigating radiation-resistant approaches for beam imaging to reduce damage to cameras and peripheral electronics [1]. In previous work, a single large-core multimode fiber (MMF) relay was developed to transport light from a Chromox scintillating screen in the beam pipe to a CMOS camera located in a low-dose area [2]. The main difficulty is that propagation modes inside the MMF experience coupling, scattering, and dispersion, which alter the input power distribution [3]. The output intensity is therefore a scrambled mixture rather than an image directly usable for transverse beam parameter extraction. Although fiber-based image reconstruction has been studied, the optimal approach for incoherent, screen-based light remains open. This paper evaluates seven representative reconstruction models—covering image-to-image, generative, and image-to-parameter formulations—using paired data from an MMF-relayed screen campaign. To address practical deployment, we also report model size, training time, and convergence speed, relevant for periodic fine-tuning under drifting MMF conditions (temperature, vibration).

* qiyuan.xu@liverpool.ac.uk

During the campaign, light from a 30 mm × 30 mm Chromox scintillating screen was transmitted through a multimode fiber (FP1500ERT, Ø1.5 mm). Lenses coupled light into and out of the fiber, with a beam splitter sending a fraction to a reference camera for direct screen images. Each dataset entry therefore consists of a paired image (reference screen, MMF output), stored as 256 × 256 single-channel arrays. In total, 6257 pairs were acquired using a triplet scan technique; the main parameters are listed in Table 1.

Table 1: Transverse Beam Image Dataset Collected at CERN

Dataset Configuration	Specifications
Facility	CLEAR
Particle source	e^-
Energy	~ 150 MeV
Screen type	Chromox (Al_2O_3:CrO_2)
Central λ	693 nm
Number of samples	6257

MACHINE LEARNING MODELS AND TRAINING

Four model families are considered for MMF image reconstruction. First, the vectorized models, such as the transmission matrix (TM) [4] and a single-hidden-layer dense neural network (SHL-DNN) [5], operate on flattened input and output images. Their model parameter count scales as $\mathcal{O}((HW)^2)$, where H and W denote the image height and width, so for these models we restrict the resolution: the SHL-DNN input is reshaped to 64×64, and both SHL-DNN and TM outputs are constrained to 32×32.

The convolutional image-to-image models include U-Net [6], a convolutional autoencoder (CAE), and the conditional generative adversarial network Pix2Pix [7], all operating on native 256 × 256 input and output images. These models share an encoder–decoder structure, where the input is progressively compressed into abstract feature representations and then decoded back to an image. U-Net preserves fine detail through skip connections; CAE removes all skip connections and pooling layers, instead using stride-2 convolutions for improved generalization; Pix2Pix extends the U-Net by adding an adversarial discriminator to enhance reconstruction of fine structures.

ISBN: 978-3-95450-262-2 ISSN: 2673-5350 doi: 10.18429/JACoW-IBIC2025-TUPCO11

For the regression model, the encoder–regressor network (ERN) [8] retains only the encoder branch and attaches a multilayer perceptron head to directly predict four transverse beam parameters from the MMF output. Finally, the Swin Transformer (Swin-T) model [9, 10] employs windowed self-attention to capture both local and long-range correlations introduced by MMF scrambling, offering strong global modeling at higher computational cost. All image-based models, including the vectorized ones, are trained with a pixel-wise mean squared error (MSE) loss, whereas the ERN is trained with a beam parameter-wise MSE loss.

All models were trained under identical conditions on a GPU-based Linux HPC using PyTorch. Each model was trained for up to 100 epochs, with early stopping to prevent overfitting, using the Adam optimizer at a learning rate of 1×10^{-4}. Input image pixels were normalized to the range [0, 1]. To assess stability, each model was run three times with different random seeds affecting weight initialization and dataset partition. For a unified evaluation, the final performance metric is the normalized four transverse beam parameters (horizontal and vertical beam centroids and widths) extracted from the reconstructed images using Gaussian fitting. The dataset was split into training, validation, and test sets in an 8:1:1 ratio. Because beam images were acquired sequentially, temporally adjacent frames could be highly similar. Conventional random splits would therefore risk near-duplicate leakage between training and test sets. To prevent this, we applied a time-based split, ensuring the test set has no temporal overlap with training data. The resulting training history for all models is summarized in Fig. 1.

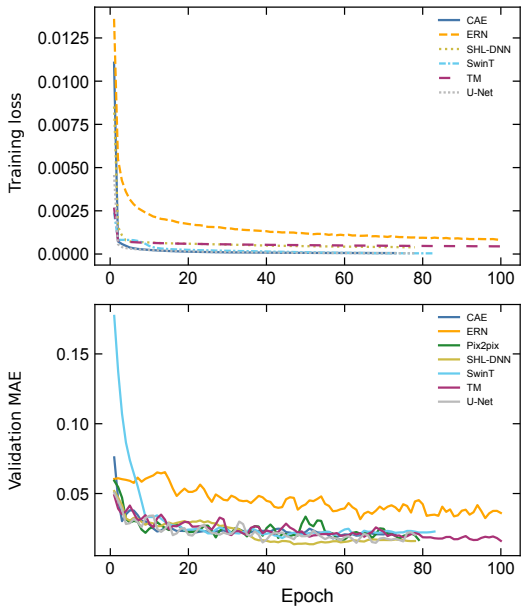

Figure 1: Training history over 100 epochs: (a) training loss and (b) validation set MAE on four beam parameters.

The top plot illustrates the convergence of training loss over epochs, while the bottom plot shows the mean absolute error (MAE) of four beam-parameter predictions on the validation set, where each sample point represents the average of multiple runs. All models converge rapidly within the first 20 epochs. Among the convolutional models, CAE and U-Net achieve very low training loss, whereas Pix2Pix exhibits larger fluctuations, likely due to adversarial training instabilities. The ERN converges more slowly and stabilizes at a higher loss than the others, consistent with its beam parameter-wise loss function.

The model size and training time are summarized in Table 2. TM and SHL-DNN have simple structures but their parameter count scales quadratically with resolution. For example, a full 256×256 image for the TM would require on the order of 17 GB, and the SHL-DNN is even larger. The SHL-DNN design follows the original paper, where it was optimized for coherent light sources, but it appears less suitable for incoherent MMF image reconstruction. By contrast, convolutional models are more compact due to weight sharing in the kernels, and the ERN is the most lightweight model because it omits the decoder branch. Training times are reported as the mean ± standard deviation over three runs with different random seeds, measured until early stopping. The U-Net required the longest training, mainly due to the heavy computation from concatenated skip connections, while the other models remained within a manageable range.

Table 2: Model Size and Total Training Time on the HPC

Model	Size [MB]	Training Time [min]
SHL-DNN	80	17 ± 1
ERN	46	54 ± 25
CAE	71	59 ± 18
TM	256	61 ± 6
Swin-T	150	153 ± 31
Pix2pix	104	155 ± 89
U -Net	83	311 ± 13

TRANSVERSE DISTRIBUTION RECONSTRUCTION

Representative reconstruction samples are shown in Fig. 2. Each row corresponds to one MMF input–output pair, where the first two columns show the MMF output (input to the model) and the corresponding ground-truth beam distribution. These comparisons illustrate how different models balance fidelity to fine details versus recovery of the main structure. For the TM and SHL-DNN approaches, the limited model size constrains the reconstruction quality. While they can reproduce the approximate beam position and general blob shape, the output resolution is lower and background noise remains high, reflecting an averaging effect over many training samples. In contrast, the convolutional models achieve improved performance. They not only recover the beam position but also reproduce shape information with reduced noise. Among them, the CAE provides the best overall quality: its compressed latent representation and simplified architecture allow more generalized reconstruction of both centroid and width. The ERN, which directly predicts beam parameters, is not included in this figure.

ISBN: 978-3-95450-262-2 ISSN: 2673-5350 doi: 10.18429/JACoW-IBIC2025-TUPCO11

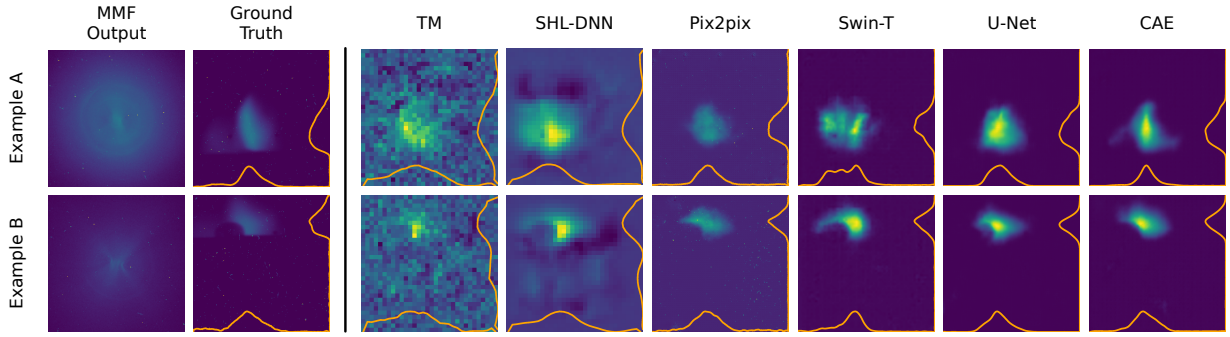

Figure 2: Representative transverse beam samples with their corresponding MMF output reconstructed by all evaluated models, with orange curves showing the transverse beam profiles.

Figure 3 summarizes the overall performance of each model on the test set. For each reconstructed image (or directly from the ERN), four transverse beam parameters were extracted, and the MAE across these parameters was computed. The x-axis lists the models, and the y-axis gives the per-sample MAE histograms. The box represents the interquartile range (Q1–Q3), whiskers extend to the furthest data points within 1.5 × IQR, the orange line indicates the median, and the green triangle shows the mean. A small number of outliers beyond 0.1 are omitted for clarity. Consistent with the qualitative reconstructions, TM and SHL-DNN yield higher errors and broader distributions, indicating larger variability across samples. The convolutional models show lower MAE and tighter ranges, reflecting both improved accuracy and more stable performance.

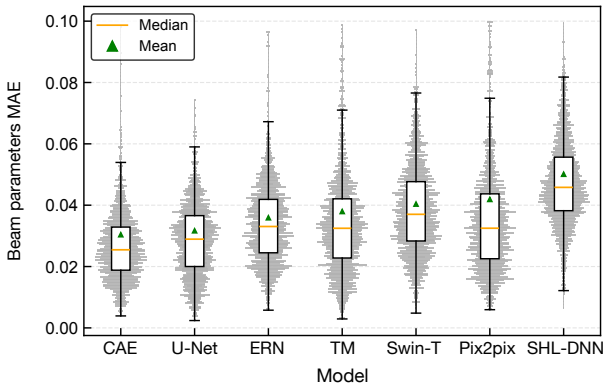

Figure 3: Box plots of MAE distributions and statistical comparison across models.

The final performance scores of all models, measured in MAE, MSE and root mean squared error (RMSE), and evaluated over three runs on the test set, are summarized in Fig. 4. Two horizontal dashed lines indicate the lowest MAE and RMSE values. CAE achieves these minima, U-Net performs comparably well, SHL-DNN yields the highest errors, and the other models (ERN, TM, Swin-T, Pix2Pix) lie in between, confirming CAE as the most reliable approach.

CONCLUSION

This work presented a systematic study of machine-learning-based reconstruction for radiation-tolerant transverse beam imaging using the scintillating screen and a multimode fiber. Unlike most earlier studies on coherent laser sources, this work evaluated model performance with incoherent scintillation light. Seven models were benchmarked by reconstruction accuracy, size, and training time. Lightweight designs such as ERN and CAE are advantageous for practical fine tuning under changing MMF conditions, while convolutional encoder–decoder structures repeatedly outperformed the TM and SHL-DNN baselines. Among them, the CAE offered the best overall balance between accuracy, stability, and computational cost, making it a robust candidate for MMF-based diagnostics.

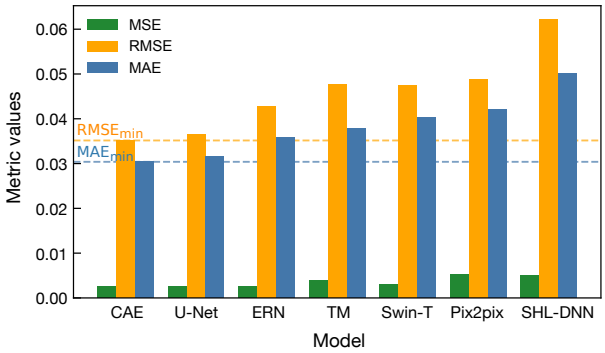

Figure 4: Average test-set performance over models in terms of MSE, RMSE, and MAE on beam parameters prediction.

Future work will address the dataset side of the problem, in particular how to achieve good performance with minimal experimental data. Approaches such as synthetic augmentation, transfer learning, or active sampling may help reduce data requirements while improving generalization under varying MMF conditions. These directions represent a step toward deploying radiation-tolerant imaging systems in high-radiation accelerator environments.

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 also acknowledge the support of the CERN Beam Instrumentation group. Model training was performed on Barkla, part of the High Performance Computing facilities at the University of Liverpool, UK.

ISBN: 978-3-95450-262-2 ISSN: 2673-5350 doi: 10.18429/JACoW-IBIC2025-TUPCO11

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] H. Cao, T. Čižmár, S. Turtaev, T. Tyc, and S. Rotter, "Controlling light propagation in multimode fibers for imaging, spectroscopy, and beyond", *Adv. Opt. Photon.*, vol. 15, no. 2, pp. 524–612, 2023. doi:10.1364/AOP.471055

[4] S. M. Popoff *et al.*, "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

[5] C. Zhu *et al.*, "Image reconstruction through a multimode fiber with a simple neural network architecture", *Sci. Rep.*, vol. 11, no. 896, pp. 1–9, 2021. doi:10.1038/s41598-020-79646-8

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation", in *Proc. MICCAI'15*, Munich, Germany, Oct. 2015, pp. 234–241. doi:10.1007/978-3-319-24574-4_28

[7] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks", in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1125–1134. doi:10.1109/CVPR.2017.632

[8] Q. Xu *et al.*, "Estimation of beam transverse parameters through a multimode fiber using deep learning", in *Proc. IBIC'24*, Kraków, Poland, Sep. 2024, pp. 170–173. doi:10.18429/JACoW-IBIC2024-MOPP30

[9] Z. Liu *et al.*, "Swin transformer: Hierarchical vision transformer using shifted windows", in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, Canada, Oct. 2021, pp. 10012–10022. doi:10.1109/ICCV48922.2021.00986

[10] Y. Chen, B. Song, J. Wu, W. Lin, and W. Huang, "Deep learning for efficiently imaging through the localized speckle field of a multimode fiber", *Appl. Opt.*, vol. 62, no. 2, pp. 266–274, Jan. 2023. doi:10.1364/AO.472864