Bessel Equivariant Networks for Inversion of Transmission Effects in Multi-Mode Optical Fibres

Joshua Mitton*
School of Computing Science
j.mitton.1@research.gla.ac.uk

Simon Peter Mekhail
School of Physics & Astronomy,
Simon.Mekhail@glasgow.ac.uk

Miles Padgett
School of Physics & Astronomy,
Miles.Padgett@glasgow.ac.uk

Daniele Faccio
School of Physics & Astronomy,
Daniele.Faccio@glasgow.ac.uk

Marco Aversa
School of Computing Science
marco.aversa@glasgow.ac.uk

Roderick Murray-Smith
School of Computing Science
roderick.murray-smith@glasgow.ac.uk

University of Glasgow, Glasgow, Scotland, UK.

Abstract

We develop a new type of model for solving the task of inverting the transmission effects of multi-mode optical fibres through the construction of an $\mathrm{SO}^+(2,1)$-equivariant neural network. This model takes advantage of the of the azimuthal correlations known to exist in fibre speckle patterns and naturally accounts for the difference in spatial arrangement between input and speckle patterns. In addition, we use a second post-processing network to remove circular artifacts, fill gaps, and sharpen the images, which is required due to the nature of optical fibre transmission. This two stage approach allows for the inspection of the predicted images produced by the more robust physically motivated equivariant model, which could be useful in a safety-critical application, or by the output of both models, which produces high quality images. Further, this model can scale to previously unachievable resolutions of imaging with multi-mode optical fibres and is demonstrated on 256×256 pixel images. This is a result of improving the trainable parameter requirement from $\mathcal{O}(N^4)$ to $\mathcal{O}(m)$, where N is pixel size and m is number of fibre modes. Finally, this model generalises to new images, outside of the set of training data classes, better than previous models.

1 Introduction

Multi-mode fibres (MMF) have many potential applications in medical imaging, cryptography, and communications. In the medical domain, the use of multi-mode fibre imaging has potential to create hair-thin endoscopes for imaging sensitive areas of the body. However, to achieve these applications, the fibre transmission properties must be compensated for to return a clear image (Stasio, 2017). A MMF has multiple different fibre modes, each of which propagates at a different velocity. This leads to an amplitude and phase mixing of the image as it propagates through the fibre (Mitschke, 2016). As a result, an input image creates a complex-valued speckled pattern on the output of the MMF. The ability to accurately and in a scalable way learn to invert the transmission effects would unlock MMF imaging as a useful tool across a range of domains. This work concerns the use of a single multi-mode fibre and not fibre bundles.

*https://github.com/JoshuaMitton

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Inverting a speckled image is challenging for multiple reasons. Firstly, the speckled images have a non-local relationship with respect to the original images. As a result, solely local patch-based models, such as convolutional neural networks, do not make sense as a solution without some dense mapping function. Therefore, the non-locality necessitates mapping the speckled images into a spatial arrangement similar to the original images before typical image-based deep learning techniques can be used, such as convolutions and pooling. In addition, the speckled images have circular correlation, which could be taken advantage of, although as noted by Moran et al. (2018), finding these requires solving the inversion, creating a chicken-and-egg problem. Finally, the fibre is equivalent to an unknown complex transmission matrix (TM) so, the inverse of this could be found using a complex-valued linear model, although this presents challenges in terms of memory requirements. A mapping between 350×350 original and speckled images would result in a TM with $350^4 \approx 15$ billion entries, requiring a linear model with as many parameters.

Previous work in inverting the transmission effects of MMFs has either required extensive experimentation to characterise the TM of the fibre (Čižmár & Dholakia, 2011, 2012; Choi et al., 2012; Mahalati et al., 2013; Papadopoulos et al., 2012; Plöschner et al., 2015; Leite et al., 2021), where the number of experimental measurements required for re-calibration was reduced by Li et al. (2021) by exploiting sparstiy in the TM; made use of dense linear models (Moran et al., 2018; Fan et al., 2019; Caramazza et al., 2019); or made use of convolutional models (Borhani et al., 2018; Rahmani et al., 2018). For the machine learning approaches to tackling the inversion task, those which make use of a dense linear model (Moran et al., 2018; Fan et al., 2019; Caramazza et al., 2019) can naturally account for the difference in spatial arrangement between speckled and original images, although they scale badly with the resolution of the images considered ($\mathcal{O}(N^4)$ for $N \times N$ resolution). On the other hand, in theory, the convolutional neural network (CNN) models (Borhani et al., 2018; Rahmani et al., 2018) improve upon the scalability issue, but in practice due to the need to approximate the transmission matrix and its non-local effects, the models require a large number of layers in order to be able to effectively map every pixel in the speckled image to every pixel in the original image, and hence in practice do not over come the scalability issues. All of these approaches are mostly expected to work for classes of objects that belong to the class that was used for training (Borhani et al., 2018), with Rahmani et al. (2018) making initial steps towards general imaging and Caramazza et al. (2019) demonstrating this for a more diverse testing dataset. We provide further details on each of the previous methods in Appendix A.2.

In this work we present a model which naturally accounts for the difference in spatial arrangement between speckled and original images and scales more efficiently than previous methods to higher resolution images. We believe this is the first method to demonstrate an ability to invert 256×256 pixel speckled images into 256×256 pixel original images. Our approach also takes advantage of the circular correlations in the speckled images, and improves upon previous general imaging results. Concerning the equivariance literature, we develop a model comprising of cylindrical harmonic basis functions, a basis set which has seen little attention in the equivariance literature, and make the connection between the transmission of light through a fibre and the group theoretic understanding used in developing equivariant neural networks. Our contributions are:

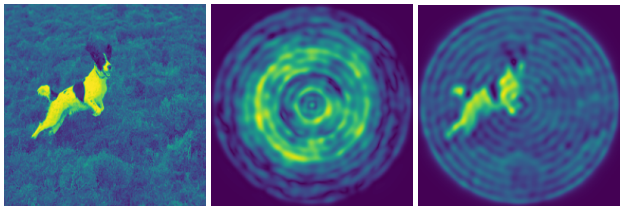
1. A more data-efficient, scalable model to solve the inversion of MMF transmission effects.
2. A model that provides better generalisation to out-of-training domain images.
3. A connection between group theoretic equivariant neural networks and the inversion of MMF transmission effects, providing a new type of model to tackle the problem.

2 Background

2.1 Multi-Mode Fibres

Multi-mode fibres present a clear advantage over single-mode fibre bundles due to having 1-2 orders of magnitude greater density of modes than a fibre bundle (Choi et al., 2012). Despite this, the different propagation velocities of each mode, which result in the fibre producing scrambled images, presents a significant challenge. If each propagation mode and velocity was known the TM could be computed, providing a linear system that inverts the transmission effects, although in general this is not known. Further, mode specific losses and imperfect mapping between the input pixels and fibre modes can lead to information loss such that inverting the transmission does

not yield the true original input image, as is demonstrated in Figure 1. Therefore, a model which correctly models the physics of inverting the transmission effects should produce the inverted image in Figure 1 and producing the original image requires addition information to be learned. We discuss the propagation of light through optical fibres and how we construct theoretical TMs in Section 2.2 and further details on the inversion of the TM in Appendix A.4.



(a) Original (b) Speckled (c) Inverted

Figure 1: (a) Original image. (b) Speckled image from passing original images through a theoretical TM. (c) Inverted image created by passing speckled images through the inverted theoretical TM.

2.2 Generation of Theoretical Transmission Matrices

Light propagation through a MMF is characterised by the transmission modes of the fibre. The fibres considered in this work have a step index profile with the refractive index within the core being constant and at a higher value than the cladding. Analytical solutions for the fibre modes can be determined by solving the Helmholtz equation in cylindrical coordinates (details in Appendix A.1). The problem can therefore be expressed as an eigenvalue eigenfunction problem, where the eigenfunctions are the fibre modes and the eigenvalues are the propagation constants, β, of the modes. The solutions for the electric field within the fibre are comprised of Bessel functions of the first kind inside the core and modified Bessel functions of the second kind in the cladding as follows:

$$f_l^{core}(r,\theta) = \frac{J_l(u(\beta)\cdot\frac{r}{R})}{J_l(u(\beta))}e^{\pm il\theta}, \qquad f_l^{clad}(r,\theta) = \frac{K_l(w(\beta)\cdot\frac{r}{R})}{K_l(w(\beta))}e^{\pm il\theta}, \tag{1}$$

where r and θ are the radial and azimuthal coordinates, respectively, l is the azimuthal index of the mode, R is the fibre core radius, and u and w are normalised frequencies defined as follows:

$$u(\beta) = R\sqrt{k_0^2 n_{core}^2 - \beta_{lm}^2}, \qquad w(\beta) = R\sqrt{\beta_{lm}^2 - k_0^2 n_{clad}^2}, \tag{2}$$

where n_{core} and n_{clad} are the refractive indices of the core and cladding, respectively, k_0 is the vacuum wave number, and m is the radial mode index. Further details on the connection between propagation of light through MMFs and group theoretic equivariant neural networks are provided in Appendix A.1. Taking the derivative of Equations 1, and equating the resulting functions asserts a smoothness condition on the electric field across the core-cladding boundary. Solving for this for all values of β over all possible integer values of l gives the propagation constants. These constants correspond to the fibre modes given by the now appropriately parameterised Equations 1. Examples of the mode fields are given in Figure 2.

A diagonal fibre propagation matrix, F, can then be made from the eigenvalues, β, of the Helmholtz problem and the corresponding eigenfunctions can be recorded in a matrix, M, which maps the image space to the fibre mode, or group space. We can find the inverse mapping bases from the fibre or group space back to the image space, $M^\dagger$, by taking the conjugate transpose. These three components allow us to construct the transmission matrix as: $\mathrm{TM} = MFM^\dagger$.

2.3 Equivariance

Equivariance in neural networks concerns the study of symmetries and uses these to build an inductive bias into a model (Cohen & Welling, 2016a; Bronstein et al., 2021). Equivariance as a property of a neural network guarantees that a transformation of the input data produces a predictable transformation of the predicted features (Worrall et al., 2017). Formally, we say that, given two transformations T and T' which are linear representations of a group G and a network or layer Φ, the network is equivariant if applying transform T_g to an input x and then passing it through the

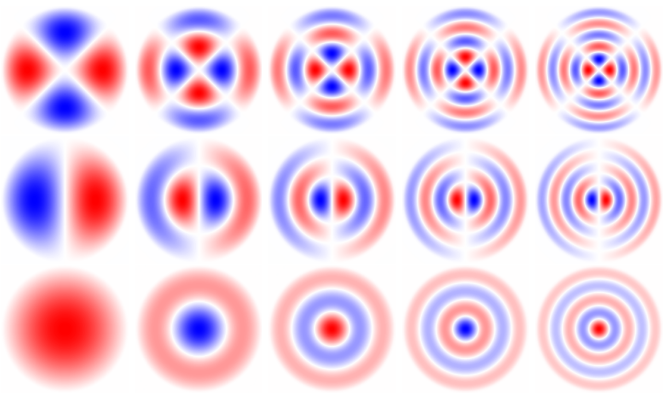


Figure 2: Electric field amplitude profiles of the Bessel bases used within our model. Here red is positive, blue is negative and the colour saturation represents the electric field strength. These bases are often known as LP modes.

network Φ yields the same result as first passing x through Φ and then transforming by T'_g (Cohen & Welling, 2016a); that is,

$$\Phi T_g(x) = T'_g(\Phi(x)). \tag{3}$$

Convolutional neural networks (CNNs) are an early example of equivariance on images, where by design they are translationally equivariant (LeCun et al., 1998). More recently, equivariance has been considered for other group actions than that of the translation group. The majority of works on images involve incorporating rotation or reflection equivariance into the model, where a variety of different group transformations have been considered (Cohen & Welling, 2016a,b; Weiler & Cesa, 2019; Murugan et al., 2019; Wiersma et al., 2020; De Haan et al., 2020; Franzen & Wand, 2021).

We are not interested in developing a group equivariant convolutional model here due to the non-similar spatial arrangements between the speckled and original image domains. Despite this, the concept of equivariance still applies, as the concept of learning a model from a fixed basis set, which guarantees symmetry properties are conserved, is relevant due to the nature of transmission through multi-mode optical fibres.

2.4 Circular Harmonics

Of the works that incorporate rotation equivariance into CNNs we are particularly interested in those using a continuous rotation group $\mathrm{SO}(2)$. In the case of discrete rotation groups a different feature space is associated to each rotation angle. Storing features for an infinite number of rotation angles is not computationally tractable. One option is to turn to a Fourier representation, whereby instead of choosing a discrete number of rotations a maximum frequency can be chosen. To use a Fourier representation for the weights of the model the domain $\mathbb{R}^2$ is decomposed into a radial profile and angular function. Solving for the kernel space of permissible filters that can be used within a CNN and ensuring that the model maintains rotation equivariance yields only the spectrally localised circular harmonics (Worrall et al., 2017; Weiler & Cesa, 2019; Franzen & Wand, 2021). Therefore a rotation equivariant CNN can be created by solving for the circular harmonics up to a certain rotational frequency, combining this with a radial profile function, and sampling a basis of resolution given by the CNN filter size. This yields a set of bases which can be linearly combined with a learnable weighting applied to each to form the kernel for the CNN.

2.4.1 Cylindrical Harmonics

While the circular harmonics have seen some attention in the deep learning community due to the use in constructing rotational equivariant CNNs, the cylindrical harmonics have seen less attention (Klicpera et al., 2020). The cylindrical harmonics appear as a solution to Bessel functions for integer α and are therefore of interest for problems where information transformation is characterised by such functions. For example Bessel functions are used when solving wave or heat propagation.

Bessel functions are solutions for different complex numbers α of Bessel's differential equation:

$$x^2 \frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2 - \alpha^2)y = 0. \tag{4}$$

For integer values of α the Bessel function solutions are a linearly independent set of functions expressed in cylindrical coordinates. Each function consists of the product of three functions. The radially dependent term is typically called the cylindrical harmonics. Further details on the connection between group theory and the Bessel basis functions is provided in Appendix A.1.

3 Our Method

Our method comprises two models: a Bessel equivariant model and a convolutional post-processing model. The Bessel equivariant model uses a basis set comprising cylindrical harmonics and learns a complex mapping function to use the known symmetries in optical fibres. The Bessel equivariant model is constrained by the basis choice to produce circular images, which can have circular artifacts also seen in a true inversion, see Figure 1. The post-processing model is used to fill gaps in the corners, remove circular artifacts, and sharpen the images. The two stage approach is useful in safety-critical applications as users can inspect the output of both models, where the Bessel equivariant model output is less likely to depend on the training data due to the choice of inductive bias.

3.1 Bessel Basis Equivariant Model

The core of our model has an inductive bias that better replicates the physics of the task of inverting the transmission effects of MMF imaging. To achieve this we utilise prior knowledge of MMFs and the modes with which an image propagates through the fibre. Similar to the successes of rotation equivariant neural networks, which make use of circular harmonics to achieve rotation equivariance, we utilise cylindrical harmonics as an inductive bias of the model. This is motivated by the propagation modes of the fibre being characterised by Bessel functions, where the radially dependent solution is given by the cylindrical harmonics. We make the connection to the group theoretic development of equivariant neural networks in Appendix A.1, demonstrating the connection between the group $\mathrm{SO}^+(2,1)$, the light cone, and Bessel function solution to the wave equation.

The first stage of the model is the computation of two sets of basis functions. The first set transforms the input image into a function that lives on the group and the second transforms back into the image domain. The model's weight matrix linearly transforms the function mapped by the first basis set. The basis filters are computed in accordance with equations in Appendix 2.2.

These functions are sampled on a grid given by the size of the speckled images for the first basis set and the original images for the second basis set. An example of the bases produced is given in Figure 2. The Bessel function bases are computed in a pre-training stage and therefore this step only has to be completed once when creating the model. Further, these bases are not trainable parameters and are not updated during training of the model. These basis functions are an alternative to those used in general rotation equivariant neural networks, which typically offer equivariance to the group $\mathrm{SO}(2)$, except they correctly model the symmetry group, $\mathrm{SO}^+(2,1)$ of the task of inverting transmission effects of a MMF due to the added time dimension.

The model trainable parameters are complex-valued weights of size equal to the number of bases. The complex weight matrix is a diagonal matrix. The images transformed by the first basis set are multiplied by the weight matrix before being transformed by the second basis set to produce the predicted image. The overall model architecture is given in Figure 3.

3.2 Post-Processing Model

As demonstrated in Section A.4 the limited number of modes means that the transmission effects of a MMF are typically not fully invertible leading to information loss. We would therefore not expect an equivariant model informed by the physics of the task to be able to fully invert the transmission effects. To overcome this we add a second model which is similar to a super-resolution model, except that it does not change the resolution and instead learns to remove circular artifacts, fill in gaps, and sharpen the images. It is a fully convolutional model comprising of convolutional layers and non-linearities. This model takes as input the output of our Bessel equivariant model, which we suspect

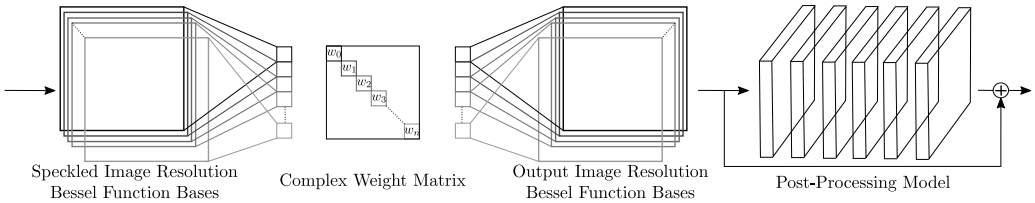


Figure 3: Our two stage model architecture comprising of a Bessel equivariant model and post-processing model. Input (left) is speckled images. Output (right) is predicted image. The speckled image resolution Bessel function bases transform the speckled image into a mode space of the group $\mathrm{SO}^{+}(2,1)$. The complex weight matrix is diagonal. The original image resolution Bessel function basis provides a mapping from the mode space to the output image space. The trainable parameters of the Bessel equivariant model are the diagonal complex weight matrix only. The post-processing model is a convolutional model.

will be similar to what is achievable by passing the speckled images through an inverted TM, and predicts the original images. This model is composed of eight convolutional layers with ReLU non-linearities. The motivation behind the explicit separation of the inversion and enhancement models is that in some applications, especially safety-critical applications, it is important to be able to see the results of an inversion conditioned on the observations alone, regardless of prior expectations.

4 Experiments

We investigate performance with both data generated with a theoretical TM and experimental data collected using a real MMF fibre. We provide details of the theoretical TM generation in Section 2.2. We use the data from a real fibre, as described in Moran et al. (2018), to validate that our new approach can model the same data, despite using significantly fewer trainable model parameters. Further, we generate data using a theoretical TM as this allows us to flexibly adapt the parameters of the experiment and create new datasets to validate the new model design. Finally, the theoretical TM allows us to experiment with images with higher resolution than has been previously been considered. We provide details on the training of the models and datasets in Appendix A.3.

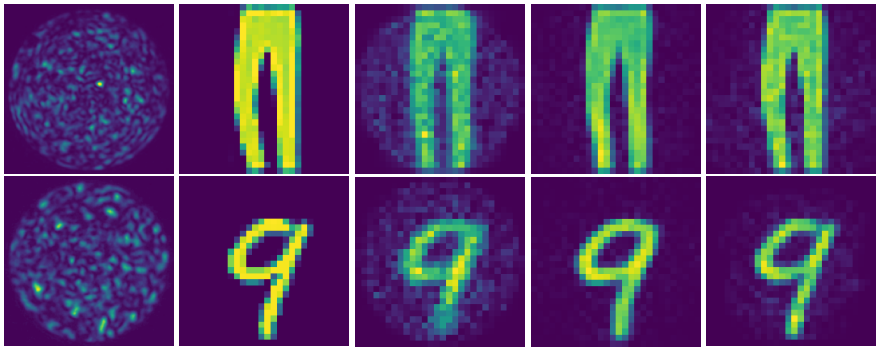
4.1 Real Multimode Fibre

We compare our model to the complex-valued linear model developed by Moran et al. (2018) by considering both MNIST and fMNIST data. For this we train and test separate models for the MNIST and fMNIST datasets. In addition, we compare multiple different versions of our model where the diagonal restriction of the mapping between bases is relaxed to allow for a block diagonal structure. We do this to account for manufacturing defects, sharp bends, dopant diffusion, elliptical cores which causes the diagonal mapping of the fibre matrix within the TM to be block diagonal (Carpenter et al., 2014). This relaxation allows for x-offsets above and below the main diagonal.

Table 1 shows that our model with a full mapping matrix between bases, i.e. full relaxation of the diagonal constraint, outperforms all other models. In addition, with a relaxation to allow for a 10 block diagonal structure our model performs comparably. On the other hand, the complex linear model outperforms our Bessel equivariant model when the diagonal restriction on the mapping function between bases is enforced. Despite this, our model still provides clear and accurate reconstructions of

Table 1: Comparison of the loss values of each model trained with MNIST or fMNIST data.

	MNIST		fMNIST	
Model	Train Loss	Test Loss	Train Loss	Test Loss
Complex Linear	0.00396	0.00684	0.00509	0.01061
Bessel Equivariant Diag	0.03004	0.03125	0.02903	0.03140
Bessel Equivariant Diag + Post Proc	0.01317	0.01453	0.01609	0.01749
Bessel Equivariant 10 Off Diag + Post Proc	0.00488	0.00576	0.00943	0.01162
Bessel Equivariant Full	0.00300	0.00684	0.00548	0.01380
Bessel Equivariant Full + Post Proc	**0.00145**	**0.00378**	**0.00306**	**0.00964**



(a) Input (b) Target (c) BEM (d) BEM+PP (e) Complex

Figure 4: Comparison of predicted images from inverting transmission effects of a MMF. The upper row is fMNIST data, lower row MNIST data. (a) Input speckled image, (b) the target original image to reconstruct, (c) Output of the Bessel equivariant model, (d) Output of the combination of Bessel equivariant and post-processing model, and (e) the output of the complex-valued linear model.

the target images whilst using orders of magnitude fewer trainable parameters. Additional results are presented in Tables 6 and 7. Further to a comparison of the loss values, we visually inspect the reconstruction quality in Figure 4, and we provide further visualisations in Figures 16 and 17. Thus despite our Bessel equivariant diagonal model achieving a larger loss value, the digit or item is still clearly visible in the prediction and is correctly sharpened to a realistic prediction of the original image by our post-processing model. The complex linear model can also be seen to predict a realistic looking output of the correct digit and object, although some noise does feature in the prediction. In addition, we visually compare each of our Bessel equivariant models with the diagonal restriction relaxed in Figures 10 and 11, which demonstrates that as the diagonal restriction is relaxed the model can better predict the original images with less noise in the prediction from the Bessel equivariant model.

We also compare the number of trainable parameters within the model in Table 2. This highlights that our model requires two orders of magnitude fewer parameters that the complex linear model when using the diagonal Bessel equivariant model. Therefore, our model has the potential to scale to higher resolution images where the complex linear model will run out of GPU memory. As the diagonal mapping between Bessel bases is relaxed to include 10-block diagonal structure the model still requires two order of magnitude fewer parameters. In the most flexible version of our model the number of trainable parameters is comparable with the complex Linear model, although we have demonstrated that this level of flexibility is not required to achieve good image reconstruction.

Table 2: Comparison of # trainable parameters in each model trained with MNIST or fMNIST data.

Model	Number of Trainable Parameters (Millions)
Complex Linear	78.826
Bessel Equivariant Diag + Post Proc	0.500
Bessel Equivariant 10 Off Diag + Post Proc	0.617
Bessel Equivariant Full + Post Proc	43.108

4.2 Theoretical TM

We compare our model to the complex linear model developed by Moran et al. (2018) with data created using a theoretical TM. We create multiple datasets one using 28×28 fMNIST images, with 180×180 speckled images, another with MNIST images, of the same size, and one with a subset of images from the ImageNet dataset (Deng et al., 2009) where we fix the resolution to be 256×256 of both the images and speckled images. Using fMNIST and MNIST with a theory TM is useful for developing understanding, due to the ease of creating different experiments and datasets, and for testing the generalisability of the model between the two datasets. We include the ImageNet-based dataset as these are higher resolution images containing more challenging information than fMNIST. To the best of our knowledge, demonstrating the ability to invert transmission effects of such high resolution images has not been previously achievable with a machine learning based approach.

Table 3: Comparison of the loss values of each model trained with fMNIST data.

Model	fMNIST Train Loss	fMNIST Test Loss	MNIST Test Loss
Complex Linear	0.0149	0.0146	0.0363
Bessel Equivariant	0.0141	0.0139	**0.0026**
Bessel Equivariant + Post Proc	**0.0032**	**0.0032**	0.0028

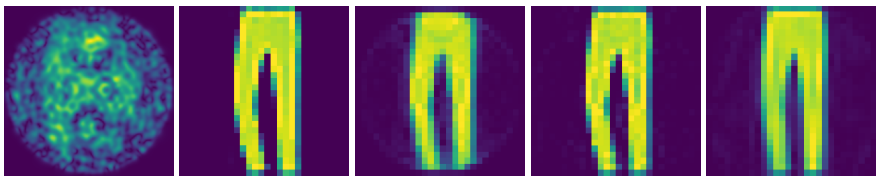
4.2.1 Generalisability of Models

We provide a comparison of the predictions of a complex valued linear model (Moran et al., 2018), our equivariant model only, and our equivariant model coupled with a post-processing model trained on fMNIST images and tested on both fMNIST and MNIST to analyse to what extent the models can generalise to a new dataset. Table 3 presents the loss values for both the training and testing of each model. Considering the fMNIST data this shows that the complex linear and Bessel equivariant models achieve similar loss values. Although this is a result of the Bessel equivariant model under-performing due to not being able to reconstruct pixel values close to the edge of the image as a consequence of the circular nature of the Bessel bases functions, while the complex linear model under-performs due to failing to reconstruct higher frequency information and over-fitting to the general clothing categories. Finally, when our Bessel equivariant model is combined with the post-processing model it outperforms all other models. The ability to generalise to a new data is assessed through the MNIST dataset. This shows that our Bessel equivariant model significantly out-performs the complex linear model in generalising to a new dataset.

Figure 5 shows reconstructions produced by the three models, with further reconstructions provided in Figure 18. This demonstrates visually that our model, even without the post-processing part, is able to better reconstruct the original images. The post-processing model refines the output of the equivariant model and fills in gaps outside the fibre. Despite it being possible to estimate the general category of clothing from the outputs of the complex linear model, the higher frequency information has been lost, i.e. the pattern on the jumper or the curve in the trouser leg. On the other hand, our model is able to reconstruct this higher frequency information.

Previous works have demonstrated some ability to generalise to new data classes (Rahmani et al., 2018; Caramazza et al., 2019), although in each case the results are not perfect. Here, we compare the different methods trained on fMNIST and then tested on MNIST data. Figure 6 demonstrates that our Bessel equivariant model generalises well to a new data domain, with further reconstructions provided in Figure 19. On the other hand, the complex linear model somewhat predicts the correct digits. In this out of training domain our post-processing model does not add any value to the original Bessel equivariant model. This highlights the benefit of our two-stage modelling approach as we have one robust model and a second model to sharpen the images, as a result if a user believes the situation could be unusual they could reliably use the output of our Bessel equivariant model and not the post-processing part. Further to achieving strong generalisation results, we also consider the effect of noise in the speckled images in Appendix A.6.

Figure 8 shows how the number of trainable model parameters scale. Even for 28×28 images our model requires two orders of magnitude fewer trainable parameters than the complex linear model.



(a) Input (b) Target (c) BEM (d) BEM + PP (e) Complex

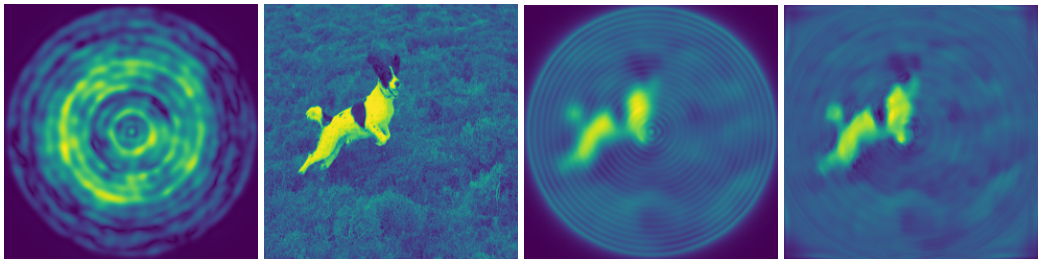
Figure 5: Comparison of predicted images from inverting transmission effects of a MMF. (a) Input speckled image, (b) Target original image to reconstruct, (c) Output of the Bessel equivariant model, (d) Output of the Bessel equivariant and post-processing model, and (e) Output of the complex valued linear model.

(a) Input (b) Target (c) BEM (d) BEM + PP (e) Complex

Figure 6: Comparison of predicted images from inverting transmission effects of a MMF. (a) Input speckled images, (b) Target original images, (c) Output of the Bessel equivariant model, (d) Output of Bessel equivariant and post-processing model, and (e) Output of the complex valued linear model.

4.2.2 Scaling to Larger Images – ImageNet

We now experiment with a subset of images from the ImageNet dataset, where we fix the resolution to be 256×256. Inverting the transmission effects of such higher resolution images has not been previously achievable. We only compare our equivariant model with our equivariant model coupled with a fully convolutional post-processing model. The complex-valued linear model cannot fit in the GPU memory of an A6000 with 48.7GiB of memory with this resolution of image.



(a) Input (b) Target (c) BEM (d) BEM + PP

Figure 7: Comparison of predicted images from inverting transmission effects of a MMF using high resolution ImageNet data. (a) Input speckled image, (b) Target original image to reconstruct, (c) Output of Bessel equivariant model (d) Output of Bessel equivariant and post-processing model.

We analyse our model visually, with Figure 7 showing that our Bessel equivariant model produces a reconstruction from which one can determine the type of dog and its activity, although the reconstruction does not capture all the high frequency information, and information in the corners is lost due to the circular nature of the fibre. Further reconstructions are provided in Figure 20. When we combine the two models some of the artifacts of the Bessel equivariant model are removed as the post-processing model fills in information towards the corners and sharpens the image.

We present the loss values for both models in Table 4. This shows a similar result as for fMNIST, that our Bessel equivariant model can solve the task well, but requires the post-processing model to sharpen the image and fill in gaps in the corners. We compare the models' memory requirements in Figure 8. Note how fewer trainable parameters translate into significantly less memory use than the complex linear model. This effect is seen more drastically when scaling to high resolution images, where the complex linear model runs out of memory on a 24.2GiB Titan RTX for original and speckled images of resolution 180×180 pixels. Our model, in contrast, has been tested on 256×256 pixel images, requiring only 2.1GiB, and can scale to megapixel images.

Finally, we explore the impact on each model of reducing the size of the training dataset. Requiring a reduced dataset size minimises the time taken in a lab collecting data. Figure 8 demonstrates that our model outperforms the complex linear model with $10\times$-less training data, reducing the need to collect 12000 samples to 1200. The reconstructions are visualised in Appendix A.12.

Table 4: Comparison of the loss values of each model trained with ImageNet data.

Model	Train Loss	Test Loss
Bessel Equivariant	0.0541	0.0574
Bessel Equivariant + Post Proc	**0.0161**	**0.0159**

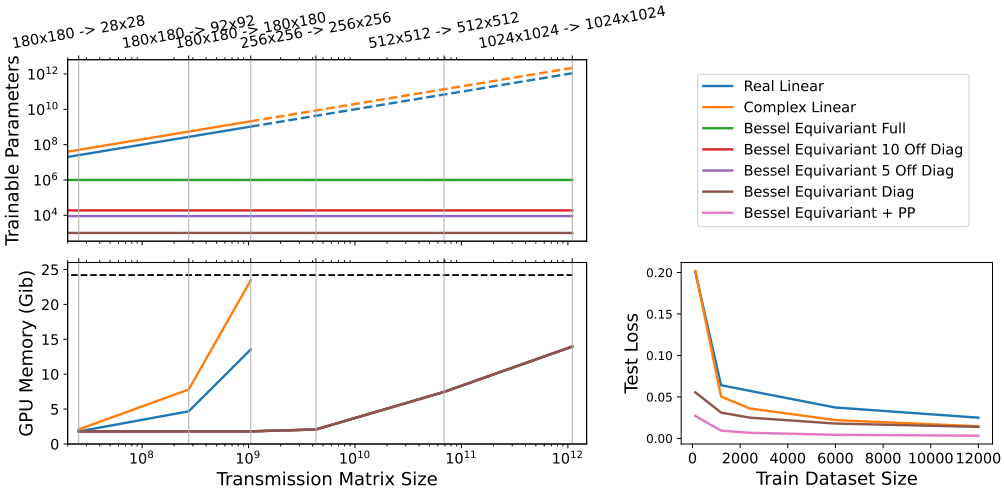


Figure 8: (Upper Left) Comparison of number of trainable parameters v. TM size (dashed line indicates model cannot in practice be built due to large memory requirements). (Lower Left) Comparison of required GPU memory against TM size. Vertical grey lines indicate a given image size – our model can process 1024×1024 images on a single consumer-level Titan RTX GPU. (Lower Right) Comparison of reducing the size of the training dataset. All plots for a 1000 mode fibre.

5 Conclusions

We present a new type of model for solving the task of inverting the transmission effects of a MMF through developing a $\mathrm{SO}^+(2,1)$-equivariant neural network and combining this with a post-processing network. This improves upon complex linear models through incorporating a useful physically informed inductive bias into the equivariant network. The equivariant network is shown to perform well on new image classes, providing a general model for fibre inversion. The post-processing network is specific to an image domain, as it generalises to new image classes as well as a general convolutional network, although it improves the quality of returned images. The use of a theoretical TM allows us to compare an 'ideal' inverse with the output of previous models based on learning a full transmission matrix. This suggests that previous learned transmission matrices were combining elements of the actual transmission matrix with elements of the post-processing network, which would affect their ability to generalise to new images. We also anticipate that in interactive safety-critical applications users might want the ability to switch between modes, to be sure that the evidence was there, and not overly influenced by the priors in the training data.

This new model significantly improves the ability to scale to higher resolution images by improving the scaling law from $\mathcal{O}(N^4)$ to $\mathcal{O}(m)$, where N is pixel size and m is the number of fibre modes. We provide a comparison between our model and prior works on both data created using a real fibre and a theoretical TM, demonstrating in both cases that our model solves the task while using significantly fewer trainable parameters than the complex and real linear models. In addition, we demonstrate results on high resolution 256×256 images, which has previously been unachievable due to the growth of parameters with previous models. Furthermore, we demonstrate the ability of our model to generalise to new data classes outside of the training data, outperforming prior works. The dramatic reduction in the number of parameters for each fibre configuration opens the way for future models which can learn mappings for high-resolution images, from a wider set of perturbed fibre poses and combine these using architectures such as VAEs.

Acknowledgments and Disclosure of Funding

JM is supported by a University of Glasgow Lord Kelvin Adam Smith Studentship. RM-S, MP, DF and SPM are grateful for EPSRC support through grants EP/T00097X/1 and RM-S for EP/R018634/1. MA is funded by dotPhoton and a UofG scholarship. DF acknowledges funding through the Royal Academy of Engineering Chair in Emerging Technologies programme.

References

Navid Borhani, Eirini Kakkava, Christophe Moser, and Demetri Psaltis. Learning to see through multimode fibers. *Optica*, 5(8):960–966, 2018.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Piergiorgio Caramazza, Oisín Moran, Roderick Murray-Smith, and Daniele Faccio. Transmission of natural scene images through a multimode fibre. *Nature communications*, 10(1):1–6, 2019.

Joel Carpenter, Benjamin J Eggleton, and Jochen Schröder. Maximally efficient imaging through multimode fiber. In *CLEO: Science and Innovations*, pp. STh1H–3. Optical Society of America, 2014.

Youngwoon Choi, Changhyeong Yoon, Moonseok Kim, Taeseok Daniel Yang, Christopher Fang-Yen, Ramachandra R Dasari, Kyoung Jin Lee, and Wonshik Choi. Scanner-free and wide-field endoscopic imaging by using a single multimode optical fiber. *Physical review letters*, 109(20): 203901, 2012.

Tomáš Čižmár and Kishan Dholakia. Shaping the light transmission through a multimode optical fibre: complex transformation analysis and applications in biophotonics. *Optics express*, 19(20): 18871–18884, 2011.

Tomáš Čižmár and Kishan Dholakia. Exploiting multimode waveguides for pure fibre-based imaging. *Nature communications*, 3(1):1–9, 2012.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016a.

Taco S. Cohen and Max Welling. Steerable CNNs. *arXiv preprint arXiv:1612.08498*, 2016b.

Pim De Haan, Maurice Weiler, Taco Cohen, and Max Welling. Gauge equivariant mesh CNNs: Anisotropic convolutions on geometric graphs. *arXiv preprint arXiv:2003.05425*, 2020.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Pengfei Fan, Tianrui Zhao, and Lei Su. Deep learning the high variability and randomness inside multimode fibers. *Optics express*, 27(15):20241–20258, 2019.

Daniel Franzen and Michael Wand. General nonlinearities in SO(2)-Equivariant CNNs. *Advances in Neural Information Processing Systems*, 34, 2021.

Johannes Klicpera, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. *arXiv preprint arXiv:2003.03123*, 2020.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Ivo T. Leite, Sergey Turtaev, Dirk E. Boonzajer Flaes, and Tomáš Čižmár. Observing distant objects with a multimode fiber-based holographic endoscope. *APL Photonics*, 6(3):036112, 2021.

Shuhui Li, Charles Saunders, Daniel J Lum, John Murray-Bruce, Vivek K Goyal, Tomáš Čižmár, and David B Phillips. Compressively sampling the optical transmission matrix of a multimode fibre. *Light: Science & Applications*, 10(1):1–15, 2021.

Reza Nasiri Mahalati, Ruo Yu Gu, and Joseph M Kahn. Resolution limits for imaging through multi-mode fiber. *Optics express*, 21(2):1656–1668, 2013.

Fedor Mitschke. *Fiber optics*. Springer, 2016.

Oisín Moran, Piergiorgio Caramazza, Daniele Faccio, and Roderick Murray-Smith. Deep, complex, invertible networks for inversion of transmission effects in multimode optical fibres. *Advances in Neural Information Processing Systems*, 31, 2018.

Muthuvel Murugan, KV Subrahmanyam, and Siruseri IT Park. SO (2)-equivariance in neural networks using tensor nonlinearity. *British Machine Vision Conference*, 2019.

Ioannis N. Papadopoulos, Salma Farahi, Christophe Moser, and Demetri Psaltis. Focusing and scanning light through a multimode optical fiber using digital phase conjugation. *Optics express*, 20(10):10583–10590, 2012.

Martin Plöschner, Tomáš Tyc, and Tomáš Čižmár. Seeing through chaos in multimode fibres. *Nature Photonics*, 9(8):529–535, 2015.

Babak Rahmani, Damien Loterie, Georgia Konstantinou, Demetri Psaltis, and Christophe Moser. Multimode optical fiber transmission with a deep learning network. *Light: Science & Applications*, 7(1):1–11, 2018.

Nicolino Stasio. Multimode fiber optical imaging using wavefront control. Technical report, EPFL, 2017.

Maurice Weiler and Gabriele Cesa. General E(2)-equivariant steerable CNNs. *Advances in Neural Information Processing Systems*, 32, 2019.

Ruben Wiersma, Elmar Eisemann, and Klaus Hildebrandt. CNNs on surfaces using rotation-equivariant features. *ACM Transactions on Graphics (TOG)*, 39(4):92–1, 2020.

M.K.F. Wong. Unitary representations of $SO(n, 1)$. *Journal of Mathematical Physics*, 15(1):25–30, 1974.

Daniel E Worrall, Stephan J Garbin, Daniyar Turmukhambetov, and Gabriel J Brostow. Harmonic networks: Deep translation and rotation equivariance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5028–5037, 2017.

Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] See Sections 3.2 and A.4.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] See Section A.14.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] See Sections 2, 2.2, and A.4.
   (b) Did you include complete proofs of all theoretical results? [Yes] See Sections A.4 and A.1.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Section A.3.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No] Due to compute constraints we did not run with multiple random seeds, although we do note that training was very stable and so the results will be reproducible.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Section A.3.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] See Section 4.1.
   (b) Did you mention the license of the assets? [Yes] See Section A.3.
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] Consent provided by the authors who collected the data.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] It does not contain any sensitive information.
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Appendix

A.1 Group Theoretic Understanding of Optical Fibre Transmission Modes

When a light beam propagates in free space or in a transparent homogeneous medium, its transverse intensity profile generally changes. Despite this, there exist certain distributions that do not change intensity profile as they traverse. These fixed profiles are the transmission modes of the space.

The development of group equivariant networks exploits a similar principle, where these networks are constructed under the more general principle of finding basis functions called irreducible representations of some group. Some examples of this principle include: for the circle S^1 or line $\mathbb{R}$ the irreducible representations are given by complex exponentials $\exp(in\theta)$, for the group $\mathrm{SO}(2)$ the irreducible representations are given by the circular harmonics, for the group $\mathrm{SO}(3)$ the irreducible representations are given by the Wigner D-functions, and for S^2 the irreducible representations are given by the spherical harmonics. Thus a function on the group can be composed as a linear combination of the corresponding irreducible representations. On the other hand, when learning a function it is possible to learn the weightings of each of the irreducible representations of the group to learn that function. Building a model such that its feature space comprises the irreducible representations of the group provides a method of constructing a model that is equivariant to the underlying group of the representations.

Here we seek to show a connection between the known properties of optical fibres and group equivariant networks. We start by providing some details on the propagation of light through a fibre. The propagation of light is governed by the time-independent (Helmholtz) equation, which in cylindrical coordinates is given by Equation 5.

$$\frac{\partial^2 E}{\partial r^2} + \frac{1}{r}\frac{\partial E}{\partial r} + \frac{1}{r^2}\frac{\partial^2 E}{\partial \phi^2} + q^2 E = 0. \tag{5}$$

The standard approach to solving the above equation is to use the separation-of-variables procedure, which assumes a solution of the form given in Equation 6.

$$E_z = AF_1(r)F_2(\phi). \tag{6}$$

Due to the circular symmetry of the fibre, each component must not change when the coordinate ϕ is increased by a multiple of 2π. Therefore, we make the following assumption:

$$F_2(\phi) = e^{i\nu\phi}, \tag{7}$$

where $\nu \in \mathbb{Z}$. Substituting into Equation 5 yields a wave equation of the following form:

$$\frac{\partial^2 F_1}{\partial r^2} + \frac{1}{r}\frac{\partial F_1}{\partial r} + \left(q^2 - \frac{\nu^2}{r^2}\right) F_1 = 0, \tag{8}$$

which is a differential equation for Bessel functions. Solving this both inside the core of the fibre and in the cladding of the fibre provides two solutions. In the core of the fibre, as $r \to \infty$ the guided modes must remain finite. Thus for $r \leq a$ for core radius a the solution is a Bessel function of the first kind:

$$E_z(r \leq a) = AJ_\nu(ur)e^{i\nu\phi}e^{i(\omega t - \beta z)} \tag{9}$$

and outside of the core the solution is a modified Bessel function of the second kind:

$$E_z(r \geq a) = CK_\nu(wr)e^{i\nu\phi}e^{i(\omega t - \beta z)}. \tag{10}$$

Solving these equations provides the transmission modes of the fibre, i.e. the non-changing distributions.

Next we consider the propagation of light through a fibre from a group theoretic perspective. The indefinite special orthogonal group $\mathrm{SO}(2,1)$ is the group considering two spatial dimensions and one time dimension, and can be realised as:

$$SO(2,1) = \{X \in \mathrm{Mat}_3(\mathbb{R}) | X^t \nu X = \nu, \det(X) = 1\} \tag{11}$$

where,

$$\nu = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{12}$$

To identify the Lie algebra $\mathfrak{so}(2,1)$ we use the tangent space $T_1 SO(2,1)$ to $\mathrm{SO}(2,1)$ at the identity 1. We then choose a curve $a : L \to SO(21)$ such that $a'(0) = A$ Then, the characterising equation of A gives:

$$a(t)^T \nu a(t) = \nu. \tag{13}$$

Taking the derivative with respect to t gives:

$$a'(t)^T \nu a(t) + a(t)^T \nu a'(t) = 0. \tag{14}$$

Then, evaluating the expression at $t = 0$ gives:

$$A^T \nu + \nu A = 0. \tag{15}$$

Now we check the linear conditions determined by the above characterisation to then write out the Lie algebra, with ν having a natural block decomposition. Therefore, for a general element $A \in \mathfrak{so}(2,1)$, given as:

$$A = \begin{pmatrix} W & x \\ y^T & z \end{pmatrix}, \tag{16}$$

where $W \in M(2,\mathbb{R})$, $x, y \in \mathbb{R}^2$, and $z \in \mathbb{R}$. In this block decomposition $\nu = \begin{pmatrix} \mathbb{I}_2 & 0 \\ 0 & -1 \end{pmatrix}$. Then, Equation 15 becomes:

$$\begin{pmatrix} W^T & y \\ x^T & z \end{pmatrix} \begin{pmatrix} \mathbb{I}_2 & 0 \\ 0 & -1 \end{pmatrix} + \begin{pmatrix} \mathbb{I}_2 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} W & x \\ y^T & z \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \tag{17}$$

Therefore, the following conditions are imposed:

$$\begin{aligned} W^T &= -W \\ y &= x \\ z &= 0. \end{aligned}$$

Hence, the Lie algebra $\mathfrak{so}(2,1)$ is given by:

$$\mathfrak{so}(2,1) = \left\{ \begin{pmatrix} W & x \\ x^T & 0 \end{pmatrix} : W^T = -W \right\} = \left\{ \begin{pmatrix} 0 & -w & x_1 \\ w & 0 & x_2 \\ x_1 & x_2 & 0 \end{pmatrix} \right\}. \tag{18}$$

As a result we can see that the condition on W is such that $W \in \mathfrak{so}(2,\mathbb{R})$. Therefore, the condition on W characterises the rotational symmetry of the group. Given the goal is make the connection between the group theoretic understanding of group equivariant neural networks and solution to the wave equation in cylindrical coordinates, this is promising as the solutions to Bessel functions have rotational symmetries.

Further, the group $\mathrm{SO}(2,1)$ is that of two spatial dimensions and one time dimension. The connection between the three spatial dimensional case and the view as a light cone was made in the development of Minkowski spacetime. Here you imagine a cone where the time axis runs from the point of a cone through the centre of the plane drawn on the open end and the two spatial axes form a plane which intersects the cone and is parallel to the plane drawn on the open end. This is known as the future light cone and is an interpretation of how light spreads out after an event occurs. The group actions of the group $\mathrm{SO}^+(2,1)$, which is the group $\mathrm{SO}(2,1)$ with the requirements that $t > 0$, act on this space and can be viewed as rotations of the three dimensional Euclidean sphere. A connection can be drawn between this view and the fact that non compact generators of $\mathrm{SO}(n,1)$ differ from corresponding matrix elements of the same generators of $\mathrm{SO}(n+1)$, the group of rotations in $n+1$-dimensional space, by a factor of $\sqrt{(-1)}$ (Wong, 1974). Therefore, the connection can be made between the group actions of $\mathrm{SO}^+(2,1)$ and the light cone view of light propagation.

A final connection can be made between the light cone and the wave equation, given the connection between the group action and the light cone. Returning to the wave equation we note that it describes waves travelling with frequency independent speed. The character of the solution is different in odd and even dimensional spaces. In an odd dimensional space a disturbance propagates on the light cone and vanishes elsewhere. On the other hand, in an even dimensional space a disturbance propagates inside the entire light cone. Therefore, in an even dimensional space a disturbed medium

never returns to rest. This phenomena is known as geometric dispersion. Here we are interested in even dimensional space as we have two spatial dimensions governing transmission through the fibre along with a third time dimension. We therefore expect the propagation of light through the fibre to be understood by propagation inside the entire light cone. Given the wave equation:

$$u_{tt} = c^2(u_{xx} + u_{yy}). \tag{19}$$

A solution is possible by considering the three-dimensional theory, if we regard u as a function in three dimensions and that the third dimension is independent. So, if we require:

$$\begin{aligned} u_0(0, x, y) &= 0 \\ u_t(0, x, y) &= \phi(x, y), \end{aligned}$$

then the three-dimensional solution equation becomes:

$$u(t, x, y) = tM_{ct}[\phi] = \frac{t}{4\pi} \iint_S \phi(x + ct\alpha, y + ct\beta) d\omega, \tag{20}$$

where α and β are the first two coordinates on the unit sphere and $d\omega$ is the area element on the sphere. The integral can be written as a double integral over disc D with centre x, y and radius ct:

$$u(t, x, y) = \frac{1}{2\pi ct} \iint_D \frac{\phi(x + \xi, y + \nu)}{\sqrt{(ct)^2 - \xi^2 - \nu^1}} d\xi d\nu. \tag{21}$$

It becomes clear that the solution does not only depend on the data on the light cone where:

$$(x - \xi)^2 + (y - \nu)^2 = c^2t^2, \tag{22}$$

but on the entire data inside the light cone. Therefore, it can be seen how the solution to the wave equation yields the interpretation of the light cone in the precise way that was yielded by the analysis of the group $\mathrm{SO}^+(2, 1)$.

This completes the connections between the group action of $\mathrm{SO}^+(2, 1)$ and the solution to the Bessel functions that govern the transmission modes within optical fibres. Now we can compose a model in a similar way to how circular harmonics are used to construct rotation equivariant neural networks under the group $\mathrm{SO}(2)$, but with utilising the basis functions found by solving the Bessel function in Equation 8 to compose a network which is equivariant to the group $\mathrm{SO}^+(2, 1)$. This is useful as it positions the model with respect to other group equivariant neural networks and provides a model with suitable inductive bias for the task of learning a function approximation to transmission through optical fibres.

A.2 Related Work

One method for inverting the transmission effects is to find the underlying TM of the multi-mode fibre. In general this is not known, although in principle the TM can be found by acquiring the output amplitude and phase relative to each mode (Čižmár & Dholakia, 2011, 2012). Choi et al. (2012) present an approach to construct the TM in a scanner-free method based on measurement of amplitude and phase of the output. Although the method requires 500 measurement repetitions at different incidence angles. Mahalati et al. (2013) develop a method in which the number of resolvable image features approached four times the number of spatial modes. Papadopoulos et al. (2012) develop a digital phase conjugation technique to restore images without the requirement of calculating the full TM. Despite this, the solution of finding the TM, or part of it, has to be repeated for every different fibre, for every different length, under each different bending scenario, and for each different temperature, reducing the practicality of this solution. Plöschner et al. (2015) developed a procedure that could also incorporate bending through a precise characterisation of the fibre and a theoretical model.

Another approach to inverting the transmission effects is to use machine learning to solve the inverse problem. Moran et al. (2018) and Caramazza et al. (2019) approached solving the inverse problem through the use of either a Real and Complex linear model with a Hadamard layer to model the power drop off of the spatial light modulator. At its core this approach requires the use of a linear model, which maps from speckled images to the original images and scales as $\mathcal{O}(N_s^2 N_o^2)$, where

N_s is the resolution of the speckled image and N_o is the resolution of the original image. This is a very memory-expensive operation and restricts the scalability of the approach to relatively low resolution images. Further, Borhani et al. (2018) use a convolutional U-Net model to invert the transmission effects, which although using spatially localised convolutions is more memory efficient than a linear model, the speckled images are not in a similar spatial arrangement. An indication of this is likely in the fact their model requires 14 hidden layers to learn the inversion of 32×32 resolution images. Fan et al. (2019) use a convolutional neural network to invert transmission effects, where the convolutions and pooling reduce the speckled resolution down before a dense linear layer predicts the original image. As a result the model will be sensitive to the learned down-sampling of the convolutions and suffer from similar scalability issues to Moran et al. (2018) due to the inclusion of a dense linear layer. Further, Rahmani et al. (2018) also use a convolutional model and despite not having a dense linear layer, the model does comprise 22 convolutional layers, so, for low resolution, 28×28, images it is not surprising the model can overcome the mapping between two image domains with non-similar spatial arrangements as a mapping from each speckled pixel to each original image pixel is possible. There is therefore no evidence that the convolutional based models truly improve the scalability issue as these models are very large considering the resolution of images. Finally, of the machine learning based methods for inverting the transmission effects Rahmani et al. (2018) make the first attempt to demonstrate generalisation outside of the training domain, although the results demonstrate limited evidence of true generalisation due to the choice of images featuring limited high frequency features and low quality reconstructions, while Caramazza et al. (2019) demonstrate stronger generalisation due to the testing data choice, but there is still scope to improve the reconstruction quality outside of the training domain.

A.3 Training Details

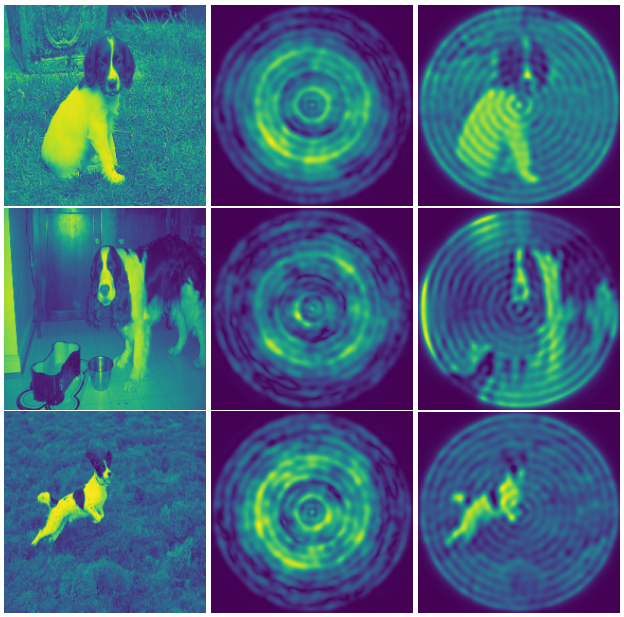
Throughout the paper we make use of three different datasets that are commonly used in machine learning, namely MNIST (LeCun et al., 1998), fMNIST (Xiao et al., 2017), and ImageNet (Deng et al., 2009). Where MNIST and fMNIST are also commonly used in the study of inverting transmission effects of MMFs due to their low resolutions. MNIST images are 28×28 images of handwritten digits between 0 and 9. fMNIST images are 28×28 images of items of clothing, such as trousers, jumpers, or shoes. ImageNet is a large scale dataset of higher resolution images which we size to be $256x256$. We only use a subset of the imageNet dataset to demonstrate the ability to invert such high resolution images. For each of these datasets we split into a training and testing dataset. The training dataset is used for training, while the testing dataset is not used during training and we use this to test the model after training. We use the testing dataset to generate all the predicted images throughout the paper.

We train each model for 200 epochs with the stochastic gradient descent algorithm. We use the mean squared error as a loss function between the original images and the predicted images. We do not use any regularisation on the model weights. For the Real and Complex linear models and the Bessel equivariant model we train each model for 200 epochs. For the post-processing model, as it is an addition to the Bessel equivariant model, we train this model for 200 epochs after the Bessel equivariant model has already been trained for 200 epochs and the model weights frozen. For the training, all models are trained on 1 Titan RTX GPU with 24GiB of memory. We did attempt training the complex linear model on a A6000 GPU with 48GB of memory on ImageNet to check if this was possible, but it required more memory.

We utilise two different sources of data throughout the experimental section. The first of these is collected using a theoretical TM for which we have both the amplitude and phase information for the speckled patterns. In addition, we also have access to the original and inverted images, where the inverted images are created by first passing an image through the TM and then back through its inverse. The second source of data is that collected by Moran et al. (2018), which is data collected in the lab using a real fibre. For this data we only have the amplitude information for the speckled patterns. In addition, we also have the original images which have constant phase. This data has no licence and it accessible on the projects GitHub page.

A.4 Invertibility of TMs

Given a TM, which models how an image propagates through a fiber, it is possible to invert the matrix and get the corresponding mapping back from the speckled image space to the original image space. We provide details of the construction of TMs in Appendix 2.2. Utilising this we can inspect information loss due to the limited number of modes of the fibre by passing the image through the TM and back through the inverse to create an inverted image. Therefore, if we knew the TM this inverted image would be the recoverable information. We show an example of the original image, the speckled image, and the inverted image in Figure 9.



(a) Original (b) Speckled (c) Inverted

Figure 9: (a) The original image. (b) The speckled image created by passing the original images through a theoretical TM. (c) The inverted image created by passing the speckled images through the inverted theoretical TM.

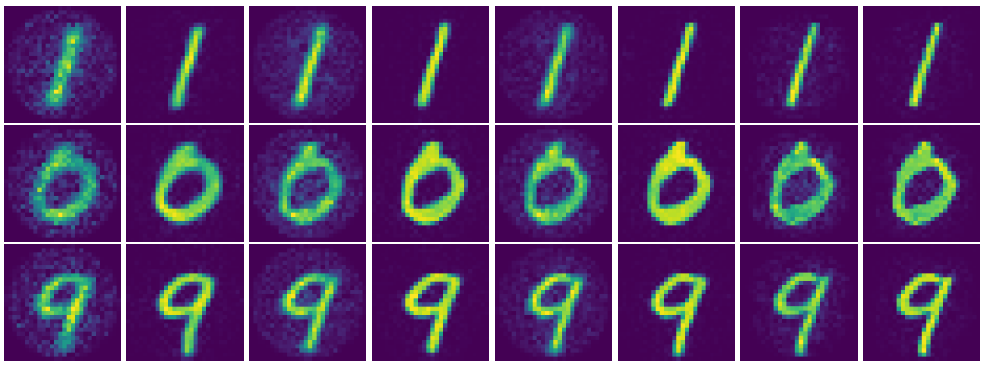
This is useful as it allows us to separate the task of inverting the transmission effects of a fibre into that which could be reconstructed by understanding the physics of the TM and that which requires generating due to being lost information. We believe this could therefore be viewed as two tasks: (1) a task of inverting the transmission effects which would have the aim of generating the inverted images and (2) a task, similar to a super-resolution task, of predicting the original images from the output of the first task.

A.5 Results for Accounting for Losses in a Real System

As we noted in Section 4.1 the assumption that the mapping between Bessel bases is diagonal holds in theory, although in practice an imperfect system could lead to a necessity to relax this assumption. Here we explore this for the MNIST data collected using a real fibre by building three different versions of our model (1) the strong assumption of a diagonal matrix, (2) relaxing the assumption to allow five elements either side of the diagonal to be populated, (3) relaxing the assumption to allow ten elements either side of the diagonal to be populated, and (4) one with a full matrix mapping between bases. If the fibre and experimental set-up was in an ideal setting we would expect model (1) to produce results as strong as those demonstrated in Section 4.2 where data was created with a theoretical TM. On the other hand, model (2) allowing 5 elements to be populated either side of the diagonal would allow the model to capture some of the block diagonal structure seen in Carpenter et al. (2014), although this would still occlude the mapping between modes further apart in our mapping space. Similarly, (3) allows for capturing more of the block diagonal structure by allowing 10 elements to be populated either side of the diagonal. Finally, model (4) allows the greatest flexibility and has the potential to capture deviations from theory as seen in practice, although, this

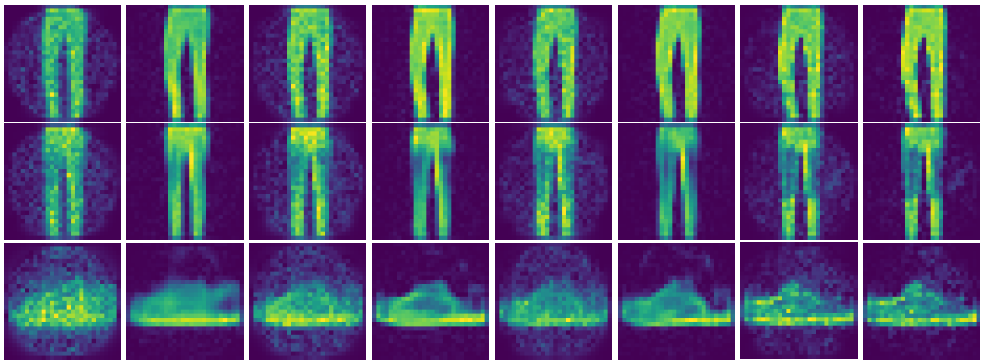
model does not take advantage of the known diagonal structure that this mapping function has and is therefore over-parameterised.

Figure 10 demonstrates that as the diagonal assumption of the mapping function between Bessel bases is relaxed, our Bessel equivariant model produced images closer to that original image through capturing more high frequency detail and predicting less noise. Our model with five elements either side of the diagonal populated with trainable parameters achieves a comparable result to the Real and Complex linear models, and our most constrained model is able to generate digits that like the correct digit. This is a promising result as it demonstrates that our model allows for a design choice of complexity, with the option to trade off performance for a reduced memory requirement whilst maintaining accurate results even in the lowest memory configuration; this is something that is not possible with the Real or Complex linear models. Choosing a model with a low number of off-diagonal trainable elements, we believe, is the best trade off between considering losses and imperfections of the real-world while benefiting from the known sparsity of the mapping function between Bessel bases.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 10: Comparison of predicted images from inverting transmission effects of a MMF with varying relaxations on the assumption the fibre has a diagonal fibre propagation matrix. (a) the output of the Bessel equivariant model with a diagonal propagation matrix assumption between Bessel bases, (b) the output of the combination of Bessel equivariant with a diagonal propagation matrix assumption between Bessel bases and post-processing model, (c) the output of the Bessel equivariant model with 5 diagonal offsets in the propagation matrix between Bessel bases, (d) the output of the Bessel equivariant model with 5 diagonal offsets in the propagation matrix between Bessel bases and post-processing model, (e) the output of the Bessel equivariant model with 10 diagonal offsets in the propagation matrix between Bessel bases, (f) the output of the Bessel equivariant model with 10 diagonal offsets in the propagation matrix between Bessel bases and post-processing model, (g) the output of the Bessel equivariant model with a full propagation matrix between Bessel bases, and (h) the output of the Bessel equivariant model with a full propagation matrix between Bessel bases and post-processing model.



(a) (b) (c) (d) (e) (f) (g) (h)

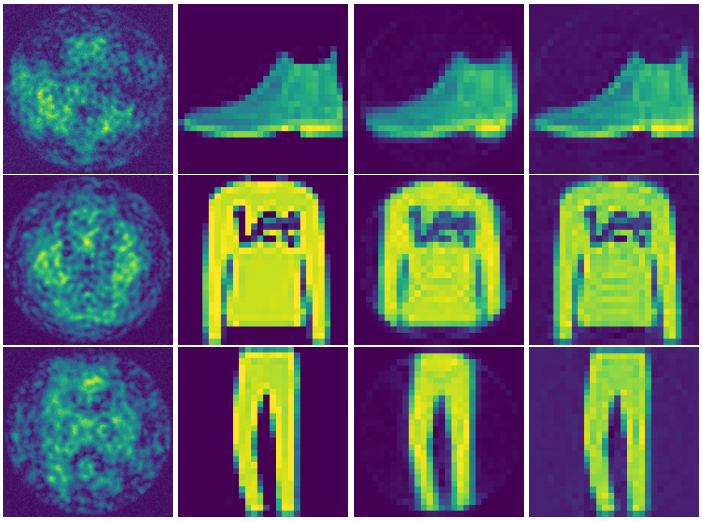
Figure 11: Comparison of predicted images from inverting transmission effects of a MMF with varying relaxations on the assumption the fibre has a diagonal fibre propagation matrix. (a) the output of the Bessel equivariant model with a diagonal propagation matrix assumption between Bessel bases, (b) the output of the combination of Bessel equivariant with a diagonal propagation matrix assumption between Bessel bases and post-processing model, (c) the output of the Bessel equivariant model with 5 diagonal offsets in the propagation matrix between Bessel bases, (d) the output of the Bessel equivariant model with 5 diagonal offsets in the propagation matrix between Bessel bases and post-processing model, (e) the output of the Bessel equivariant model with 10 diagonal offsets in the propagation matrix between Bessel bases, (f) the output of the Bessel equivariant model with 10 diagonal offsets in the propagation matrix between Bessel bases and post-processing model, (g) the output of the Bessel equivariant model with a full propagation matrix between Bessel bases, and (h) the output of the Bessel equivariant model with a full propagation matrix between Bessel bases and post-processing model.

A.6 Effects of Noise on Speckled Images

Here we compare the effect of noise on the speckled images when using a theoretical TM. In each case the speckled images are saturated at 90% of there maximum value and Gaussian noise is added with different standard deviation values. We demonstrate the reconstruction ability of our model in Figure 12 when Gaussian noise with 0.01 standard deviation, in Figure 13 when Gaussian noise with 0.05 standard deviation, in Figure 14 when Gaussian noise with 0.1 standard deviation, and in Figure 15 when Gaussian noise with 0.5 standard deviation. Further, we provide the loss values in Table 5. This demonstrates that our approach is robust to noise.

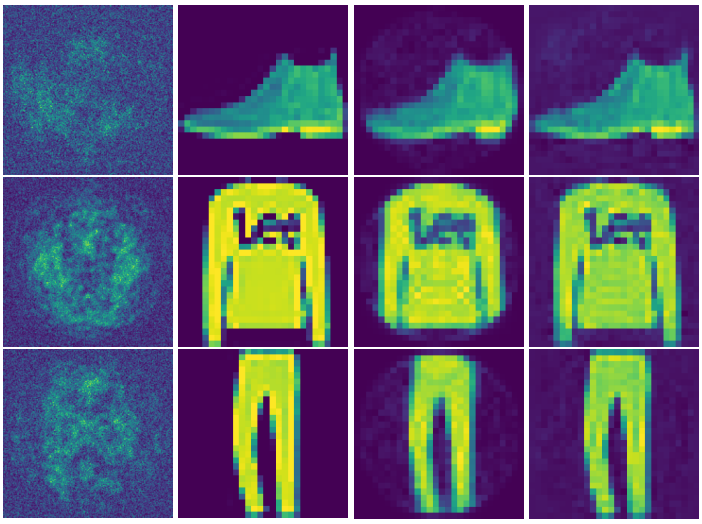
Table 5: Comparison of the loss values of each model trained with F-MNIST data when the speckled images are saturated at 90% and Gaussian noise is added with standard deviation given in the column Noise Level.

Noise Level	Model	Train Loss	Test Loss
0.01	Bessel Equivariant	0.0140	0.0142
	Bessel Equivariant + Post Proc	0.0033	0.0033
0.05	Bessel Equivariant	0.0147	0.0149
	Bessel Equivariant + Post Proc	0.0043	0.0043
0.1	Bessel Equivariant	0.0166	0.0168
	Bessel Equivariant + Post Proc	0.0049	0.0049
0.5	Bessel Equivariant	0.0373	0.0377
	Bessel Equivariant + Post Proc	0.0164	0.0166



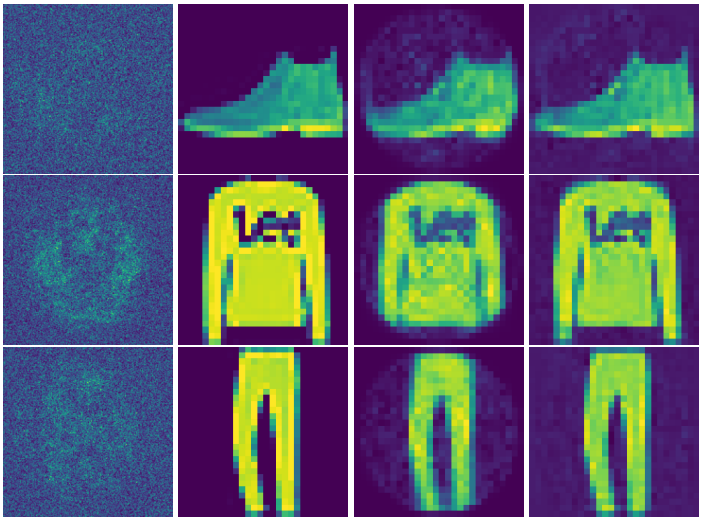
(a) Input (b) Target (c) BEM (d) BEM + PP

Figure 12: Comparison of predicted images from inverting transmission effects of a MMF using FMnist data created with a theoretical TM when the speckled images are saturated at 90% of their maximum value and Gaussian noise with 0.01 standard deviation added. (a) The input noisy speckled image, (b) the target original image to reconstruct, (c) the output of the Bessel equivariant model, and (d) the output of the combination of Bessel equivariant and post-processing model.



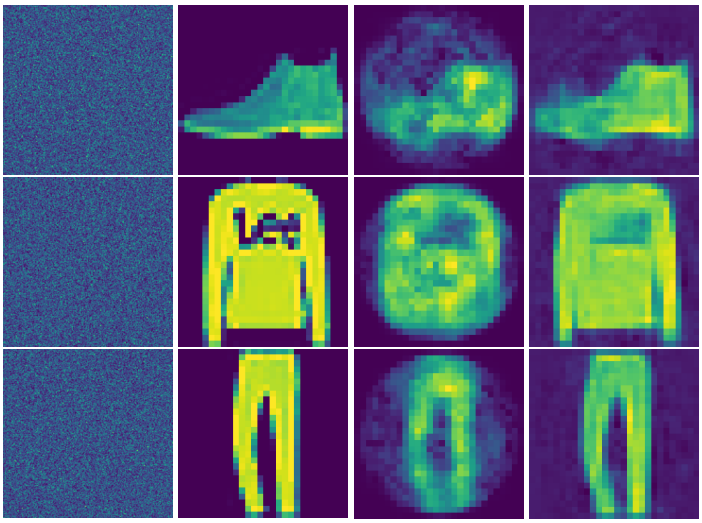
(a) Input (b) Target (c) BEM (d)BEM + PP

Figure 13: Comparison of predicted images from inverting transmission effects of a MMF using FMnist data created with a theoretical TM when the speckled images are saturated at 90% of their maximum value and Gaussian noise with 0.05 standard deviation added. (a) The input noisy speckled image, (b) the target original image to reconstruct, (c) the output of the Bessel equivariant model, and (d) the output of the combination of Bessel equivariant and post-processing model.



(a) Input (b) Target (c) BEM (d) BEM + PP

Figure 14: Comparison of predicted images from inverting transmission effects of a MMF using FMnist data created with a theoretical TM when the speckled images are saturated at 90% of their maximum value and Gaussian noise with 0.1 standard deviation added. (a) The input noisy speckled image, (b) the target original image to reconstruct, (c) the output of the Bessel equivariant model, and (d) the output of the combination of Bessel equivariant and post-processing model.



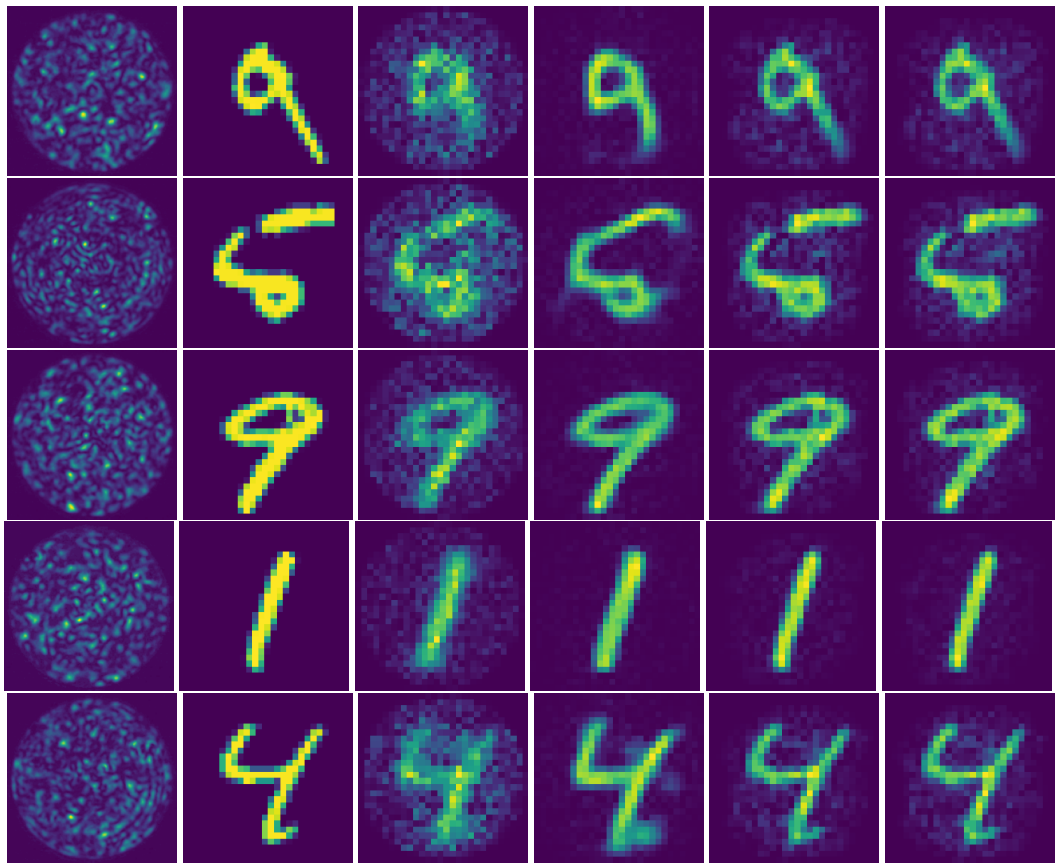
(a) Input (b) Target (c) BEM (d) BEM + PP

Figure 15: Comparison of predicted images from inverting transmission effects of a MMF using FMnist data created with a theoretical TM when the speckled images are saturated at 90% of their maximum value and Gaussian noise with 0.05 standard deviation added. (a) The input noisy speckled image, (b) the target original image to reconstruct, (c) the output of the Bessel equivariant model, and (d) the output of the combination of Bessel equivariant and post-processing model.

A.7 Additional Reconstructions - Real Fibre MNIST

Table 6: Comparison of the loss values of each model trained with MNIST data.

Model	MNIST Train Loss	MNIST Test Loss
Real Linear	0.00466	0.00683
Complex Linear	0.00396	0.00684
Bessel Equivariant Diag	0.03004	0.03125
Bessel Equivariant Diag + Post Proc	0.01317	0.01453
Bessel Equivariant 5 Off Diag	0.01782	0.01931
Bessel Equivariant 5 Off Diag + Post Proc	0.00658	0.00740
Bessel Equivariant 10 Off Diag	0.01362	0.01521
Bessel Equivariant 10 Off Diag + Post Proc	0.00488	0.00576
Bessel Equivariant Full	0.00300	0.00684
Bessel Equivariant Full + Post Proc	**0.00145**	**0.00378**



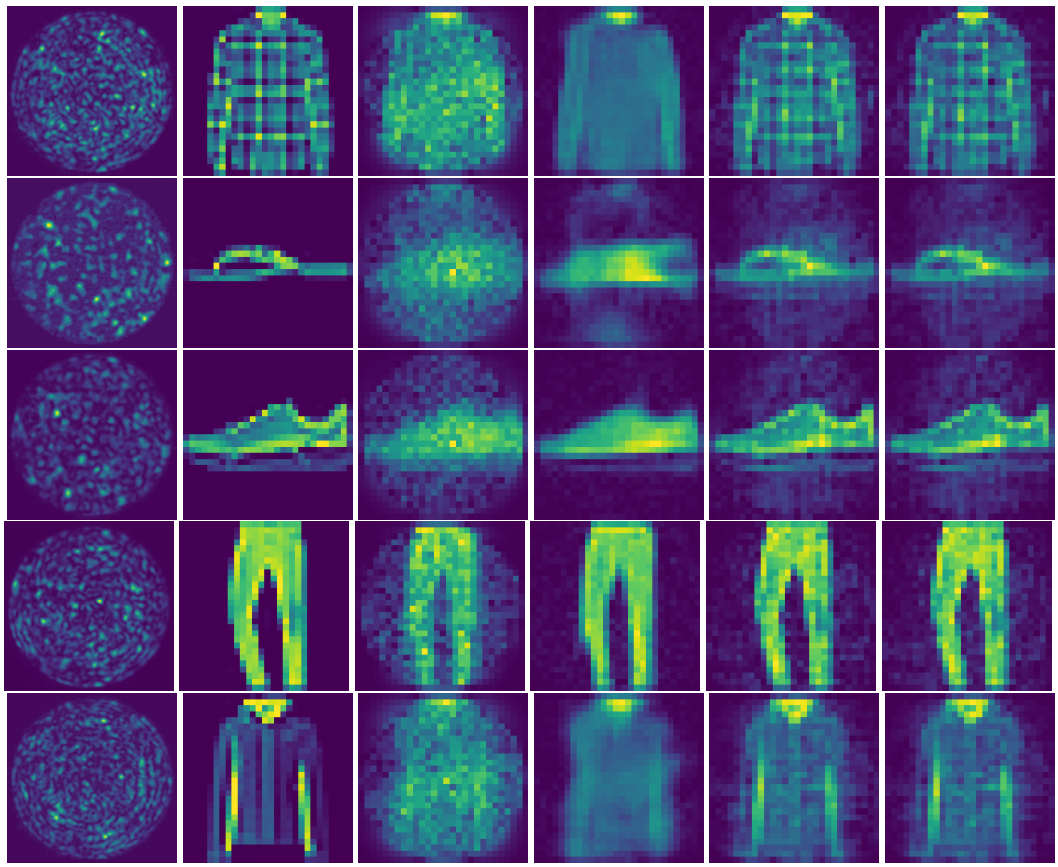
(a) Input (b) Target (c) BEM (d) BEM + PP (e) Real (f) Complex

Figure 16: Comparison of predicted images from inverting transmission effects of a MMF. (a) The input speckled image, (b) the target original image to reconstruct, (c) the output of the Bessel equivariant model, (d) the output of the combination of Bessel equivariant and post-processing model, (e) the output of the Real valued linear model, and (f) the output of the Complex valued linear model.

A.8 Additional Reconstructions - Real Fibre fMNIST

Table 7: Comparison of the loss values of each model trained with fMNIST data.

Model	fMNIST Train Loss	fMNIST Test Loss
Real Linear	0.00586	0.01061
Complex Linear	0.00509	0.01061
Bessel Equivariant Diag	0.02903	0.03140
Bessel Equivariant Diag + Post Proc	0.01609	0.01749
Bessel Equivariant 5 Off Diag	0.02057	0.02354
Bessel Equivariant 5 Off Diag + Post Proc	0.01138	0.01315
Bessel Equivariant 10 Off Diag	0.01788	0.02140
Bessel Equivariant 10 Off Diag + Post Proc	0.00943	0.01162
Bessel Equivariant Full	0.00548	0.01380
Bessel Equivariant Full + Post Proc	**0.00306**	**0.00964**



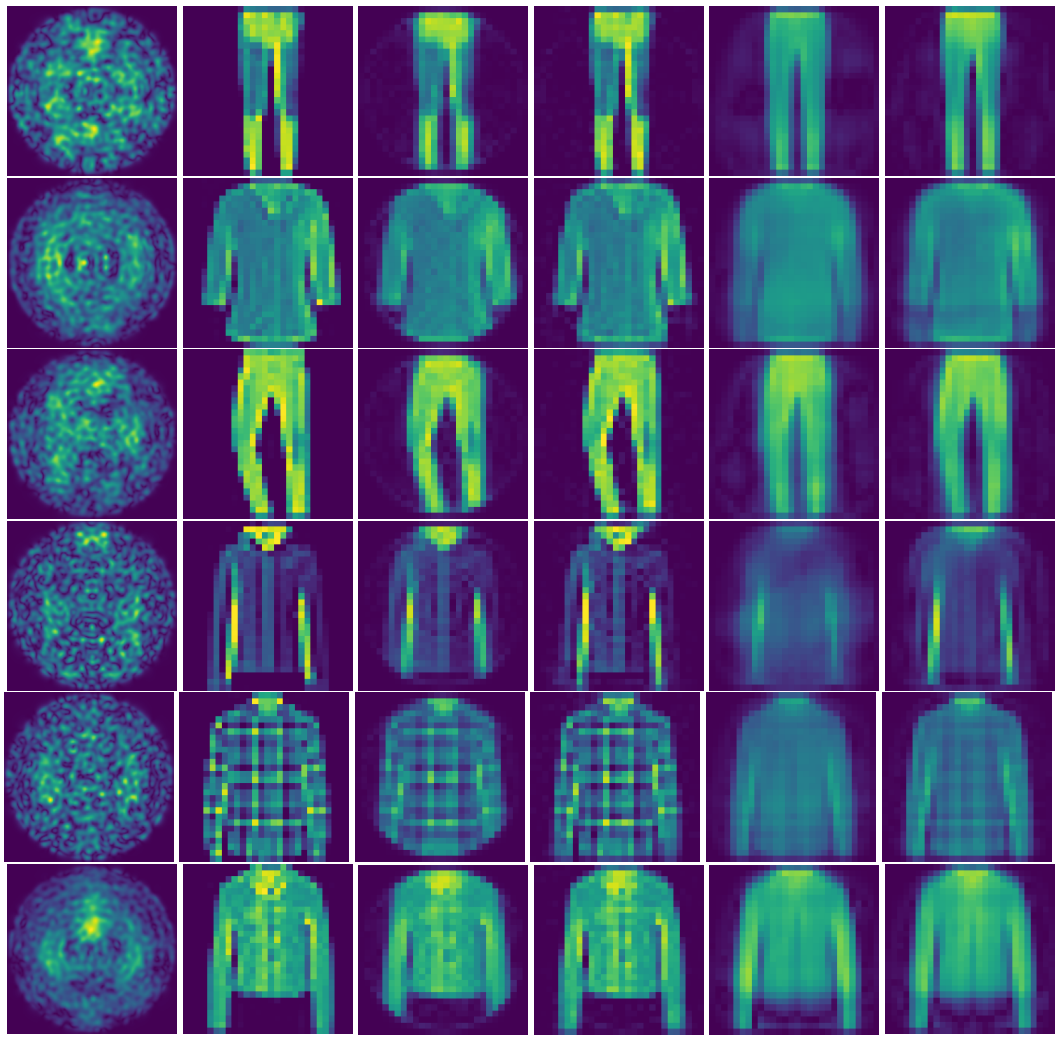
(a) Input (b) Target (c) BEM (d) BEM + PP (e) Real (f) Complex

Figure 17: Comparison of predicted images from inverting transmission effects of a MMF. (a) The input speckled image, (b) the target original image to reconstruct, (c) the output of the Bessel equivariant model, (d) the output of the combination of Bessel equivariant and post-processing model, (e) the output of the Real valued linear model, and (f) the output of the Complex valued linear model.

A.9 Additional Reconstructions - Theory TM fMNIST

Table 8: Comparison of the loss values of each model trained with fMNIST data.

Model	fMNIST Train Loss	fMNIST Test Loss
Real Linear	0.0256	0.0250
Complex Linear	0.0149	0.0146
Bessel Equivariant	0.0141	0.0139
Bessel Equivariant + Post Proc	**0.0032**	**0.0032**



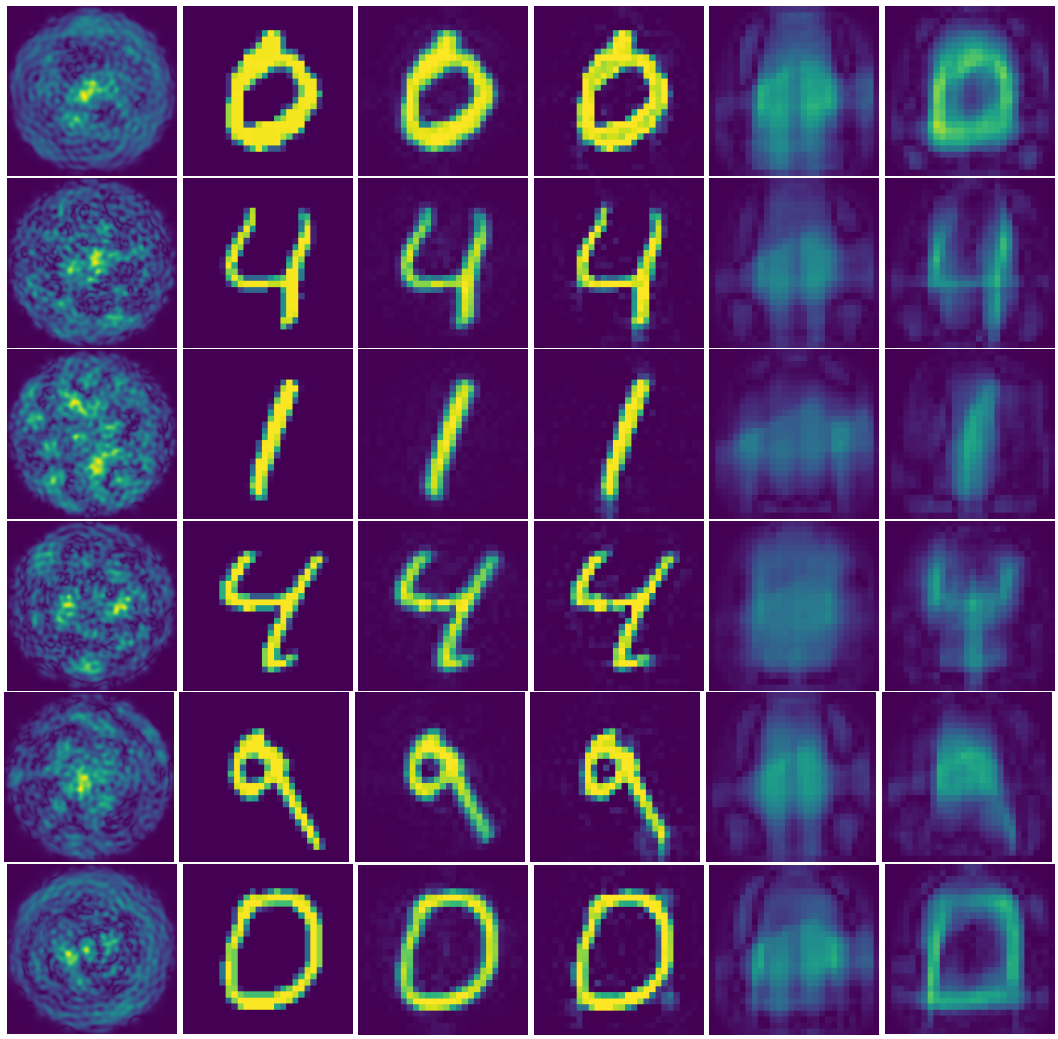
(a) Input (b) Target (c) BEM (d) BEM + PP (e) Real (f) Complex

Figure 18: Comparison of predicted images from inverting transmission effects of a MMF. (a) The input speckled image, (b) the target original image to reconstruct, (c) the output of the Bessel equivariant model, (d) the output of the combination of Bessel equivariant and post-processing model, (e) the output of the Real valued linear model, and (f) the output of the Complex valued linear model.

A.10 Additional Reconstructions - Theory TM MNIST

Table 9: Comparison of the loss values of each model trained with fMNIST data.

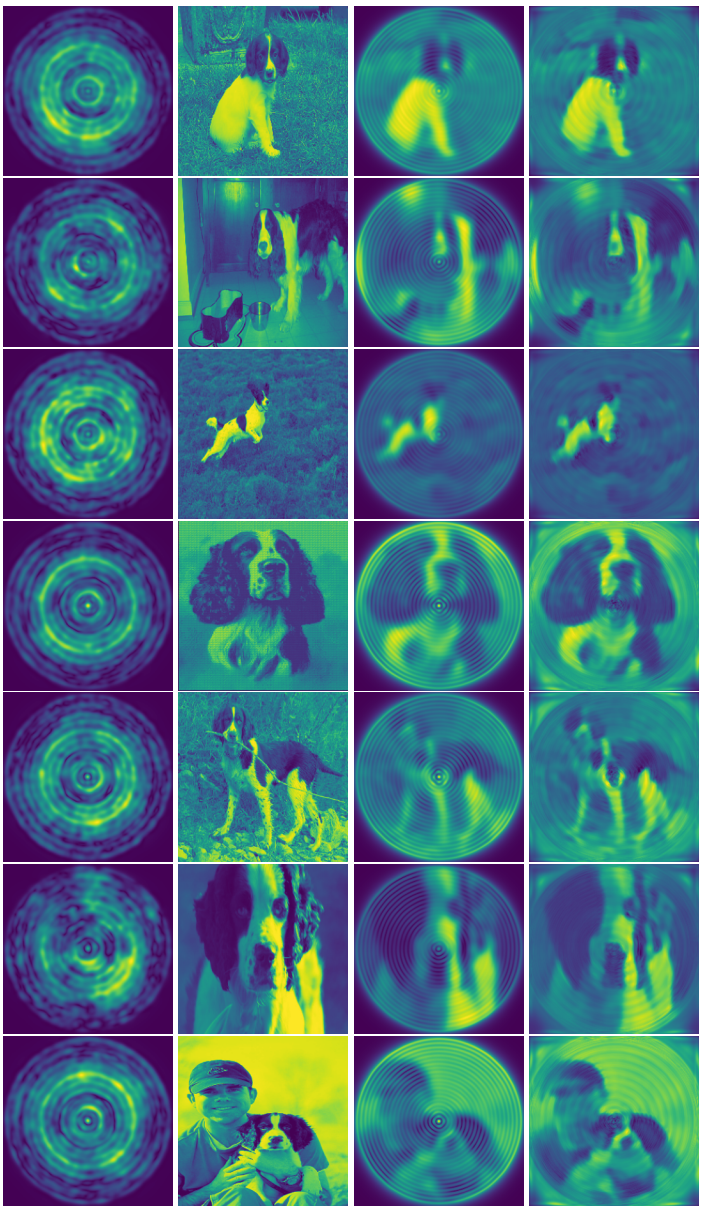
Model	MNIST Test Loss
Real Linear	0.0641
Complex Linear	0.0363
Bessel Equivariant	**0.0026**
Bessel Equivariant + Post Proc	0.0028



(a) Input (b) Target (c) BEM (d) BEM + PP (e) Real (f) Complex

Figure 19: Comparison of predicted images from inverting transmission effects of a MMF. (a) The input speckled image, (b) the target original image to reconstruct, (c) the output of the Bessel equivariant model, (d) the output of the combination of Bessel equivariant and post-processing model, (e) the output of the Real valued linear model, and (f) the output of the Complex valued linear model.

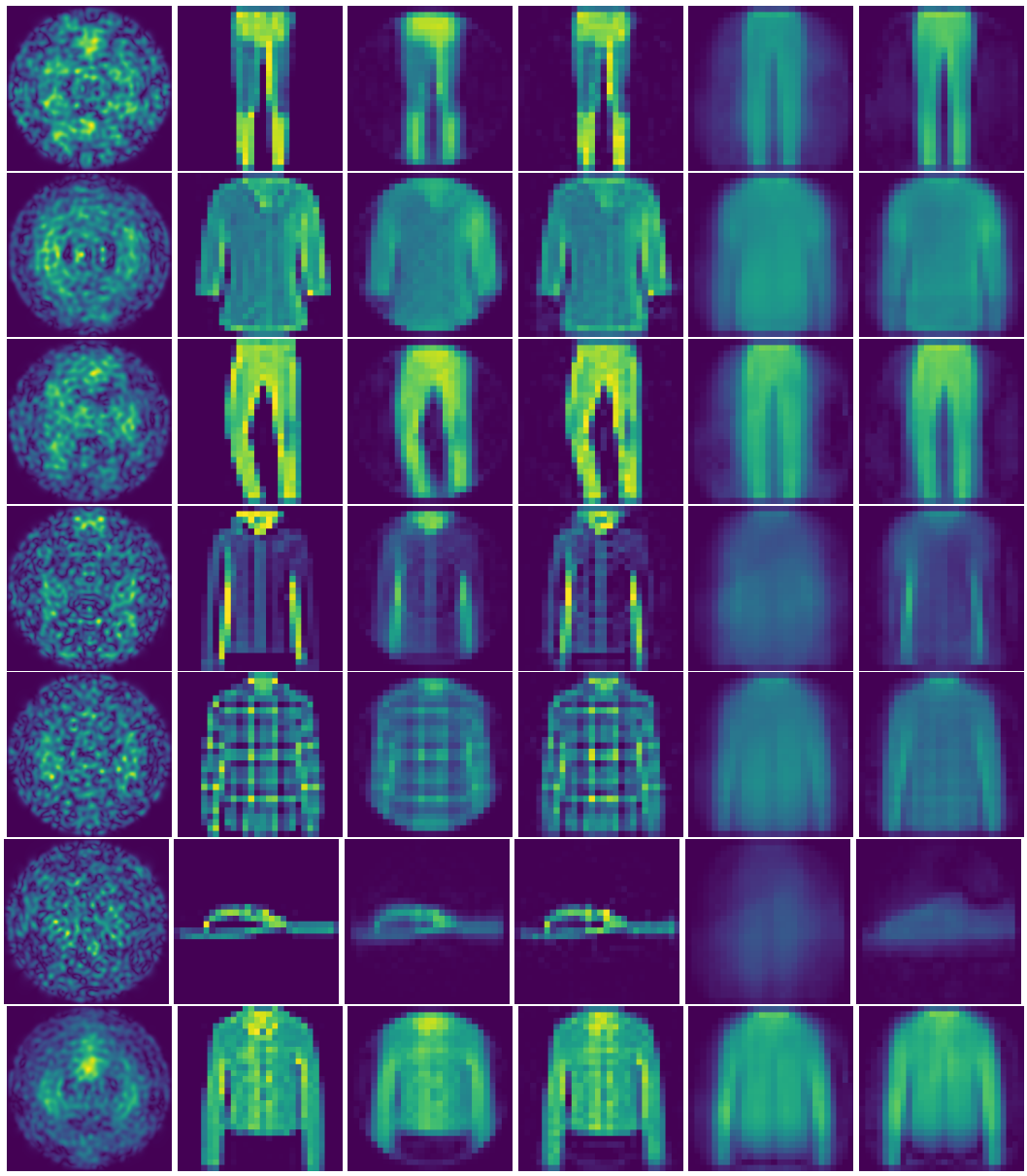
A.11 Additional Reconstructions - Theory TM ImageNet



(a) Input (b) Target (c) BEM (d) BEM + PP

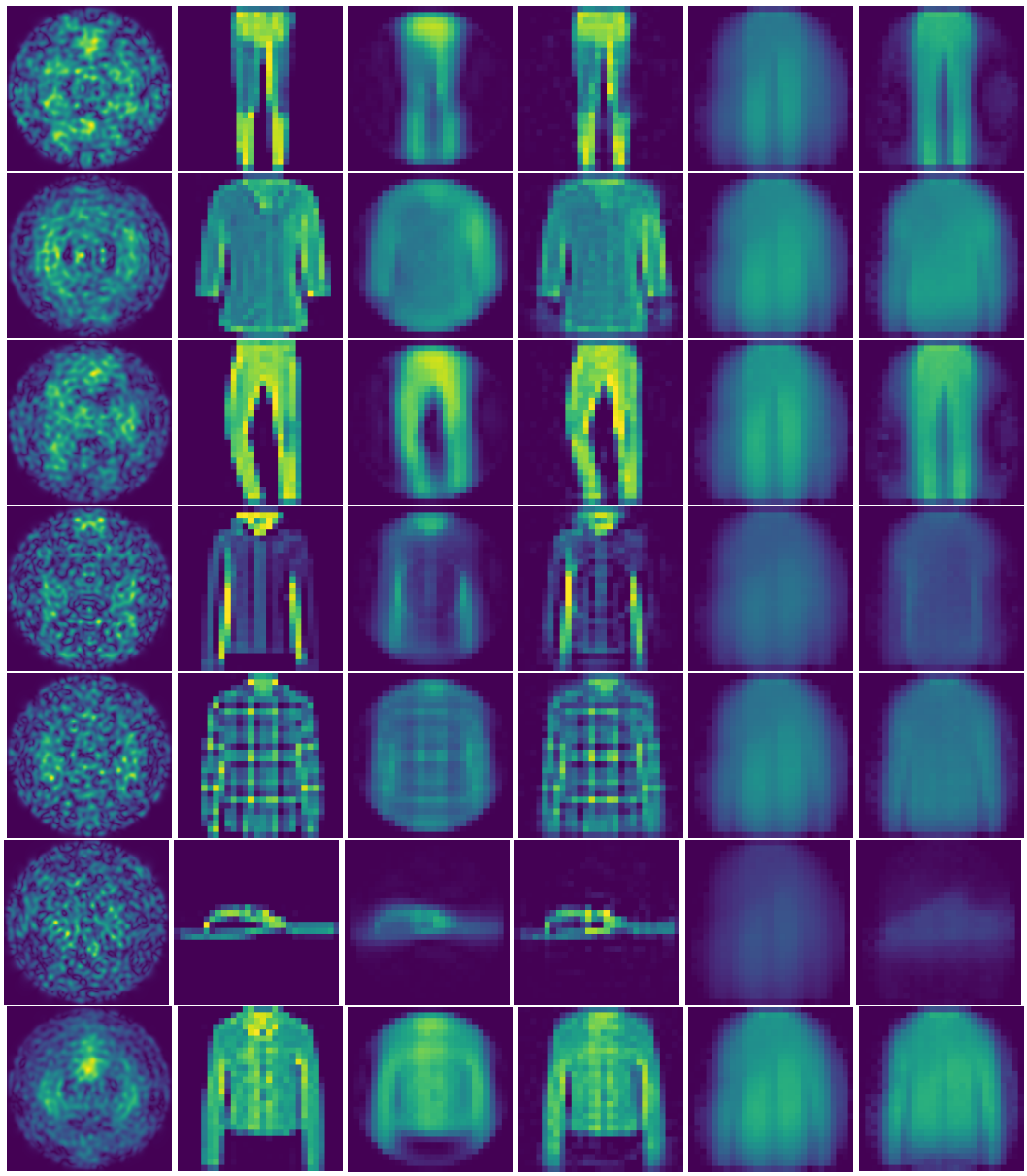
Figure 20: Comparison of predicted images from inverting transmission effects of a MMF. (a) The input speckled image, (b) the target original image to reconstruct, (c) the output of the Bessel equivariant model, and (d) the output of the combination of Bessel equivariant and post-processing model.

A.12 Impacts of Reducing Training Dataset Size - Theory TM fMNIST



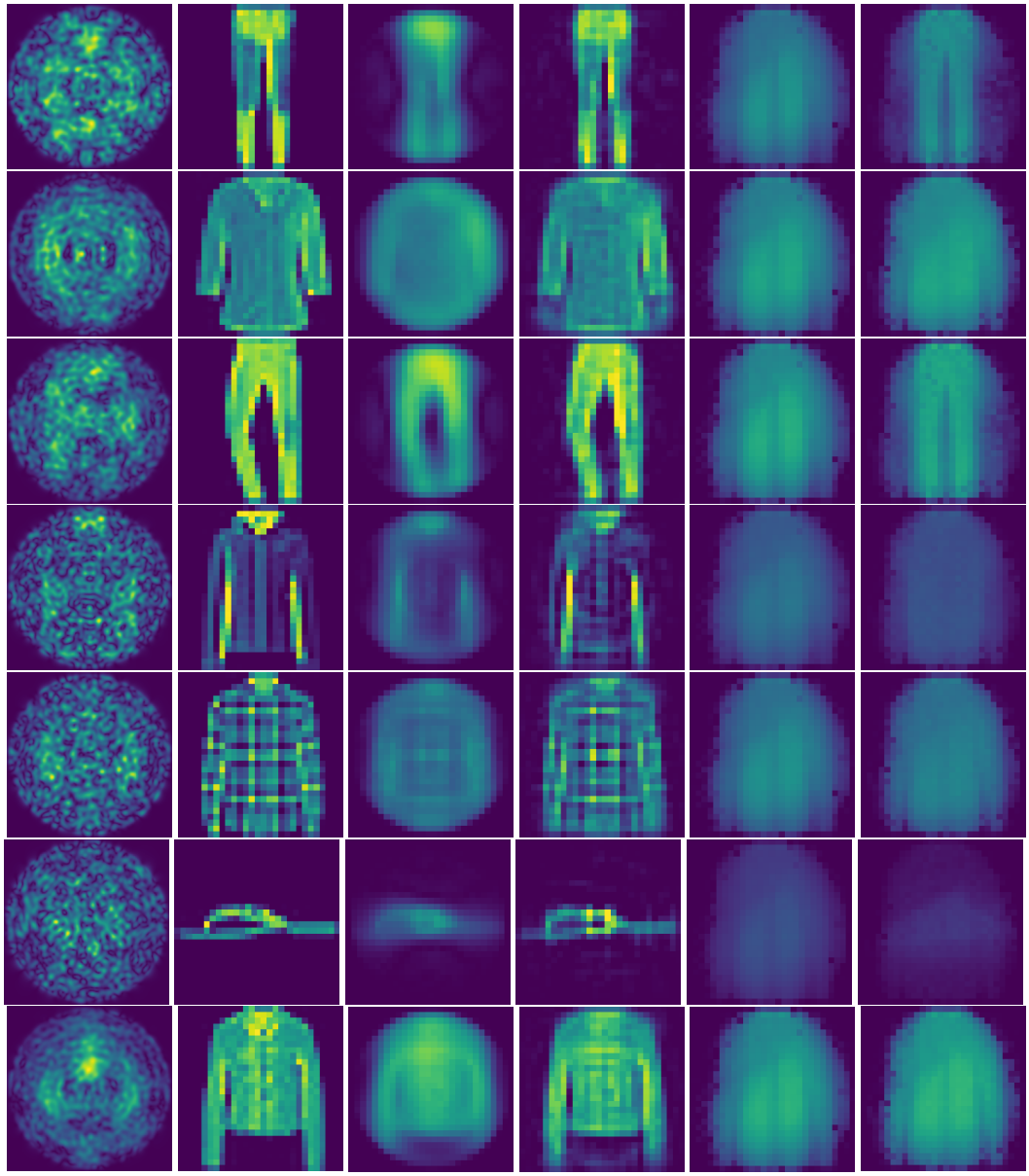
(a) Input (b) Target (c) BEM (d) BEM + PP (e) Real (f) Complex

Figure 21: Comparison of predicted images from inverting transmission effects of a MMF. The training dataset was reduced from the original size of 12000 to 6000. (a) The input speckled image, (b) the target original image to reconstruct, (c) the output of the Bessel equivariant model, (d) the output of the combination of Bessel equivariant and post-processing model, (e) the output of the Real valued linear model, and (f) the output of the Complex valued linear model. The data used was fMNIST with speckled patterns created with a theoretical TM.



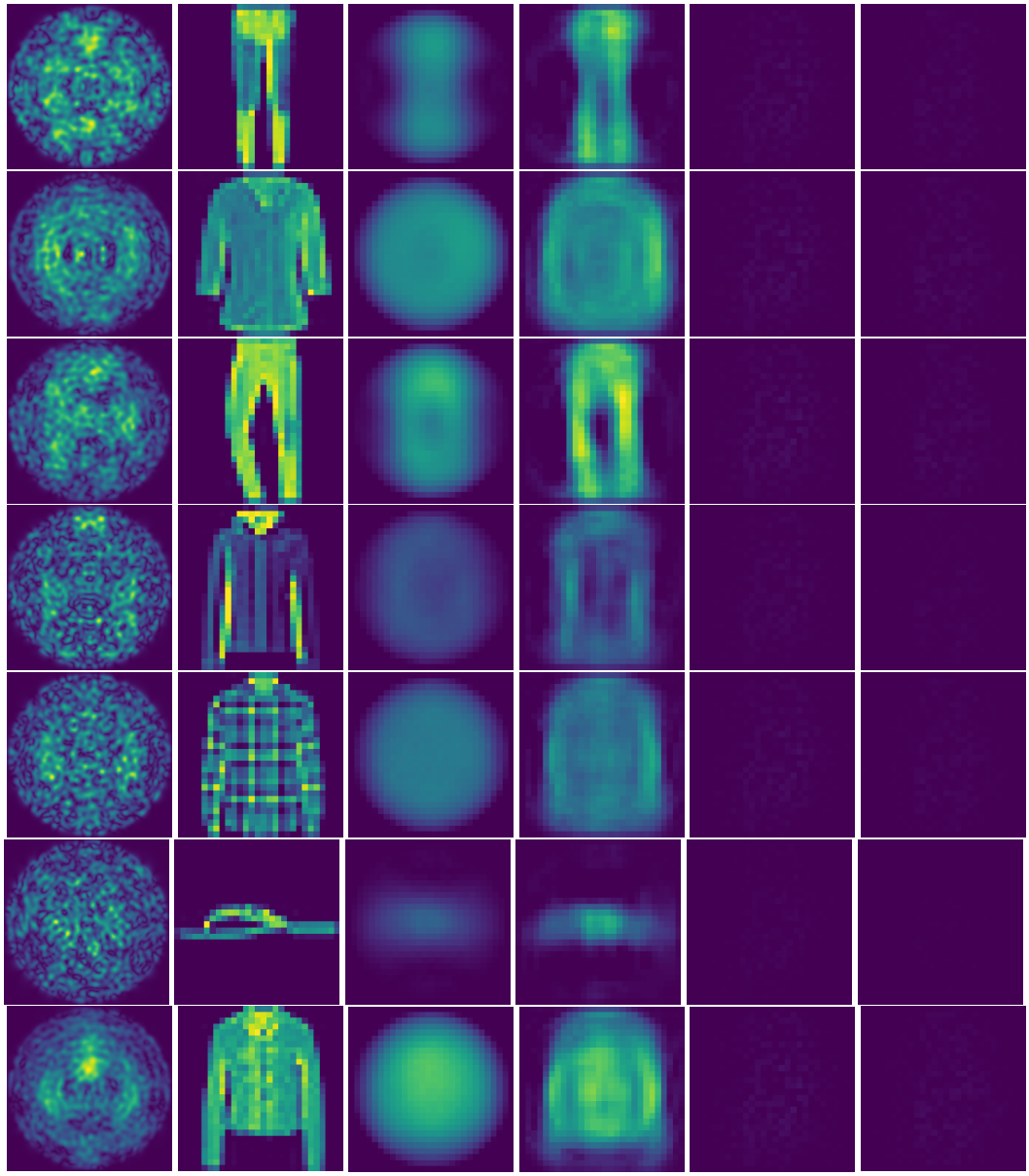
(a) Input (b) Target (c) BEM (d) BEM + PP (e) Real (f) Complex

Figure 22: Comparison of predicted images from inverting transmission effects of a MMF. The training dataset was reduced from the original size of 12000 to 2400. (a) The input speckled image, (b) the target original image to reconstruct, (c) the output of the Bessel equivariant model, (d) the output of the combination of Bessel equivariant and post-processing model, (e) the output of the Real valued linear model, and (f) the output of the Complex valued linear model. The data used was fMNIST with speckled patterns created with a theoretical TM.



(a) Input (b) Target (c) BEM (d) BEM + PP (e) Real (f) Complex

Figure 23: Comparison of predicted images from inverting transmission effects of a MMF. The training dataset was reduced from the original size of 12000 to 1200. (a) The input speckled image, (b) the target original image to reconstruct, (c) the output of the Bessel equivariant model, (d) the output of the combination of Bessel equivariant and post-processing model, (e) the output of the Real valued linear model, and (f) the output of the Complex valued linear model. The data used was fMNIST with speckled patterns created with a theoretical TM.



(a) Input (b) Target (c) BEM (d) BEM + PP (e) Real (f) Complex

Figure 24: Comparison of predicted images from inverting transmission effects of a MMF. The training dataset was reduced from the original size of 12000 to 120. (a) The input speckled image, (b) the target original image to reconstruct, (c) the output of the Bessel equivariant model, (d) the output of the combination of Bessel equivariant and post-processing model, (e) the output of the Real valued linear model, and (f) the output of the Complex valued linear model. The data used was fMNIST with speckled patterns created with a theoretical TM.

A.13 Impacts of Under Parameterising the set of Bessel Function Bases - Theory TM fMNIST

Here we consider the impact on under parameterising the Bessel function basis. For this we reduce the number of radial frequencies that feature in the set of bases functions. The original basis set we utilised for the data collected using a theoretical fibre for fMNIST comprises of 21 radial frequencies and 1061 bases. Here we show the original results using this full bases set in Figure 26 column (b). Next we show the results of removing the high frequency bases by only considering the first 14 radial frequencies, which amounts to having 932 bases in Figure 26 column (c). In addition, we show the results of removing further high frequency bases by only considering the first 7 radial frequencies, which amounts to having 567 bases in Figure 26 column (d). Finally, we show the results of removing further high frequency bases by only considering the first 4 radial frequencies, which amounts to having 322 bases in Figure 26 column (e).

Table 10: Comparison of the loss values of each model trained with fMNIST data created with a theoretical TM.This shows the impact of under parameterising the Bessel function basis set by removing higher frequency modes. 21 radial frequencies is the natural choice for the fibre considered, hence all values less than 21 give an under parameterised model.

Model	Radial Frequencies	# Modes	Train Loss	Test Loss
Bessel Equivariant	21	1061	0.0141	0.0139
Bessel Equivariant + Post Proc	21	1061	**0.0032**	**0.0032**
Bessel Equivariant	14	932	0.0158	0.0156
Bessel Equivariant + Post Proc	14	932	0.0036	0.0036
Bessel Equivariant	7	567	0.0203	0.0201
Bessel Equivariant + Post Proc	7	567	0.0053	0.0053
Bessel Equivariant	4	332	0.0299	0.0296
Bessel Equivariant + Post Proc	4	332	0.0090	0.0090

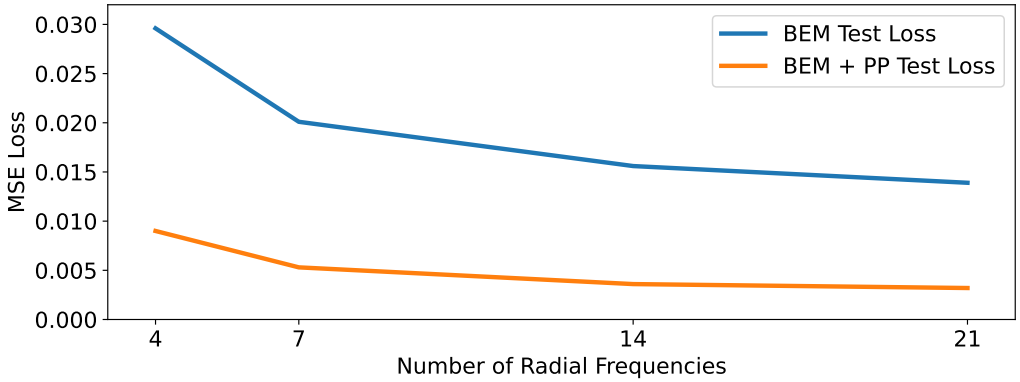
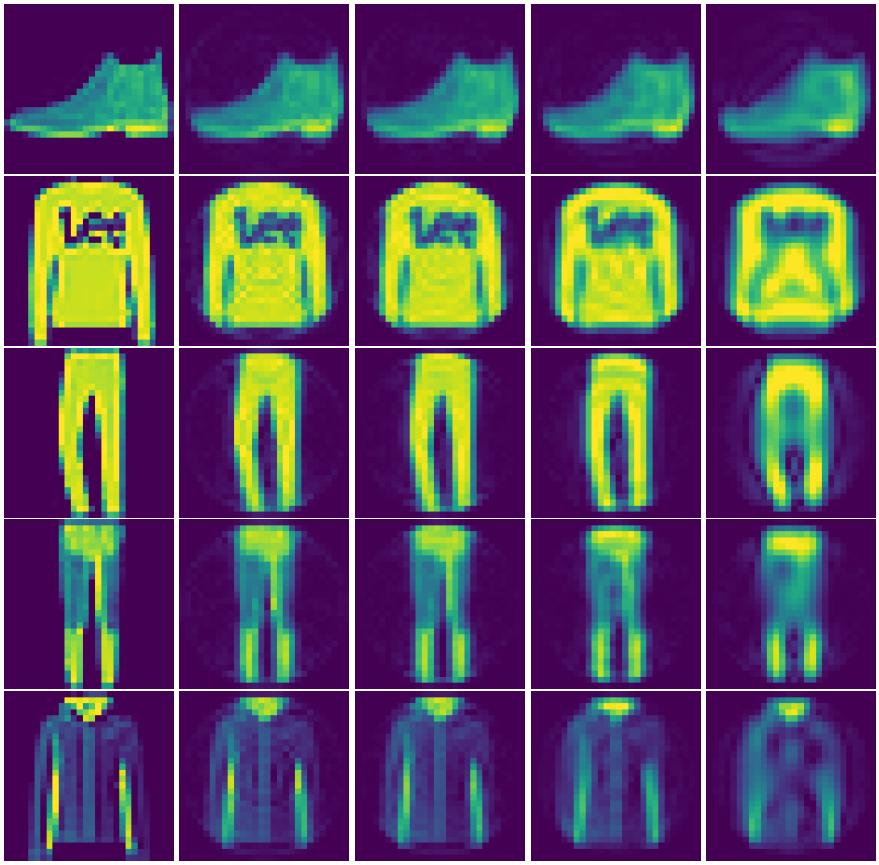


Figure 25: Comparison of the loss values of each model trained with fMNIST data created with a theoretical TM. This shows the impact of under parameterising the Bessel function basis set by removing higher frequency modes. 21 radial frequencies is the natural choice for the fibre considered, hence all values less than 21 give an under parameterised model.



(a) Target (b) BEM (21) (c) BEM (14) (d) BEM (7) (e) BEM (4)

Figure 26: Comparison of predicted images from inverting transmission effects of a MMF, when reducing the number of radial frequencies present in the Bessel function basis. (a) The input speckled image. (b) The number of radial frequencies present in the Bessel function baiis was left at the original value of 21, which represents 1061 bases. (c) The number of radial frequencies present in the Bessel function basis was reduced from the original value of 21 to 14, which represents a reduction in the number of bases from the original value of 1061 to 930. (d) The number of radial frequencies present in the Bessel function basis was reduced from the original value of 21 to 7, which represents a reduction in the number of bases from the original value of 1061 to 567. (e) The number of radial frequencies present in the Bessel function basis was reduced from the original value of 21 to 4, which represents a reduction in the number of bases from the original value of 1061 to 322. The data used was fMNIST with speckled patterns created with a theoretical TM.

A.14 Negative Societal Impacts

Our work enables imaging using a multi-mode optical fibre at higher resolutions than has been previously been achievable with a machine learning based approach. In addition, our new method generalises better to new data classes out of the training data classes. Therefore, it better enables imaging using multi-mode fibres to be used practically. Thus, this technology could potentially be used in a way that negatively impacts society through any negative use case of multi-mode optical fibre imaging (e.g. possible uses in espionage). Never-the-less the technology developed during this paper does not have a direct negative impact.