Artificial Neural Networks for Scattered Light Imaging

Piergiorgio Caramazza

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Physics and Astronomy
College of Science and Engineering
University of Glasgow

November 2019

Abstract

Image formation is one of the most important aspect of our everyday life. Conventional optical Imaging (and Sensing) exploits light, reaching the detection system from a target or a scene of interest, mainly unscattered. However, there are many practical situations in which unscattered light may be undetectable, insufficient or mispresented. Nonetheless, if the considered system allows it, it could be still possible to exploit scattered light in order to extract relevant information. Problems arise from the fact that, in these cases, light propagation may undergo severe alterations, thus leading to challenging, and sometimes ill-posed, problems.
In this thesis, two main scenarios involving scattered light are studied and addressed by means of artificial neural networks. Over the last period, these powerful data-driven algorithms have been extensively employed in many scientific contexts for their ability to solve even complex problems implicitly. Precisely this characteristic is exploited, in the present work, in a non-line-of-sight scenario in order to simultaneously locate and identify people hidden behind a corner. Moreover, a complex-valued neural network algorithm is implemented and applied to the problem of transmission of images through a multimode fibre, demonstrating high-speed and high-resolution image restoration even without the need for any phase measurements. Finally, due to its formulation based on the physics of multimode fibres, a direct comparison is proposed between the same algorithm and a more standard approach.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my supervisor Daniele Faccio for all his support and his interesting humanity. I would like to thank also my second supervisor Roderick Murray-Smith for the fruitful collaboration and the time dedicated to our meetings.
I would like to express my gratitude to Dr. Kali Wilson, Dr. Alessandro Boccolini, Dr. Catherine Higham, Dr. Alejandro Turpin and Dr. Yoann Altmann for their guidance, support, collaboration and the time invested in our scientific discussions, without which all the work carried out during this PhD would not have been possible. I would like to thank also Oisín Moran for his collaboration and Dr. Hugo Defienne for the fruitful discussions.

Declaration

The work contained in Chapters 2, 3 and 4 was carried out by the author. At the beginning of these chapters there are sections dedicated to the review of previous works, that serve as introduction. Chapter 1 contains introductory material only.

The work presented in Chapter 2 has been published in [1]. The experiment was designed and performed along with Dr. Alessandro Boccolini, who has also participated in the analysis and interpretation of data. Within this chapter, the analysis relative to the last two figures (Fig. 2.10 and Fig. 2.11) was carried out by Murray-Smith's group (in particular, by Dr. Catherine Higham).
The work presented in Chapter 3 has been published in [2]. The neural network was implemented by Oisín Moran. Within this chapter, two figures are adopted from [3] and one from [4], as indicated in the manuscript.
The experiment presented in Chapter 4 was designed and performed along with Dr. Alejandro Turpin, who has also participated in the analysis and interpretation of data.

The codes (and the relative neural network architectures) illustrated in Appendix A, published in [1], and Appendix C, published in[2], have been originally conceived and implemented by Murray-Smith's group, then customized and employed by the author. The network shown in Appendix A has been employed in Chapter 2. The network shown in Appendix C has been employed in Chapter 3 and Chapter 4.
The analysis proposed in Appendix B was carried out entirely by Murray-Smith's group. Experimental data were provided by the author. The results have been published in [3].

List of acronyms

AI . . . . . Artificial Intelligence

ANN . . . . . Artificial Neural Network

CCD . . . . . Charge-Coupled Device

CMOS . . . . Complementary Metal–Oxide–Semiconductor

CNN . . . . . Convolutional Neural Network

CPU . . . . . Central Processing Unit

CUDA . . . . . Compute Unified Device Architecture

CUP . . . . . Compressed-Sensing Ultrafast Photography

CW . . . . . Continuous Wave

DDM . . . . . Data-Driven Modeling

DMD . . . . . Digital Micromirror Device

DNN . . . . . Deep Neural Network

FWHM . . . . Full Width Half Maximum

GPU . . . . . Graphics Portable Unit

GRIN . . . . . Graded-Index

iCCD . . . . . intensified Charge-Coupled Device

IRF . . . . . Impulse Response Function

LHC . . . . . Large Hadron Collider

MCP . . . . . Micro-Channel Plate

MMF . . . . . Multi-Mode Fibre

MNIST . . . . Modified National Institute of Standards and Technology

MSE . . . . . Mean Squared Error

NA . . . . . Numerical Aperture

NLOS . . . . . Non-Line-Of-Sight

PBS . . . . . Polarizing Beam Splitter

PCC . . . . . Parson Correlation Coefficient

PMD . . . . . Photon Mixer Device

PMT . . . . . Photo-Multiplier Tube

PReLU . . . . Parametric Rectified Linear Unit

RAM . . . . . Random Access Memory

RMT . . . . . Random Matrix Theory

ReLU . . . . . Rectified Linear Unit

SLM . . . . . Spatial Light Modulator

SPAD . . . . . Single-Photon Avalanche Diode

SSE . . . . . Error Sum of Squares

SSIM . . . . . Structural Similarity Index

STAMP . . . . Sequentially Timed All-Optical Mapping Photography

STEAM . . . . Serial Time-Encoded Amplified Imaging/Microscopy

TCSPC . . . . Time-Correlated Single Photon Counting

TM . . . . . Transmission Matrix

Introduction

Recently, the technology advancements and the development of novel mathematical optimisation methods have revitalized the interest towards artificial neural networks. Thus, current computer performances allow to explore a new route for the investigation of physical reality: no longer exclusively based on the formulation of explicit mathematical laws, but on the identification of statistical correlations among the data that represent a certain phenomenon of interest. In a sense, a network is programmed to "make sense" of the observed data by formulating an internal, and mostly inaccessible, representation of the phenomenon itself. Typically, this makes the possibility to draw a reliable interpretation for the results delivered, and for the procedure followed to obtain them, more complicated.
The work presented in this thesis aims to contribute to the research area of imaging and sensing in presence of scattered light by employing artificial neural network algorithms. The general approach followed here is to rely on neural networks aided by the physics involved in the problems addressed. In this sense, the attempt is to improve, on the one side, the performances of the network itself and, on the other one, the chances to verify the physical relevance of the obtained results, whenever applicable. In the following outline I will indicate, chapter by chapter, every subject matter.

0.1 Thesis Outline

In Chapter 1, I will offer an introduction to artificial neural network algorithms. Firstly, I will retrace the history of the development of these algorithms. Then I will illustrate the basic principles of artificial neural networks, how these are structured and what the relevant elements are (for the purposes of this thesis in particular). Another section, in this way, will be dedicated to how to implement them on a computer, relying on the programming language "Python". Finally, I will emphasize the importance of this powerful tool in the scientific panorama, with a short overview focused on neural networks applied to Science and Imaging in particular, thus describing the general context of this thesis.

In Chapter 2, I will apply neural network algorithms to the non-line-of-sight imaging context. Prior art techniques will be reported and, within this context, a method for identifying people

hidden behind corners will be proposed. Here, the idea is to study single temporal signals, i.e. histograms built collecting, from a wall, the pulsed light back-scattered from the bodies, using a neural network for classification purposes. Thus, this experiment will provide an implicit solution to a problem for which an explicit forward model (and its respective inversion) would require a far too sophisticated description.

In Chapter 3, another ill-posed problem is considered switching from the temporal domain to the spatial domain. The problem addressed will be the transmission of images through a multimode fibre, without any phase measurements involved. As it will be shown in the introduction to the physics of multimode fibres, since the propagating light field is complex, phase information is typically essential to control light at the output of a fibre or for image restoration. Nonetheless, the approach proposed here will still be able to retrieve the input images at high speed (limited by the light modulator device) and high resolution. This result is made possible thanks to a complex-valued neural network.

Chapeter 4 will investigate further on multimode fibre imaging, by comparing the introduced complex-valued network to a commonly adopted technique for complex media, i.e. the transmission matrix. In this case, phase measurements will be required and will be acquired thanks to a phase-shifting interferometry method. This will also give the opportunity to study the behaviour of the complex-valued algorithm when trained with full complex field input. As a result, I will report the comparison between the performances of three different approaches: transmission matrix inversion, complex-valued network trained with full-complex field and complex-valued network trained with amplitude-only measurements (analogous to the ones used in Chapter 3). Furthermore, the singular value decomposition will be considered as a tool for extracting possible physical information from the transmission matrix and the complex-valued network approach.

Finally, Chapter 5 will draw the conclusion to this thesis leaving the possibility also to discuss possible interesting directions for future research. In Appendix A, the code of the neural network employed in Chapter 2 will be presented. In Appendix B, the first approach followed for the restoration of images transmitted through a multimode fibre, employing the complex-valued neural network method and relying on intensity-only images, will be reported. Finally, Appendix C will describe how to implement the complex-valued neural network used in this thesis.

Chapter 1

A whole data-driven world

Artificial Intelligence (AI) is shaping and deeply changing modern society along with the way we acknowledge reality. From common tools that we use in our daily life, like apps for calculating the best route to reach a certain destination, or "suggestions" offered by social media, to the ability to predict things like weather or market trends: nowadays many tasks and services are generally assigned to AI algorithms. Recently, the general excitement around the possibilities opened up by AI, along with all its achievements, has pushed many major companies around the world to heavily invest in this technology and, therefore, has contributed to AI's momentous development. In a sense, as it has been said, "AI is the new electricity" [5], such that juridical, international, political, and philosophical implication have started to be seriously taken into consideration.

Yet, it is not straightforward to formulate a precise definition of what AI is. A broad definition could be "using a computer to mimic human behavior in some way" [6]. Generally speaking, AI could be subdivided into two main categories: "model-driven" and "data-driven". In the first case, the approach is to explicitly define models and rules to gain knowledge related to a certain task. In regards to the second, assignments are solved implicitly by letting the system learn from a large amount of data.

Consider the following example: if a person is playing the piano and wants to compose a melody, on what basis should this person press a certain key instead of another? One might consider two possible ways, which are opposite to one another. For the first option, the subject could listen to a huge amount of related compositions and, by playing them, could develop a sense of right and wrong to be applied to every note to be played in the future. For the second, the person could rely on, or develop, some explicit musical knowledge such as harmony or scales, so that the action of pressing a certain key is guided by the "environment" that the person is exploring. In the first case, the person would be relying on a very well trained pitch to compose the melody. Whereas, in the second case, the person is relying on a musical theory. Although it is not impossible to think to compose complex pieces using the former way, it is evident that in this case, the latter one is preferable. Thus, this simple example serves to indicate why a model-driven approach is

preferred when a fair amount of decisions, related to a certain environment, needs to be taken. On the other hand, there are situations for which it is difficult to indicate few formal rules that should help us with a decision. For example, consider an image representing a cat and another one representing a dog. There are a set of possible features that characterises the two animals which can be specified in order to distinguish them: eyes, ears, etc. However, by adding more animals, other features will be required: each one with an appropriate mathematical formulation. Therefore, with many practical cases, adding exceptions or rules would not be the best strategy. On the contrary, one might build a data-driven algorithm that learns to develop its own features by looking at a large number of related samples. As has been demonstrated, the best strategy depends on the particular task at hand. However, it may be misleading to characterise these strategies as strictly separated. In fact, there are scientific studies [7] that provide evidence of how both the approaches can run and be exploited in parallel by our brain.

Currently, data-driven AI is receiving most of the attention for two main reasons: the unprecedented huge amount of available data and the rapid growth of computing power. In fact, many of the concepts underlying data-driven algorithms have already been implemented in the past decades. However, it is only with the recent explosion of modern computer graphic cards (GPU), which allow for powerful parallel calculations, that some data-driven algorithms can now be processed in a reasonable time frame. On the other hand, this modern digital era is indeed an extraordinary source of data. The term "big data" is not an exaggeration. The estimated worldwide daily data production is around 2.5 quintillion bytes [8]. Furthermore, it has increased the availability of "structured data", meaning the data is organized in some, possibly open source, databases or formatted repositories. Typical examples are the MNIST database [9] containing 60000 training images and 10000 testing images of hand-written digits, from "0" to "9", that come with the appropriate labels, or very large repository of images like ImageNet [10]. Structured data plays an important role in so-called "supervised" and "unsupervised" algorithms where, for the former, the learning process is shaped by a continuous interaction between the algorithm prediction and the ground truth (e.g., a label for classification tasks). Whereas, for the latter, it is not constrained by a specific output.

For the purpose of this thesis, I will be focusing on a specific area of data-driven AI: the artificial neural network (ANN) algorithms. In this chapter, I will introduce the basic concepts of artificial neural networks, underlying the parts that are more inherent to the development of the original work reported. I will show examples of usage of these algorithms on a computer through the Python programming language and, in addition, the impact of the neural networks on the scientific research. The aim of the following sections is not to provide a comprehensive description of the history of ANNs. Instead, I would like to shed a light on the key elements and challenges addressed by ANNs. Many of these concepts are indeed fundamental for the comprehension of

the neural networks implemented in this thesis.

1.1 Neural network principles and architectures

Artificial neural networks were born with the idea of mimicking the biological brain. The starting point was to introduce a model for the "units" which compose the brain, the neurons. Back in 1943, McMulloch and Pitts [11] proposed a simplified mathematical model for the neuron. This was built in order to receive binary inputs and produce a single binary output. Moreover, this system was demonstrated to be able to perform some basic logic operations. The reason why input and output are binarized is because of the observation that the output of biological neurons are, typically, spikes. Therefore, in this representation, neurons can only fire or not.
During 1958, an evolution to this model was proposed by Rosenblatt [12] and is known as the "perceptron". Here, the idea was to allow the neuron to learn the function that maps inputs into output (e.g. logic operations for the previous model) instead of implementing them explicitly. Before introducing this learning process, I show in Fig. 1.1 a representation of an "artificial neuron" and how this relates to its biological counterpart. This representation is merely a raw approximation of a neuron cell which, in nature, can appear in a number of different varieties. In this model, a neuron receives the sample input information x, with x_i indicating the i-th element of the vector x, from the connected neurons by way of their axons. I will refer to the j-th sample with the superscript $x^{(j)}$. The synapses contain the weights w_i to be applied to the inputs. Then, thanks to the dendrites, all the contributions from different inputs are fed to the cell body where the weighted sum of the inputs is performed (typically a bias value b is added). Finally, the "activation" function f, which for Rosenblatt was a simple step-function, is applied to the weighted sum in order to produce the binary output. I will refer to this output as $\bar{y}$ and to the ground truth as y (both are scalars).

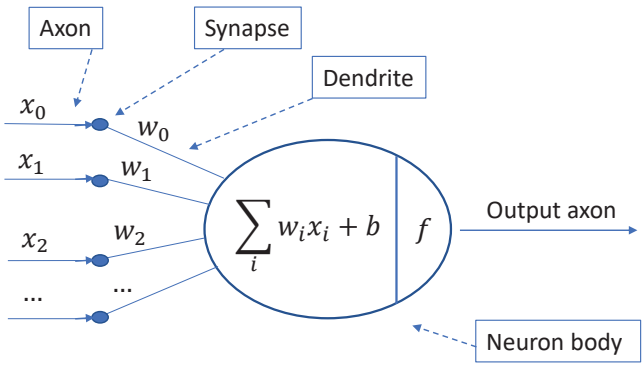

Figure 1.1: Representation of the perceptron that highlights the concepts taken from biology in order to build a model of a neuron. During the learning process, after each epoch the weights are updated with a correction related to the error estimated between output and ground truth.

The learning algorithm can be summarised in four points:

1. initialise the weights to random numbers (or to 0).
2. Calculate the output $\bar{y}^{(j)}$ of each input sample $x^{(j)}$.
3. For each sample, simultaneously update the weights.
4. Repeat steps 2-3 until there are no mistakes.

The weights are updated through: $w_i = w_i + \Delta w_i$, where Δw_i is defined by the perceptron learning rule. For the j-th sample, we will have:

$$\Delta w_i = \eta (y^{(j)} - \bar{y}^{(j)}) x_i^{(j)} \tag{1.1}$$

with η being the learning rate. To give an idea of how 1.1 works, it can be noticed that when $y^{(j)} = \bar{y}^{(j)}$ then the increment is 0, as the prediction is correct. Whereas, whenever the prediction is incorrect the weight receives a contribution directed toward the ground truth (positive or negative), proportional to $x_i^{(j)}$.
One of the main problems of this algorithm lies with point 4. In fact, it was demonstrated that the perceptron is able to perform operations like AND or OR, but not XOR, as pointed out by Minsky and Papert [13]. It turns out that the perceptron can subdivide correctly the samples in two classes only if they are linearly separable. Therefore, when this condition is not verified, the system will remain stuck in a never-ending updating loop.
Interestingly, Minsky and Papert indicated a possible solution to tackle the XOR problem: a network with multiple layers of perceptrons. However, here the limits of the perceptron arise. In fact, the learning rule puts in relationship the weights applied to the input array directly to the output value, with no information about possible intermediate layers. Thus, if there happened to be more than one layer, how could the weights from all the layers have been modified?

1.1.1 The road toward multiple-layers neural network

As was previously shown, the first problem, when trying to build up a multiple-layers network of perceptrons, is to provide a novel and more appropriate learning rule. From this point of view, the ADAptive LInear NEuron (Adaline) [14] model is useful as it allows me to introduce the concept of "cost function" for a neural network. In Fig. 1.2, the main difference with respect to the perceptron model can be appreciated: the activation step-function is replaced by a first linear activation function, $f(s) = s$ with $s = \sum_i w_i x_i$, and a quantizer , i.e. a step-function. Thus, in this case the error could be calculated on a continuous function (before the quantizer produces the binary output). If the activation function is not only continuous but also differentiable, then the

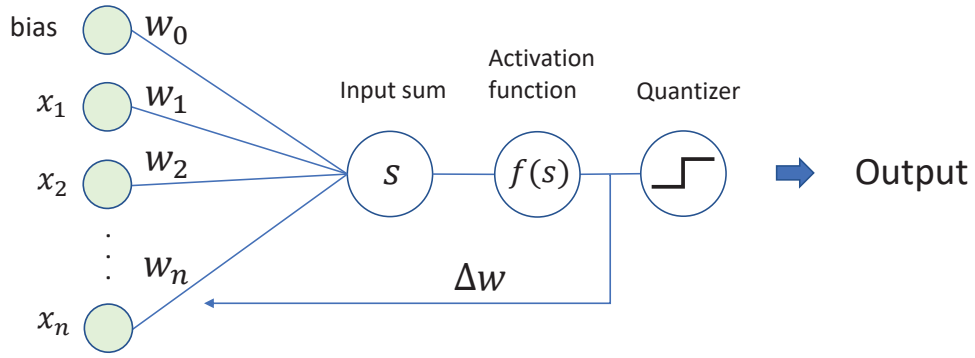

Figure 1.2: Scheme of the Adaline model. This evolution of the perceptron differentiates from its predecessor as the error is estimated before the step-function (quantizer), used to binarize. In fact, it is applied to the activation function that in this case is just linear: $f(s) = s$.

cost function could be defined as:

$$J(w) = \frac{1}{2}\sum_j \left(y^{(j)} - f(s^{(j)})\right)^2. \tag{1.2}$$

Basically, this is the sum of square error (SSE) between the ground truths and the predictions from the activation function. Here, w is the vector of the weights w_i. The basic idea is to minimize this cost function, i.e., the prediction must be as close as possible to the ground truth. Since this function is typically convex, and differentiable because of how the activation function has been defined, one might resort to the "gradient descent" algorithm in order to calculate the set of weights that minimize the cost function. Using convexity, one would aim to move in the opposite direction of $\nabla J(w)$, therefore, the increment to w will be defined as:

$$\Delta w = -\eta \nabla J(w) \tag{1.3}$$

But since $f(s)$ is linear, it can be easily calculated that:

$$\Delta w_i = -\eta \sum_j \frac{\partial}{\partial x_i} J(w) = \eta \sum_j \left(y^{(j)} - f(s^{(j)})\right) x_i. \tag{1.4}$$

As it can be observed, this learning rule looks pretty much similar to the perceptron one. Moreover, also in this case all the weights are updated simultaneously. However, I would like to remark that in (1.4) $f(s^{(j)})$ is a real number and, moreover, the updates Δw_i are not calculated after each training sample $x^{(j)}$, but rather with respect to the whole training set.
Since this rule could become quite computing-intensive when dealing with a large amount of data, other methods have been proposed such as the "stochastic gradient descent". This could be considered an approximation of the gradient descent. In fact, the learning rule is analogous but it is now calculated after each sample:

$$\Delta w = \eta \left(y^{(j)} - f(s^{(j)}) \right) x^{(j)}. \tag{1.5}$$

Since the weights are updated "on-line", usually input samples $x^{(j)}$ are shuffled after each epoch of the training process in order to avoid getting this process stuck. On the other hand, the continuous update of the weights is advantageous because it leads to a faster convergence for the algorithm. Moreover, the learning rate η is typically varied dynamically. However, since the cost function trend over the epochs is usually quite irregular, there is another popular variation in between the two techniques, called "mini-batch" which allows for the calculation of that gradient over a series of subsets of the total sample dataset.

What is the advantage with this new learning approach? The fundamental characteristic I have been looking for is the possibility to let all the weights, from each layer of the network, be corrected by the learning rule. Consider the scheme in Fig. 1.3: here, the first hidden layer could be written as $f(W1 \cdot X^T + \mathrm{B1})$, the second one as $f(W2 \cdot (f(W1 \cdot X^T + \mathrm{B1})) + \mathrm{B2})$, and so on - note that in this case W is a matrix, as there is more than one neuron in each layer. Thus, since the methods I have just defined allow the error of every single weight to be calculated thanks to a partial derivative, the chain rule can be finally applied. There are a number of techniques that can implement this type of calculation efficiently. However, the most popular is called "back-propagation" [15]. As the name suggests, the idea is to propagate the error backwards, thus preferring to implement less-expensive matrix-vector multiplications, starting from the output vector, than matrix-matrix multiplication, starting from the input.

It might be noticed that the network, built in this way, is not strictly composed by layers of perceptrons. Rather, it makes use purely of activation functions, passing from one layer to another, whereas the quantizer is employed just at the very end, when it is required. Since every neuron in each layer is connected to each neuron of the following layer, this kind of architecture takes the name "fully-connected" or, more simply-stated, "dense".

It can be easily demonstrated that when using a linear activation function hidden layers are redundant, as this would result in a series of linear operations. This introduces the need for a non linear activation function whenever hidden layers are required. Generally speaking, the choice could fall on any non linear derivable function. On the other hand, following up from the biological neuron insights concerning how neurons communicate (i.e. typically with spikes), a function similar to the step-function would be preferred, such as the sigmoid or the hyperbolic tangent. Addressing the problem of a multiclass classification (i.e. classification for a network with multiple outputs such as the one outlined in Fig. 1.3), the so called "softmax" activation function [16] - a sigmoid-like function - is the preferred one as it maps the output in a probability distribution, thus avoiding negative numbers. Whenever a sigmoid-like activation function

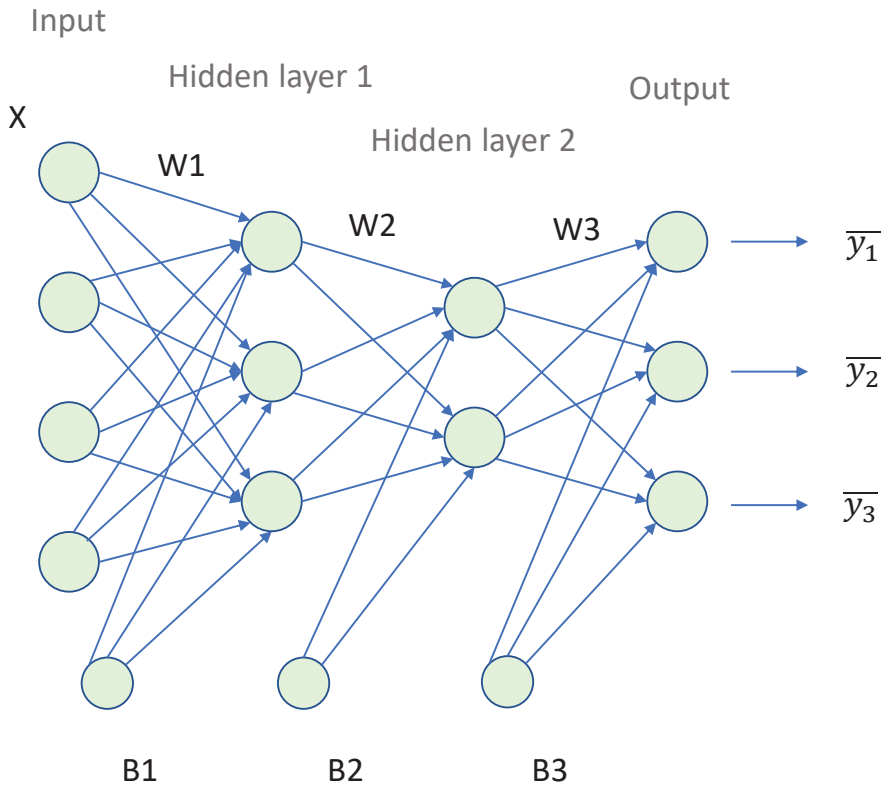

Figure 1.3: Example of multi-layer network. In particular this network presents two hidden layer and all the neurons (units) belonging to a specific layer fully connected with all the neurons of the neighbouring layers. The three output units allow a non-binarized output. Moreover, the error can be propagated through the network because, using continuous and derivable activation functions, the chain rule can be applied by expressing every layer as a function of the layer before.

is employed, cost functions such as SSE or mean squared error (MSE) should not be used because they would typically lead to a non-convex function with many local minima, thus causing difficulties for gradient descent techniques. For these situations, the choice usually falls on the so-call "cross-entropy" [17], which could be "binary" whenever considering just two classes or "categorical" for multi-class prediction.

Finally, a typical problem related to neural networks is overfitting, i.e. the incapacity of the identified representation, calculated through the leaning process ability and which might perform well on the training dataset, to generalise the result to unseen data. To overcome this, one possibility is to add a regularisation term to the cost function. For example, the norm l_2: $\mathscr{L}(w) = (1/2)\lambda||w||^2$. In this case, the idea is to help the system to learn small weights when the parameter λ is large, or, vice versa, to prefer the minimization of the cost function for small λ.

To summarise, I have retraced the development of neural networks starting from the single neuron model, the perceptron. The limits related to this model have been appreciated and the

concept of cost function, which broadens the possibilities of the learning process, has been introduced, thus paving the way to more complex networks with multiple layers of neurons. Networks with more than one hidden layer take the name of "deep neural networks" (DNNs), in this case the term "deep learning" is usually adopted.

1.1.2 Convolutional layer

In this section I will introduce another popular kind of layer, especially for DNNs, referred to as "convolutional". As the name suggests, the idea is to learn some features by convolving a series of filters to the input sample. Obviously, the size of the filters has to be smaller than the size of the sample itself. Referring to Fig 1.4, I will indicate the result of the convolution of each filter as a "feature map". After being calculated, these features maps are stacked along the depth dimension. Each element of this output volume is therefore connected to a small portion of the input sample, defined by the size of the filters. Interestingly, this concept of local connectivity is borrowed also from Biology, in particular, the architecture of the visual cortex of monkeys [18].

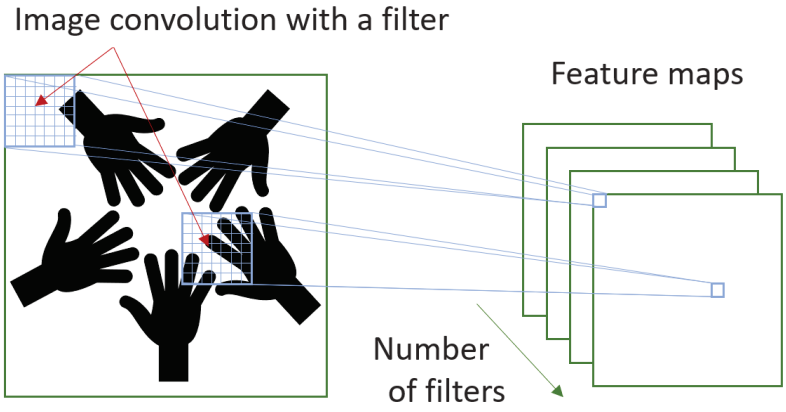

Figure 1.4: Representation of a convolutional layer operation. The input image is convolved with a series of filters. The output volume contains the resulting feature maps arranged along the depth dimension, whose size is determined by the number of filters.

Because of the different local information stored on different neurons, the network can now valorize the parameters relative to specific spatial points in the input. In fact, these types of layers are used specifically for finding features independently of scaling factors or position, which is exactly what one should be looking for in order to allow the network to gain "vision". On the contrary, fully-connected networks employed for the classification of images, for example, require images to be centered. Finally, since the convolution operation is essentially sliding the same filter to obtain one "slice" of the output volume, then the slices share the same filter weights for each position. It is precisely this characteristic which reduces the number of parameters, therefore improving the efficiency of the algorithm. To further reduce the number of parameters, "pooling" (or "subsampling") layers are typically employed between two convolutional layers.

Indeed, the job of the pooling layer is to sub-sample the convolutional output (i.e. it compresses the extracted features). This operation could be obtained by saving the maximum of each sub-sampled zone of the original layer, referred to as max pooling, which is the most popular of the techniques, or by calculating the average.
This type of architecture takes the name of "convolutional neural network" (CNN) and it is a typical example of a deep learning algorithm. Because of all the reductions of parameters involved, this deep neural network is typically lighter than the fully connected one. Finally, in Fig. 1.5, an example of image classification with a convolutional architecture is shown (credits to [19]), which is the task for which these kinds of networks became particularly famous [20].

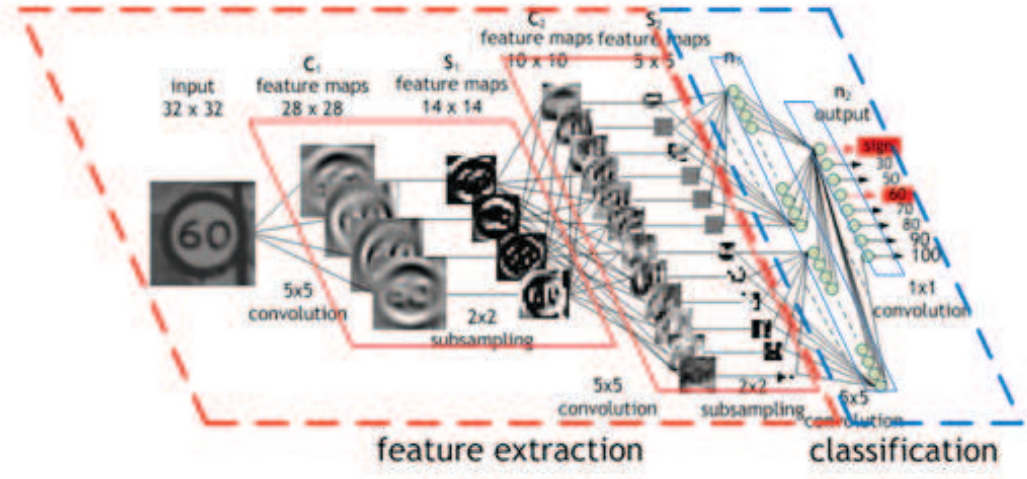

Figure 1.5: Example of classification for speed limit traffic signs. This deep network exploits a number of convolutional layers typically followed by a pooling layer in order to subsample. Figure source [19].

1.2 Using Python for neural networks

Python has been one of the main programming languages for implementing and developing neural networks. Generally speaking, a number of libraries and toolkits tailored for building and training neural networks can be easily found. Basic libraries are available, such as PyBrain [21] and Scikit-learn [22], which can provide the fundamental ANN elements introduced in the previous sections or basic ready-to-use layers and networks. However, I will be focusing on those that rely on AI frameworks created for the purpose of high performance numerical computation. The idea here is to exploit parallel computation to speed up the training process. As a matter of fact, ANNs simply imply a series of similar and repetitive operations. Indeed, Graphical Processing Units (GPUs) are built in such a way that makes it particularly easy to perform this kind of calculation, since the architecture is structured in a large number of simple cores and can easily run simultaneous threads. Therefore, with the recent rise of GPUs, frameworks were developed in order to optimise mathematical expressions when dealing with operations on large

and multi-dimensional arrays.
That being said, I opted for the library "Keras" [23] which is probably the most diffused option for artificial neural networks. Keras can be run on both frameworks Theano [24] and Tensorflow [25], although Theano is currently mostly abandoned. Throughout this thesis I used Tensorflow, which can also be imported in Python as a library. Both of the frameworks allow to run the calculation either on CPU or GPU, nonetheless, for the work of this thesis, a GPU card was used exclusively. More specifically a Nvidia GeForce 1080Ti with 11 Gb of RAM memory. Actually, RAM is one of the main limitations when dealing with neural networks as it turns out that it is not very difficult to fill the memory when, for example, using large fully-connected layers. Indeed, the considered GPU is just a gaming graphic card, although still one of the best around. Other GPU architectures tailored for high performance calculation can be found and typically come within appropriate computers, but at high prices. Another way to extend the RAM memory is to employ multiple GPUs. However, it is not trivial to parallelize calculation for neural network on multiple GPUs. This is also true for Keras, which just recently started allowing this kind of approach.

Keras exploits the object-oriented nature of Python allowing the user to interface with standalone modules each one implementing a different part of the neural network: layers, activation functions and so on. These blocks are easy to configure and link, so that an entire complex ANN model can be written compactly in a few lines of code. If, for example, I am willing to implement the neural network shown in Fig. 1.3 and supposing I am dealing with a supervised classification task, having `X_Train` as a training dataset and `y_Train` as the corresponding set of labels, I could write:

Listing 1.1: Example of keras implementation for the network in Fig. 1.3

```
from keras.layers import Input, Dense
from keras.models import Model
from keras import regularisers

# paramenters
n_epochs = 500
lambd = 0.03
learning_rate = 0.001

# model
input_v = Input(shape = (X_Train.shape[1],))
```

```
x1 = Dense(3, kernel_regulariser =
   regularisers.l2(lambd))(input_v)
x1 = Activation('sigmoid')(x1)

x2 = Dense(2, kernel_regulariser = regularisers.l2(lambd))(x1)
x2 = Activation('sigmoid')(x2)

x3 = Dense(3, activation = 'softmax',
           kernel_regulariser = regularisers.l2(lambd))(x2)

model = Model(inputs = input_v, outputs = x3)

sgd = SGD(lr = learning_rate)
model.compile(loss = 'categorical_crossentropy',
              optimizer = sgd)

model.fit(x = X_Train, y = y_Train, epochs = n_epochs)

# prediction
pred_test = model.predict(x_Test)
```

Starting from the beginning, I firstly must import from the Keras library the modules in which I am interested. Then, some parameters are defined, such as the number of training epochs. Consequently, I build up and train the model. Tensorflow works with tensors only, thus I first call `Input` to define the input tensor `input_v`. A layer can just operate on a tensor and gives another tensor, therefore I have to make explicit that the first `Dense` layer has to operate on `input_v`. The result of this operation is tensor `x1` which is composed by 3 neurons (units). In addition, another argument is passed in order to add the relative regularisation term (in this case, the norm l_2 is added for all three layers) and one more argument in order to choose the activation function. In accordance with what I discussed in the previous sections, the activation function must be non-linear for a multilayer network. In this way, the choice falls on the sigmoid activation function for the first two layer and softmax for the third (output) layer. The reason why softmax should be preferred lies in the fact that I was dealing with a multi-class classification task. In the same way, one should resort to an adequate cost function, in this case categorical cross-entropy, which is implemented in `model.compile`. This last call, as the name suggests, compiles our model which has been defined before with the call `Model`, that takes as arguments the input and output tensors. Finally, the stochastic gradient descent (`SGD` in the code) is used to minimize the cost function. At this point, after the training process is complete, the neural network can be tested on unseen data `X_Test` and compare the results to the original labels

`y_Test`.

In this example, only fully-connected layers were employed. However, many other kinds of layers are available. As an example, for convolutional layers one could use the call `Convolution1D( filters, kernel_size, padding = 'same', input_shape)(input_v)` in the 1D case, `Convolution2D` in the 2D case, and so on. The arguments that convolutional layers take must include the way in which the convolution is to be implemented including: the number of filters, the window size of these filters (i.e. `kernel_size`) and padding on the borders. Whenever a layer is not available, there is still the possibility to implement a custom one, as was done in this thesis. Furthermore, it is also possible to customize modules and different arguments to be passed to the layers, as for custom layer-initialisation and so on. Finally, a number of different types information can be extracted from these models. For example it is possible to visualise the cost function trend, over the training epochs, and the weights of the network can be saved along with the model itself.

To conclude, I would like to further stress the remarkable compactness and readability of a neural network model written with keras. As mentioned before, the Tensorflow framework is implicitly imported and exploited using keras. This means that if one is willing to exploit GPUs, as with Nvidias graphic cards, there is the possibility to program powerful parallel calculations for ANNs without the need to explicitly use a much more advanced language like "CUDA", built to program on GPUs, nor the Tensorflow language itself that works also on top of CUDA.

1.3 Neural networks in Science

Thus far, AI and in particular ANNs have had a tremendous impact on Science. One of the reasons for this is provided by the fact that there are many scientific fields for which there is an increasing amount of available data to such an extent that it is not possible for the scientists to keep up with all this information. For example, in Astronomy, this is the case with data collected by some telescopes or radioscopes. Therefore, on the whole the idea is to let these algorithms spot correlations that we are unable to highlight. In this sense, the range of possible applications is immense and crosses many different scientific areas. Particle Physics seems to be one of the first fields to which neural network were applied. Here, most of the experiments dealt with particle detectors recording a huge quantity of signals generated by different particles. Therefore, it is vital to recognize the signal of interest from the background. In this context, the clear advantage is that ANNs could lead to highly non-linear outputs, thus overcoming common approaches in terms of event selection. Firstly, ANNs were used in High-energy Physics for tracking reconstruction in some experiments [26][27]. Later, they started to be applied also to offline data analysis, as for example to the decay of Z boson [28], or the direct measurement

of the top quark mass [29]. In particular, LHC has heavily invested on ANN, with fundamental results obtained by extracting information from high-dimensionality data [30].
Biology is another field in which there is an abundance of data, with databases mainly focused on amino acids and nucleic acids. ANNs were successfully applied to predicting molecular characteristics from DNA sequences [31][32]. Once again, the non linear nature of these algorithms, such as DNNs, has provided more complete representations for some biological phenomena. Another advantage is that in many cases they can be applied directly onto the data without the need for pre-processing them in order to extract relevant features manually. Remarkably, DNNs, and in particular the ones involving convolutional layers, found fertile ground with biological image analysis. For example, they were successfully employed to detect and locate cells in microscopic images [33], remove noise from neural circuitry images [34] and classify non-membrane versus membrane pixels for neuronal structures [35].
Another relevant scientific contribution of ANNs is related to the ability to predict the evolution of certain systems. Kevin Schawinski et al. [36] exploited a "generative" neural network method, that had been previously implemented to guess the change in human faces aging over time, in order to study the properties of galaxies turning from belonging to a low-density region to being satellite galaxies. In the same context, but using a different method, Yoo-Geun Ham et al. [37] proposed a CNN architecture tailored to forecasting, with an ability to predict El Niño up to 18 months in advance.

Certainly, several papers in every scientific area - Chemistry, Neuroscience, etc. - could be found with tasks and achievements similar to the ones reported in the examples cited in this section. However, it is beyond the scope of this overview to review every type of ANN implementation within the Sciences. On the contrary, the idea was to shed light on why ANNs are growing in importance and playing a key role in modern Science. Nonetheless, it must be said that many scientists have welcomed these algorithms with caution. The reason for this potential controversy comes from the fact that even when ANNs lead to good predictions, and even when they outperform common techniques, the question around the impenetrability of the results remains. The recurrent term that is often heard in this environment is "black box". More specifically, an ANN cannot give an explanation of the procedure undertaken to reach a certain result. I avoid philosophical conjectures about black-box similarity with some aspects of our own brain. Instead, I underline the fact that, generally speaking, the only way to evaluate the job of a trained ANN is to examine the answer on the basis of the questions asked. Typically, explicit information underlying the generated function that maps input into output, such as an abstract formulation of a model or a procedure, are not accessible.
Interestingly, there are some architectures designed on purpose to directly learn an empirical model that describes a phenomenon of interest. These fall under the larger context of so-called "data-driven modeling" (DDM). A good example comes within physical dynamical systems [38]

[39], where ANNs help to discover the dynamical model that describes the system itself. What is remarkable with this kind of approach is the possibility to corroborate, either experimentally or using simulations, the physical validity of the obtained models. In this case, as well, the "procedure" to obtain a certain model is unknown, however, there might be room for an external verification. Quoting the authors: "With this explosion of interest [in DNNs applied to Science], it is imperative that we as a community seek machine learning models that favor interpretability and promote physical insight and intuition". Finally, in the same context, I would like to report the work from Raban Iten et al. [40]. Here, a neural network method is employed to discover physical concept from relatively easy (and known) physical systems. For example, the parameters that describes the pendulum motion or the conservation laws for particle collisions.

1.3.1 Imaging applications

The development of ANNs has been highly influenced by the possibility to solve tasks related to images. As a matter of fact, one of the first challenges was to provide a correct classification of hand-written digits. Moreover, it has been observed how the concept of vision in some animals has inspired the idea and realisation of convolutional layers. As a natural consequence, when these algorithms were made popular, thanks to the novel computational technology, they started to be successfully applied to many imaging problems. Indeed, possible applications have already been shown, within this section, when I highlighted the importance of ANNs for biological image analysis. Typical assignments, within fields like Biology, Neuroscience, and Medicine, are location and classification. For example, to evidence certain structures of interest for large amounts of images, or generally to improve the quality of images obtained with certain imaging techniques such as denoising or enhancing resolution.

With the advent of modern technology, imaging has become more and more dependent on electronics (i.e. cameras, devices to control illumination, etc.) to such an extent that nowadays it would sound bizarre to distinguish them separately. For this reason, computation has become of fundamental importance in order to conjugate and exploit all the factors constituting an Imaging experiment. In general, scientists refer to this evolution as "computational imaging" [41], beginning in the 1990's. In this area, the main goal is to optimise or maximize, with respect to some parameters of interest, the information that can be extracted from a designed experiment. Thus, computation could be particularly useful whenever it is not possible to achieve a certain result with common means, or in case a problem is ill-posed or whenever solutions could be found with less expensive means.

Within this context, a detailed review about the key role played by ANNs for computational imaging, in Optics, has been offered by Barbastatis et al. in [42]. Here, a schematic representation of a generic problem of computational imaging is introduced first. As reported in Fig. 1.6,

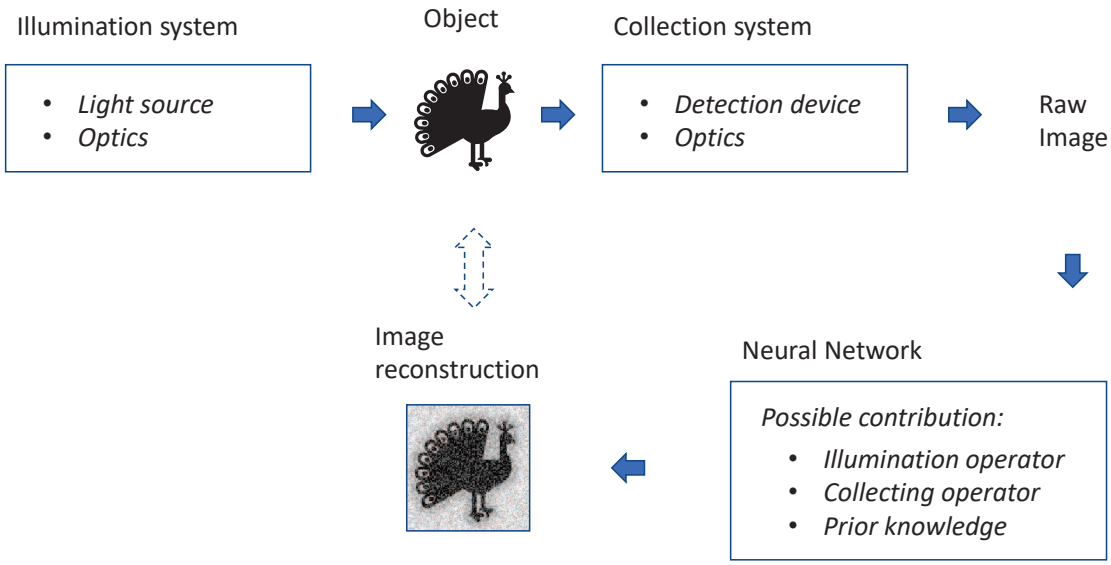

Figure 1.6: A generic computational imaging system, using a neural network to tackle the image reconstruction. The essential components are an illumination system, an object of interest, a collecting system that produces a raw image and a neural network algorithm. This could take contributions deriving from the operators relative to the illumination and collection processes, or prior knowledge about the considered system.

this is simply composed of an object of interest, an illumination system, a collecting system and the computation process that finally delivers the image, in this case a neural network. This illustration shows how the network takes as input the raw images (or signals for Sensing) in order to predict the image of the object. Therefore, the algorithm is typically trying to approximate the inverse problem. Moreover, the training process could progress in a continuous comparison between prediction and ground truth whenever an appropriate large input-output database is available. Finally, information about the system, such as illumination and collection operators, or prior knowledge, could be wisely exploited in order to shape the network in accordance to some physical and electronics aspects of the experiment, thus helping the system with finding a more convenient and realistic function that will map input into output.

Four main classes of possible applications, where ANNs have provided a crucial contribution, are identified:

- super-resolution;
- low-photon and strong noise imaging;
- retrieval of full-complex field;
- scattered light-based imaging.

Super-resolution comes into play for those ill-posed situations such as image upsampling or

deblurring (examples of ANNs related works are [43][44]). In addition, ANNs have led to important results for Imaging in dark scenes, or with a high-level of noise, where deconvolution techniques tends to struggle. Examples for low-photon imaging and denoising can be found in [45][46]. With respect to the third class, there is a whole set of different techniques that aims to restore the complex field on a certain plane of the system from a raw intensity image. However, for many practical cases, the non-linearity of certain ANNs architecture could deeply improve the reconstruction (i.e. whenever convenient assumptions on the system cannot be adopted). For instance, this is the case in digital holography techniques [47] or retrieval of a phase object (encoded on laser beam, propagated through air and collected with a camera) [48].
Finally, the fourth class includes all those systems that involve multiple light scattering, or simply strong scattering. Examples include complex media such as diffusive opaque materials or multimode fibres (due to imperfections, bending or other external agents). In these cases, the typical signature is a strong interference process (or a deviation from its ideal behaviour) that light undergoes propagating through them, sometimes leading to randomisation. Nowadays, these systems are widely studied because of their possibility to shape light at the output, transport information to not-easily accessible areas or exploit some correlation properties such as the memory effect. Recently, the interest to this field has grown progressively, particularly in the past few years, as a number of related papers that exploit ANN algorithms have been published [49][50]. Furthermore, there are other relevant scattering techniques such as the non-line-of-sight techniques (NLOS), where information about objects hidden behind a wall or around a corner is retrieved. A vast literature could be found on these approaches too, especially in recent times thanks to the crucial technological contribution of high-speed cameras. On the neural network side, the work of Satat et al. [51] exploits a CNN to classify human poses through a scattering material from time resolved measurement.

The fourth class also coincides to the part of computational imaging on which this thesis is based. In the next chapter I will introduce and discuss contributions of this work to this imaging field.

Chapter 2

Neural network algorithm for identifying and locating people hidden behind a corner

Obtaining information on objects of interest that do not fall within the visual field of view, because of the presence of obstructions, is an important challenge that has attracted the attention of many scientists over the course of time. Indeed, the demand for possible insights on such hidden objects or scenes is guided by, and might find application in, particular real-life situations such as rescue operations, self-driving cars, and surveillance. This particular field takes the name of Non-Line-of-Sight (NLOS) imaging or sensing, where the distinction imaging/sensing typically indicates what kind of information is extracted from the data, i.e. whether it is attempted to build specifically an image or not.

NLOS techniques could be basically subdivided in two main categories: looking behind the wall and looking around the corner. In practical terms, the main difference lies in the radiation employed. In fact, in order to overcome a wall the radio wavelength-range is required. However, this limits strongly the spatial resolution. Nonetheless, imaging or sensing through a wall has been demonstrated, for example in [52][53][54] .
In this sense, most of the research has been focused on looking around the corners. Even though detection is possible also through radio or acoustic waves, like with radar and sonar techniques, in this case there are no limitations related to the wavelength, such as for looking behind a wall. Thus, a better spatial resolution is in principle achievable, for example using visible light. However, electromagnetic radiation comes with the highest speed possible in nature. This might not be an issue when working with large wavelengths and thus with low spatial resolutions that correspond to easily accessible temporal scales; on the other hand, the ability to capture the motion of light, propagating through a medium, on the centimetre scale requires ultrafast detection systems with temporal resolutions of about 100 picoseconds.

Within the next sections, I will offer, firstly, a short overview of the most common detection

systems that can work at these high speeds. Then, possible phenomena and applications that have been studied thanks to the development of these powerful means will be introduced as well. In particular, the field that I will be considering takes the name of "light-in-flight" imaging, i.e. techniques where the motion of light is captured and exploited in order to extract information about the propagation itself or an illuminated object. In this way, the general context for the original work presented in this chapter will be illustrated, focusing on the specific problem of looking around corners, in order to better appreciate the novelty of the addressed challenge.

2.1 High-speed detection systems

When dealing with recording phenomena in the picosecond range, it is evident that no electronic or mechanical shutter can really work. Thus, it is necessary to rely on other acquisition techniques. Over the last three decades, but in particular over the last one, many different methods have been developed. A first distinction that might be highlighted is between "gating" methods and "continuous acquisition" methods [55]. For the former, some of the light is discarded in order to obtain a temporal sequence; for the latter, ideally all the light is collected and used for reconstructing the observed event.

There are a number of ways for creating an ultrafast shutter or gate. One way consists in creating a gating method that works in intensity. In this case, the high-speed element in the system is just the one selecting the right temporal intervals, whereas the resulting light is recorded by a low-speed imaging or sensor element. This introduces the need for a system that shifts, in time, the pulsed illumination respect to the selected temporal intervals. By scanning, the event of interest can be reconstructed. Even though these techniques are not generally efficient, in terms of the amount of light collected, it is possible to reach high temporal resolution.
Intensity gating has been achieved in different ways, for example through optical non-linearity, by time-stretching a spectrum-encoded image or resorting to a quick signal amplifier. Non-linear solutions were the first to be introduced in 1967 [56]. The light-matter interaction, for some particular material, can be exploited to create techniques able to achieve temporal resolutions of the order of picoseconds or even dozens of femtoseconds, for example by taking advantage of phenomena such as transient birefringence [57] or second harmonic generation [58].
Considering time-stretching approaches, the spatial information of an image is firstly encoded in the spectral domain; consequently, a dispersive medium is used to stretch the signal temporally in order to allow the detecting device to resolve this time-dispersed information. The idea of temporally dispersing the original signal was firstly introduced by Han et al. in [59]. Later, it was exploited for 1D and 2D imaging as well [60] [61]. In particular, the 2D imaging technique takes the name of "Serial Time-Encoded Amplified Imaging" (STEAM) and employs a fibre as dispersive medium. The same fibre is used to amplify the signal as well, allowing the capture

of a frame through just a single pulse. This is particularly important whenever dealing with non-repetitive events. For this approach, the temporal window achieved was of 440 ps, with a 163 ns frame rate. Another similar single-shot technique is the "Sequentially Timed All-optical Mapping Photography" (STAMP) [62], where the temporal resolution is significantly higher, 4.4 trillion frames per second, but limited to few frames. This method relies on separating the backscattered pulses, encoded with different wavelengths, spatially.

Amplified gating approaches resort to the introduction of a system able to amplify, selectively in time, a signal in order to achieve ultrafast acquisition. Surely, the intensified-CCD camera (iCCD) is the most common example of such imaging devices. Referring to Fig. 2.2, the iCCD presents two main components: the intensifier tube and the cooled CCD. Within the intensifier tube, a collected light photon (shown in red) is firstly converted into a photoelectron. Between the photocathode and the phosphor screen the photoelectrons are accelerated thanks to a potential gradient. Furthermore, the signal can be enhanced using a microchannel plate (MCP) by up to 6 orders of magnitude [63]. Then, photoelectrons are converted back to photons that can finally be collected by the CCD detector. ICCD models like PicoStar by LaVision have at their disposal gates of the order of hundreds of picoseconds. By varying and scanning the trigger delay between laser and camera one can easily achieve temporal resolutions on the order of dozens of picoseconds.

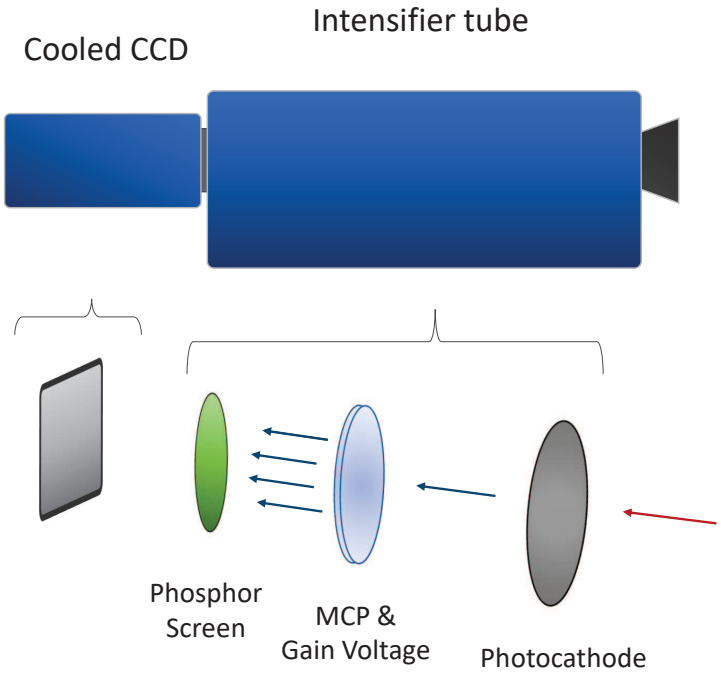

Figure 2.1: Schematic representation of an intensified CCD camera. Inside the main component of the camera, the intensifier tube, the collected photons (represented in red) are converted into photoelectrons (represented in blue) and accelerated by an electric field. The resulting signal is amplified and finally converted back into photons. Then, an image is formed on the CCD.

Gating methods have been realised by taking advantage of the coherence property of light as

well. For example, “white light interferometry” provides short temporal gates by exploiting the interference between an illuminating beam and a reference beam. In this case, the gating window coincides with the temporal window in which the two beams can actually interfere. Moreover, the shortness of this window is provided by the low coherence length of the two beams. This technique has been able to deliver a 100 fs temporal resolution [64].
Other coherence gating approaches rely on holographic methodologies, as for the one proposed by Abramson in 1978 [65]. In this case, coherent short pulses are employed to illuminate both the scene and a holographic plate, as a reference. By choosing an appropriate angle for this reference beam, the light reflected by the scene of interest interferes with the reference beam in different points on the holographic plate. This means that different zones of the plate contain different temporal information of the scene. Furthermore, short-gating is induced once again by the low coherence length. In fact, the two beams can interfere only if the distance they covered is the same.

Continuous acquisition approaches do not fragment temporally the observed event (by means of shutter or gates), on the contrary, as the name suggests, they attempt to collect the light scattered from the object of interest continually. In this case, the three main technologies are: Photon Mixer Devices (PMDs), streak camera and Time Correlated Single Photon Counting (TCSPC). The idea to avoid discarding part of the light implies that these methods are typically more efficient then the gated ones.
The approaches that exploit the PMD sensors [66] are probably the most inexpensive along with the coherent gating ones. These methods aim at inferring the time of flight of the photons bounced back from an object which has been illuminated with a modulated continuous wave light beam. To do so, multiple acquisitions are required - 2 in the absence of background light, more otherwise -, each of which corresponds to a different phase shift applied to the illuminating beam. Apart being cheap, these devices are commercially available - used for example for videogames - and can provide very high efficiencies; on the other hand, the limited sensitivity, because of the photodiode technology employed, limits its range of application to a few meters. As reported in [67], the temporal resolution can reach up to 70 picoseconds.
The components of a streak camera are similar to the ones of an iCCD. In fact, also in this case the camera disposes of a photocathode, a microchannel plate and a phosphor screen that, respectively, convert the photons into electrons, amplify the signal and converts is back to photons. Furthermore, a CCD or CMOS device captures the light after the phosphor screen. The difference lies in the fact that a slit allows only a line of the imaged scene to reach the photocathode, and, moreover, a field is applied to spatially deflect the generated electrons, before they reach the MCP. The idea is to keep the dimension parallel to the slit for the spatial information and the other axes for the temporal information. In this way, the deflections can be analysed in order to reconstruct the observed scene at ultrafast speed. Nowadays, this is the method with the best

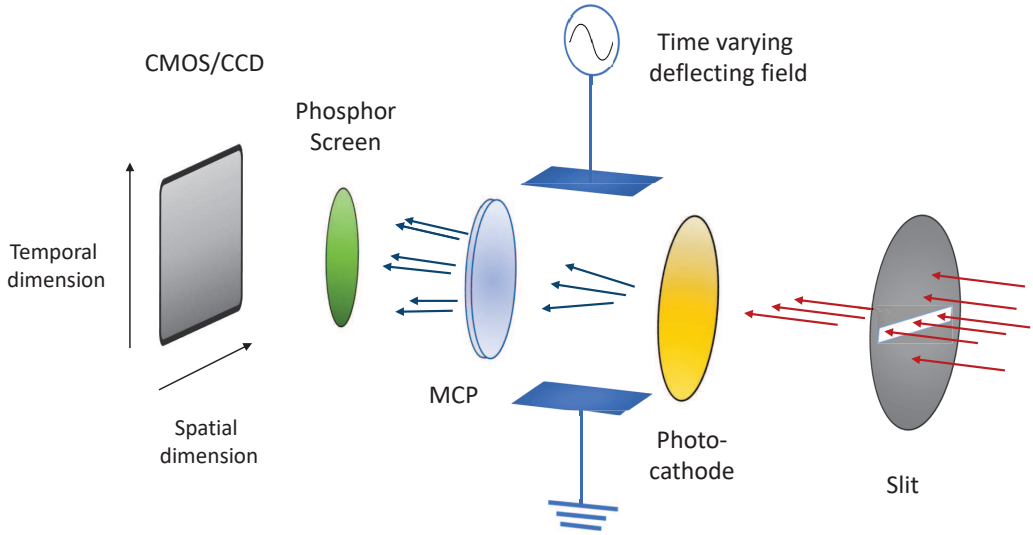

Figure 2.2: Streak camera internal scheme. This camera allows in just a line - horizontal in this example - of the total photons (indicated in red) employed to image a scene of interest. The photons are converted to electrons (indicated in blue) and deflected along the vertical axes by a field that induces a time-varying deflection on these electrons. The signal is then amplified and converted back to photons to be imaged on the sensor. Studying the deflections of the electrons it is possible to achieve impressive temporal resolutions.

performance in terms of temporal resolution. Commercially available cameras can reach a resolution of hundreds of femtoseconds, but, in principle, this is extendable to the attosecond domain [68]. Issues with the streak camera are related to portability, cost, and to the need for scanning one line at the time. However, different ways for circumventing this scanning issue have been proposed. It is worthy to mention the Compressed Ultrafast Photography (CUP) method [69]. Here, single-shot 2D acquisition is realised by resorting to compressive sensing techniques.

Finally, the last method I will overview is TCSPC. The aim of this technique is to provide a measurement of the arrival time of detected photons. A complete review of this method, its history and the related technologies, advancements and applications can be found in [70]. In order to implement photon counting one needs a trigger signal, as a reference to evaluate the time of arrival, and a collecting device able to convert a photon into a readable electronic signal. TCSPC technology works electronically, therefore the trigger signal must be electronic too (might be provided internally by the laser or it needs to be converted externally by a photodiode). Furthermore, photoelectric conversion of single photons typically provides a signal that is hardly detectable. To circumvent this issue, electron multiplications methods were implemented. Firstly, TCSPC was demonstrated employing the PhotoMultiplier Tube (PMT) [71]. Most recently, TCSPC has been relying on Single Photon Avalanche Diodes (SPADs) or superconducting detectors [72].
A PMT is a vacuum phototube commonly adopted for high-sensitivity, low-noise, ultrafast pho-

ton detection. On one side, these devices typically possess a large collection area; on the other one, they need to be used in a dark environment and tend to be small-sized. Single-photon detectors based on superconducting nanowires are devices that have recently attracted the attention of many scientists. The working wavelength range is typically near-infrared. In this range, superconducting detectors have demonstrated highefficiency and excellent temporal resolution. Clearly, they require an appropriate cooling system.

I will now briefly focus on the SPAD detectors, since these are the ones employed for the original work reported in this chapter. SPADs are semiconductive photodiodes that exploit the short-duration avalanche current triggered by a single photoelectric event. This current is generated by the fact that the diode is set close to its breakdown voltage. To achieve this, the p-n junction is used in reverse biased mode. As a result, the current can be now detected, and the initial condition needs to be restored in order to repeat the process again. This happens during the so-called "dead time", i.e. when the system is unable to detect any other incoming photon. The restoring process happens thanks to a quenching circuit that stops the avalanche by changing the voltage applied to the diode itself; this process takes around hundreds of nanoseconds.
In Fig. 2.3(a), a schematic representation of a TCSPC measurement is reported. A pulsed laser is used to illuminate a scene and to trigger the TCSPC electronic board. This could be external to the detector, for example when using a single-pixel SPAD or PMT, or integrated on the device electronic board, as for the particular example shown of a SPAD camera. Typically, the temporal window for the observation of the event of interest is set, or limited, by the repetition rate of the

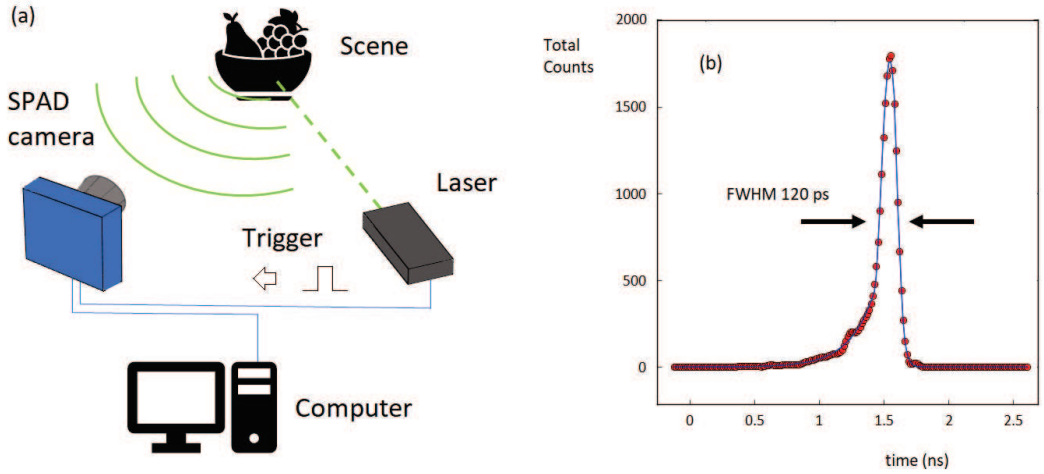

Figure 2.3: Scheme of a general TCSPC measurement. Experiments that involve measurements of the time of arrival of single photons require synchronization between the detector and the laser for the TCSPC electronics (internal to the SPAD camera for the example in (a)) to allocate correctly in time the counted photon. In (b), an example of IRF for a SPAD sensor is reported (obtained measuring a femtosecond pulsed laser beam). The red points represent the actual measure, whereas the blue is showing the fitting curve (function described in [73]).

laser. Once this window is defined, the TCSPC subdivides it into a fixed number of time bins. In order to populate these time-bins, the time of arrival of a single count is estimated comparing the trigger signal to the time of the recorded single photon event. After having repeated this operation over a number of iterations, an histogram like the one reported in Fig. 2.3(b) is obtained. The resulting Impulse Response Function (IRF) is affected by different factors: electronic jitter, the wavelength-dependent absorption coefficient and the absorption characteristics due to the design of the detector itself (size of the sensor with respect to the total pixel-size, position of the sensor etc...). The FWHM of the IRF can reach around 40 ps for the most recent SPADs and 20 ps for PMTs. Moreover, a discrete number or counts must be considered, typically around 10^2 to 10^4 photons, in order to build good temporal histograms. However, because of the presence of a dead time, the possibility to achieve a reliable distribution depends on the photon-counts-over-laser-pulses ratio. The goal is the so-called "starved regime", i.e. around 1% of laser pulses should contribute to a single count. Under these conditions, the probability distribution for a single photon count is uniform over the whole temporal window. This introduces the need for a large number of iterations. In practice, to accommodate this need one can resort to a high-repetition-rate, and sufficiently powerful, pulsed laser.

The TCSPC-related sensors analysed so far were single-pixel detectors. To achieve 2D acquisitions, one solution is to resort to a scanning system either for the illuminating or the reflected light. Another approach is to directly build a 2D detector as the previously mentioned SPAD camera. In this particular case, each pixel is a SPAD detector composed of both the semiconductive diode and the TCSPC electronics. This makes the camera, or SPAD array, easily portable. Currently, efficiency is strongly limited by the fact that the sensitive part of the pixel is relatively small. The IRFs of these detectors are analogous to the ones in the example reported in Fig. 2.3(a), with FWHM $\sim$ 120 ps. Typically, Commercially available SPAD arrays have a pixel-resolution of 32×32 or 64×64.

2.2 Light-in-flight and related applications

"Light-in-flight" is a term introduced by Abramson, when he attempted to record the motion of light with his (previously mentioned) holographic technique [65]. Generally speaking, it refers to a set of techniques that exploit the ability to observe the propagation of light at the picosecond scale. Historically, the first reason for developing ultrafast optical detection methods, as the ones presented in the previous section, was simply to display the video of light in motion itself. Light-in-flight photography was first shown by means of non-linear light-matter interaction: creating a standing wave inside a particular solution of organic molecules and observing the resulting fluorescent pattern [56] (1967), and using a shutter based on the Kerr effect in order to observe the light scattered by the propagation of a light beam through milky solution [74] (1969).

Later, holographic approaches were considered, starting from the one of Abramson during 1978 [65] [75] and followed by white-light interferometry [64]. Differently from all the other techniques, coherence gating methods do not measure the intensity of the scattered light, but its coherent amplitude field. Some of the light-in-flight photography results can now be appreciated on Youtube, because of the artistic and educational relevance of these videos [76][77]; others are displayed in Museums as well [78]. Probably the most impressive is the one obtained through the streak camera in [79][80]. Here, the video is recorded at one trillion frame per second having the light propagating through a solution that improves scattering. The first time the motion of light was recorded whilst propagating through air, therefore without resorting to another scattering medium, was thanks to a SPAD array [81]. In this case, pulses of 500 ps were observed on a simple optical path in which the light was reflected by two mirrors.

Light-in-flight photography can be used to observe different interesting phenomena. Firstly, to visualise the propagation of light in different media such as fibres and slow-light media. For example, a SPAD camera has been used to study the supercontinuum generation derived from the propagation of light pulses in a photonic crystal fibre [82]. In order to record this broad-spectrum event, a set of different filters were put in front of the camera, each one for a different wavelength. In this way, it was also possible to investigate the chromatic dispersion of this fibre [83].
Other dispersive media considered for light-in-flight photography were the slow-light ones, i.e. media that can induce very low group velocities to the light propagating in it [84]. In particular, SPAD detectors were applied to situations where sensitivity of single photons was required [85].

The fact of dealing with the light on a scale in which it cannot be considered to be propagating at infinite speed can raise interesting distortion imaging effects. In fact, whenever attempting to image an object at relativistic speed one must take into account the actual time in which the light reaches the camera, since this will be now comparable to the temporal scale of the motion of the object itself. Two main distortions have been studied: “shearing” and “temporal inversion”. The shearing effect was firstly introduced and studied by Duguay and Mattick in 1971 [57]. As an example, it is possible to give the illusion that two pulses that are reflected at two different depths, from an observed scene, will arrive in the same exact time just by accurately adjusting the position of the sources of the two pulses in such a way that these pulses will share the same optical path length. In the same way, objects propagating transversely with respect to the camera, but at different depths, will not appear overlapped because of the additional time that the light of the more distant object takes to reach the camera.
Time inversion, at these speeds, can be generated using appropriate geometries [86]. In this work, the authors were able to demonstrate also an apparent superluminal propagation and to show transition from superluminal to subluminal propagation. This transition led to the obser-

vation of image pair creation and annihilation.

The principle of light-in-flight photography can be used to image repetitive ultrafast events distinct from the propagation of light itself. Thanks to the high-speed cameras, it has been possible to image the effect of focusing a femtosecond laser pulse, thus generating an ionised plasma at the focus [81] using a SPAD camera or observing light filamentation effects [87] by means of a streak camera.

2.2.1 Looking around the corner

I will now focus on one of the main applications of light-in-flight: looking around the corner. As already mentioned in the introduction to this chapter, looking around the corner is part of a larger context, NLOS, that aims at retrieving information about objects that are not in the direct line of sight, such as targets hidden behind a corner or a wall. Nonetheless, here I will discuss those specific scenarios in which the presence of an obstacle is overcome by resorting to multiple-scattered light. Generally speaking, this might apply also to seeing-behind-walls scenarios. However, seeing-behind-walls typically refer to those techniques that rely on radiation that can directly bypass the obstacle (wall). Furthermore, since the wavefront of a laser pulse bouncing off a wall can be considered approximately spherical, the more scattering events the less detectable and more complex the resulting signal will be (because the intensity decreases with distance, and due to possible scattered-light superpositions [88]).

In Fig. 2.4, a typical looking-around-the-corner layout is illustrated. Generally, a pulsed laser, such as the one reported in the scheme, is employed for this kind of experiments. In order to overcome the obstacle, the pulsed laser beam is pointed towards a wall; in this way, the light is now scattered in all directions and thus can illuminate also the hidden object. Another scattering event occurs when the light hits the object. Part of the resulting backscattered light will be directed towards the camera field of view and can therefore be collected.
How to exploit the information encoded on the back-scattered light depends on the experimental conditions and the final goal. In this sense, research has evolved in two main directions: 3D reconstruction and object-tracking. Full-3D reconstruction typically relies on bulky cameras and time-consuming algorithms. However, there are many practical situations, such as locating the position of a hidden target, for which a 3D reconstruction is not required. In these cases, the information carried by the back-scattered light can be aptly reduced. Thus, the way in which data are collected and analysed may be very different for the two methods.

Looking-around-corners 3D reconstruction is based on finding the inversion of the operator that maps the 3D geometry of the hidden scene into the data measured by the camera. Clearly, knowing the scene and the condition under which it is illuminated, it is relatively easy to simulate the

temporal distribution of the light propagated and then collected by the camera. On the contrary, solving the inverse problem is more demanding. The first to attempt to estimate this inverse operator were Velten et al. in [89]. Their algorithm consisted of a filtered backprojection. In particular, the projection method exploited the assumption of three scattering events in order to simplify the problem, whereas the filtering operation was focused on extracting surface features. Using a streak camera, they were able to provide an effective IRF of 15 ps and, in this way, demonstrate 3D reconstruction, with a 1 cm lateral spatial resolution, of objects with size of dozens of centimetres. Further feature extractions were performed in [90] by the same group. Later, Heide et al. proposed a method that could provide 3D reconstruction relying on a less bulky detection system, the PMD, and an algorithm based on convex optimisation [91]. SPAD detectors were introduced into looking-around-corners by Buttafava et al. [92]. In the work of O'Toole et al. [93], a single-pixel SPAD detector, exploited in a confocal scanning configuration, was employed to retrieve the 3D geometry of hidden objects at relatively high speed. This convenient solution could provide fast high-resolution reconstructions thanks to a lighter computational approach. On the other hand, the object to be imaged required a sufficient reflectivity with respect to the background.

Object tracking was first performed with a streak camera in [94]. In this case, it was possible to

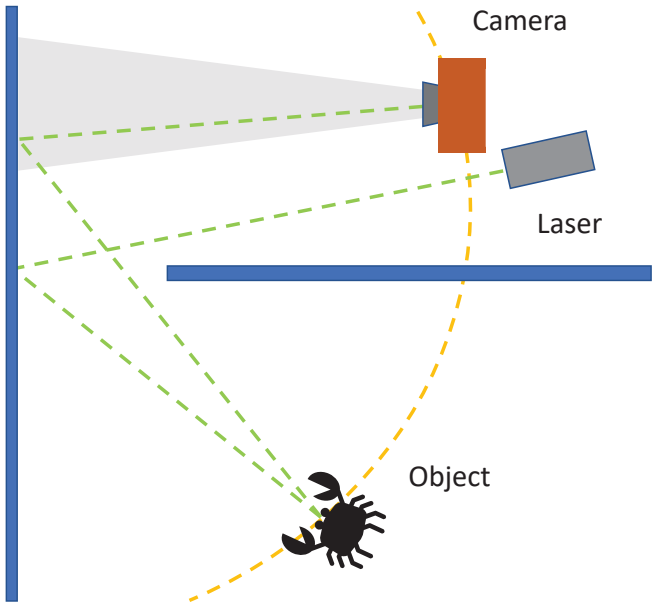

Figure 2.4: A typical looking-around-the-corner setup. A pulsed laser illuminates the hidden target exploiting the scattering that occurs on the wall. The light that bounces back undergoes a second scattering on the wall and, in this way, can be collected by the camera. Therefore, a minimum of three scattering events must occur for recording information about hidden targets using this technique. A top-view section of the ellipsoid (orange line), representing all the points that share the same time-of-flight of the light that propagates from the laser to the camera (green line) illuminating a specific point on the hidden object, is shown.

retrieve both size and direction of hidden objects by subtracting consecutive frames in order to isolate the targets from the background. Gariepy et al. introduced two new elements to locate an object in real time [88]: SPAD detectors and a triangulation algorithm. In fact, focusing on the position only, a simplified and quick algorithm can be considered. Referring to Fig. 2.4, it is evident that the symmetry of the problem implies that the signal reaching the detector can lie on an ellipsoid defined, in particular, by the position of "A" and "B" on the wall (in other words, the points belonging to this ellipsoid will share the same time-of-flight). To isolate the correct position, the authors employed more single-pixel detectors by resorting to a 32×32 SPAD array. In fact, since each pixel looks at a slightly different zone on the wall, triangulation can be achieved by intersecting the different ellipsoids obtained from the different pixel-detectors. Allowing enough distance between the single-pixel detectors can also help the triangulation calculation. Furthermore, it was shown that two detectors are enough to identify the position of certain targets when some prior knowledge about the shape of these targets is provided. SPAD detectors turned out to be particularly useful because of the poor intensity of the backscattered signal typical of this experiment. In [95], the authors achieved long range location and tracking of human targets, for distances of ~ 50 m. Moreover, these measurements were performed with an infrared laser in an environment with daylight (using an appropriate filter), thus demonstrating the applicability of such techniques in a more realistic context.
Interestingly, looking-around-the-corner was also demonstrated without the need of a pulsed laser [96]. In this case, the illumination of the hidden scene was provided by a continuous wave laser and the backscattered light recorded with a common CCD camera. The authors could achieve object tracking by developing an algorithm able to compare the simulation of the multiple scattering event to the measured field by exploiting prior knowledge about the shape of the hidden object.

The original work presented within this chapter proposes a different approach to the looking around the corner problem. Considering a situation with experimental conditions analogous to the one reported in Fig. 2.4 and employing just a single-pixel SPAD detector to collect the light, one could ask: what is the amount of information contained in a temporal histogram measured with such a single-pixel detector? What kind of information can be extracted or exploited?
Clearly, under these conditions it is extremely difficult to provide a forward model that could accurately describe this multiple-scattering problem, having 3D spatial information encoded on a 1D temporal histogram. In this work, a neural network algorithm is proposed in order to extract specific information from the single-pixel histogram without the need for solving explicitly the problem. In particular, is it possible to both locate and identify a human hidden target? Generally, looking around the corner techniques are demonstrated on targets of a relatively small size. The idea of using human individuals poses the challenge within a more realistic and unexplored scenario.

2.3 Experimental layout & procedure

The experimental layout is represented in Fig. 2.5(a). The scheme is designed as a typical looking around the corner setup. A pulsed laser source (with a repetition rate = 80 MHz and $\lambda = 808$ nm) is pointed toward a wall where a first scattering occurs. Part of the light can reach the hidden target. Thus, the light backscattered from the hidden target can be captured by the detection system.

As was mentioned earlier, the hidden targets are human individuals shown in Fig. 2.5(b)(c) (d), with relative heights 1.68 m, 1.57 m and 1.87 m. I will refer to them, respectively, as individuals "n.1", "n.2" and "n.3". The measurements are performed both with different clothing, shown in Fig. 2.5(b)(c)(d), and with the same clothing, reported in Fig. 2.5(e) (for simplicity, in this case it is only reported for the individual "n.3").

Furthermore, the subjects were placed in 7 different positions ("A", "B", "C", "D","Db", "Df", "E" and "F"), as displayed on the floor in Fig. 2.5(a). The 5 positions "A", "B", "C", "D" and "E" are located in such a way that the recorded photon-time-of-flights are similar. In practice, once the position of the target, laser and detection system are fixed, the time of arrival of the photons depends only on the location of the two points where the light is scattered from the wall. As for the triangulation example, the locations that share the same time-of-flight lie on an ellipsoid. The closer the two points on the wall, the more the ellipsoid collapses into a sphere. Therefore, these two points are kept reasonably close to each other, such that the target positions are easy to locate on the right ellipsoid. In contrast, "Df" and "Db" are put in front and behind "D" along the ideal "radial" direction.

As previously explained, this experiment aims at extracting information about position and identity of different hidden targets using just a single-pixel detector. However, the idea of relying on a neural network algorithm to achieve this goal implies the need for a sufficiently large amount of (single-pixel) measurements, i.e. temporal histograms, in order to train the network itself. To accommodate this need, a SPAD array is employed. As a matter of fact, the SPAD array considered in the experiment consists of 32×32 independent single-pixel detectors, each one with an IRF of ~ 120 ps. An objective is used to set the field of view of this array on the wall. The choice of the actual size of the field of view is driven by the need for a homogeneous training dataset. Since the IRF corresponds to a depth resolution of 3.6 cm, in order to avoid an excessive variability among the detected histograms the size of the field of view must be comparable to this value. Indeed, it has been experimentally observed that an optimal solution for the size of field of view is 3×3 cm^2. A smaller size would have led to a poor variability among the dataset. To synchronize the SPAD and the laser pulse, I made use of an external trigger (Becker & Hickl OCF-401), omitted for simplicity in Fig. 2.5, which is illuminated by a small fraction of the light at the laser output. The acquisition of a single measurement takes 2 s, thus relying on 160×10^6 laser pulses.

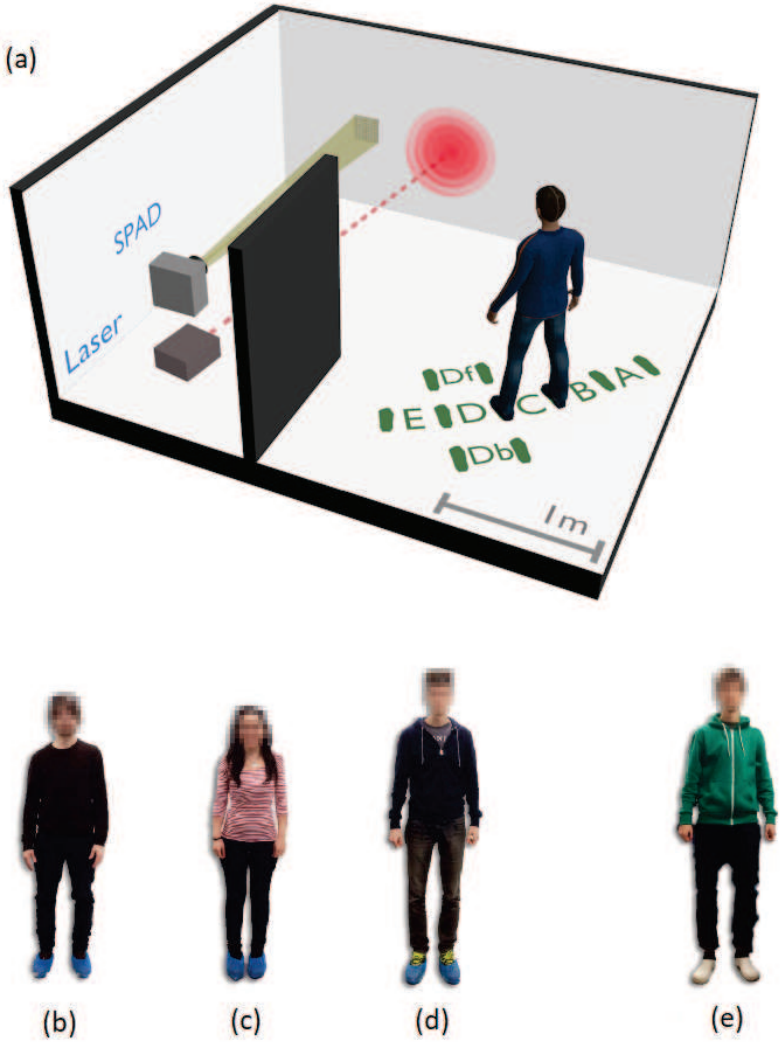

Figure 2.5: Representation of the experimental layout. Pulsed laser light, $\lambda = 808$ nm and rep. rate $= 80$ MHz, is scattered in all directions after hitting the wall. The hidden human target is thus illuminated, and the light backscattered is collected by means of a SPAD camera. Simultaneously, 800 temporal histograms are obtained on an area with size $\sim 3 \times 3$ cm^2. 5 measurements have been taken for each of the three individuals reported in (b) "n.1", (c) "n.2." and (d) "n.3", in each of the seven positions drawn on the floor. The same procedure is repeated also with the individuals wearing the clothing shown in (e). Figure adapted from [1].

Once the "hot-pixels" (pixel reporting incorrect photon-counts) are discarded, 800 histograms are obtained from each SPAD array shot. In Fig. 2.6, examples of these temporal histograms, collected in a single take, are reported. In particular, Fig. 2.6(a) shows 6×6 samples from a single take (or measurement) with the target "n.1." in position "C". Moreover, in Fig. 2.6(b) single histograms from the same pixel of the three individuals, wearing the same cloths, in position "C" are presented. The time binning for each histogram is 55 ps. As can be observed, on one side a consistent variability among the histograms of a single take can be appreciated; on the other one, the difference between the histograms of the three individuals is made evident. Nonetheless, it can be observed that there is no prior information about the individuals that can

easily address and determine these differences. As a matter of fact, the targets share similar heights, widths and photon counts. Moreover, it is not evident how these and other more subtle features will contribute to the final measured histogram. In this sense, the idea of having most of the positions sharing a similar time-of-flight - by lying on the ellipsoid - is to allow the network to focus on those features and determine an internal representation for this classification problem based more on the shape of the peak with respect to its temporal position.

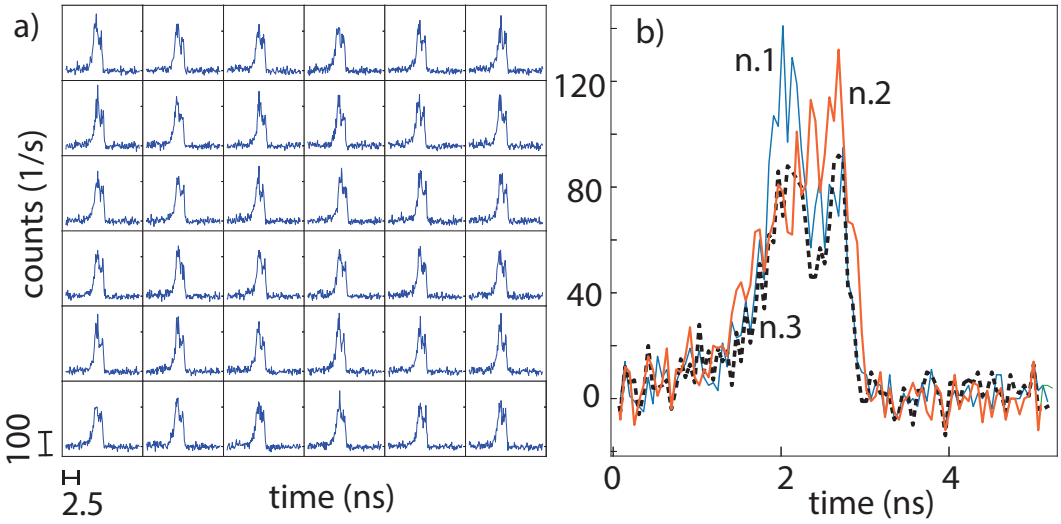

Figure 2.6: Examples of 6×6 histograms from a single measurement for individual "n.1" in position "C", reported in (a). In (b), comparison between histograms of the three individuals, same position "C", taken from the same pixel. Time binning is 55 ps for all the histograms. Figure reprinted from [1].

Five measurements (800 pixels each) are acquired for each individual in each position. The procedure is as follows: fixing a position, three measurements are taken by alternating the three individuals (facing the wall along the ideal radial direction); then, the same operation is repeated for the remaining 6 positions; finally, these steps are repeated four other times, thus obtaining the five different measurements. Furthermore, the same procedure is repeated for two different cases: same clothing and different clothing. An initial background measurement is obtained and subtracted from each other measurement. This operation is undertaken to help identify the peak. However, on one hand, it has been demonstrated in other papers [88] that the background can be easily calculated by averaging over many frames when capturing the motion of a hidden object; on the other hand, the network can, in principle, learn to isolate the actual signal from the background on its own. Nonetheless, it was judged to be beyond the scope of this project to provide an in-depth verification and analysis of these aspects.

2.4 Method

A non-linear neural network classifier is employed in order to distinguish the histograms with respect to location and identity of the corresponding hidden target. In this sense, the features that indicate the presence of a certain individual in a certain position are not found thanks to an explicit mathematical model. Instead, these are identified implicitly by the ANN algorithm as a result of a learning process based on observing a large number of histograms. In particular, in this experiment the algorithm used was characterised as supervised training. Therefore, individuals and positions are treated as classes and the network is built in such a way that the predictions about identity and location of the input histograms can be compared to the ground truth class-labels during the training process. More precisely, the classes have been encoded thanks to an "one-hot" representation [17] with $N_i = 3$ binary outputs for the individuals and $N_p = 7$ binary outputs for the positions.

The aim of the network is to provide a simultaneous classification of both location and identity. Therefore, the cost function to be minimized is defined as the junction of two distinct cost functions, one for the positions and one for the people. In particular, the single cost functions are both categorical cross-entropy losses. As described in Chapter 1, where this type of cost function has been introduced, this is a typical choice when dealing with a multi-class classification task. For the α-th input sample, being $\bar{y}^i_{\alpha,j}$ the predicted output for the j-th individual and $y^i_{\alpha,j}$ the respective ground truth class label, the loss function for the classes of individuals is expressed as:

$$\mathscr{L}^i_\alpha = \sum_{j=1}^{N_i} \left(y^i_{\alpha,j} \ln \bar{y}^i_{\alpha,j} + (1 - y^i_{\alpha,j}) \ln(1 - \bar{y}^i_{\alpha,j}) \right).$$

In the same way, for location, indicating with $\bar{y}^p_{\alpha,k}$ and $y^p_{\alpha,k}$ the predicted position and the ground truth for the k-th position, the cross-entropy loss is

$$\mathscr{L}^p_\alpha = \sum_{k=1}^{N_p} \left(y^p_{\alpha,k} \ln \bar{y}^p_{\alpha,k} + (1 - y^p_{\alpha,k}) \ln(1 - \bar{y}^p_{\alpha,k}) \right).$$

Therefore, the complete cost function $\mathscr{L}$ to be minimised over the M samples of the training dataset is:

$$\mathscr{L} = -\frac{1}{M} \sum_{\alpha=1}^{M} \left(\mathscr{L}^i_\alpha + \mathscr{L}^p_\alpha \right).$$

The architecture of the network is built allowing the input histograms to be processed in parallel by two different type of layers: fully-connected (or dense) and convolutional. The basic idea is to exploit the translation invariant and scale-free nature of the convolutional layers to focus on those features that can be extracted from the shape of the histogram peak despite its temporal location. On the contrary, the dense layer can alternatively provide more information related to

the distance as well.

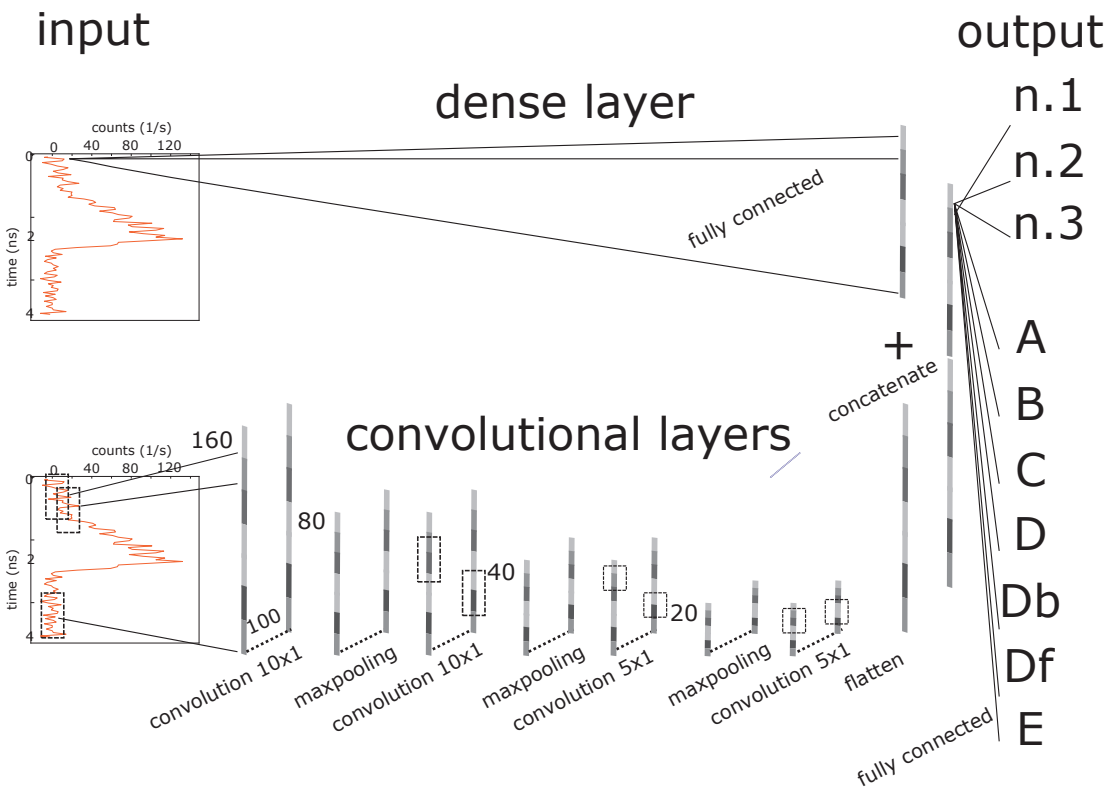

Figure 2.7: Schematical representation of the ANN architecture. Input histograms are processed in parallel using a fully-connected layer and convolutional layers (100 filters). The resulting outputs are concatenated and another pair of fully-connected layer produces the final outputs, thus obtaining, simultaneously, both position and identity. Figure reprinted from [1].

A schematic overview of the network is available in Fig. 2.7. As it can be observed the input histogram is fed, in parallel, to both groups of layers. Here, the convolutional side consists of a series of convolutional layers, normalised by batch normalisation layers [97] and followed by max-pooling layers in order to down-sample the output of each convolutional layer, apart from the last one. The convolution is mono-dimensional, having the size of filters for the first two layers 10×1 and 5×1 for the other ones. The number of filters is 100. The output of this group of layers is concatenated to the fully-connected layer. This fully-connected layer has also been batch normalised. Furthermore, dropout layers are used after each fully-connected layer in the network in order to help the generalisation of the result. All the activation functions used up to this point of the network are the non-linear "ReLu" [98]. Finally, two fully-connected layers are applied separately to the result of the concatenation. These layers produce the output predictions on location (N_p outputs) and identity (N_i outputs). In this case, the softmax activation function is applied to these layers, as for usual classification tasks. Moreover, the cost function, that compares the predicted class labels to the ground truth ones, is minimized using the stochastic gradient descent, with learning rate = 0.001 and Nesterov momentum [99].

The neural network has been implemented in Python resorting to the Keras library. In Appendix A, an example of the code is reported.

2.5 Results

After having acquired the data and defined the network model, the algorithm is ready to be tested. A "leave-one-out" cross-validation process is employed: having at its disposal 5 measurements (800 histograms each) for each individual in each position, the network is trained with only 4 out of these 5 measurements and the remaining one is left for the testing process. Then, the same procedure is repeated for all 5 of the permutations and the results are averaged.
First, the case of the three individuals wearing different clothing is considered. The results, reported in Fig. 2.8, are visualised by means of the so-called confusion matrices, one for the location and one for the identity, that compare the correct class-labels ("truth" on the vertical axes) versus the ones predicted by the neural network ("prediction" on the horizontal axes). Thus, a good classification implies a strong diagonal (with a maximum of 1). For each of the 5 permutations a confusion matrix is obtained. Therefore, the matrices shown in Fig. 2.8 are the average of the 5 matrices with relative standard deviation.

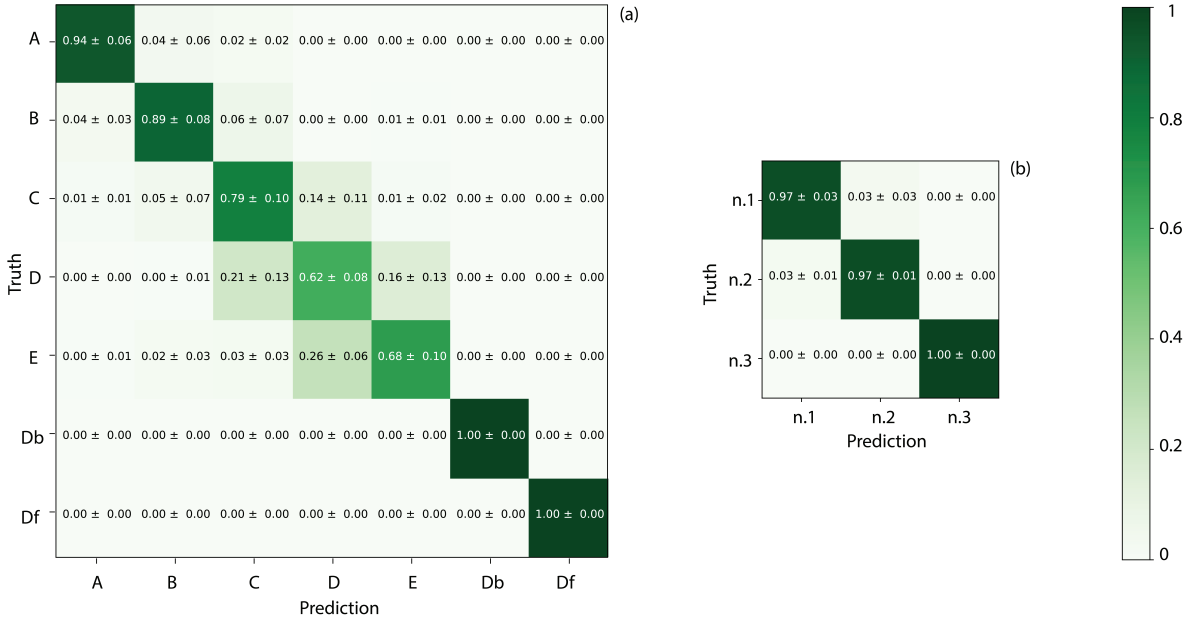

Figure 2.8: Results for the "different-clothing" case. Confusion matrices are used to help visualise the data. On the vertical axes. the ground truth classes are reported, and, on the horizontal axes, the network predictions. The results are averaged over the 5 measurement-permutation predictions. Two matrices are shown: (a) location and (b) identity. Figure reprinted from [1].

The location matrix in Fig. 2.8 (a) confirms that the two positions with different time-of-flight with respect to the other positions (i.e. "Db" and "Df") are the easiest to be classified correctly.

Clearly, this happens because of an evident temporal shift for the peak. The network performance decreases slightly for all other positions. However, it can be observed that, the highest percentages still lie on the diagonal and the largest misclassifications (a maximum of 26%), occur only for contiguous positions. Among the positions that share similar time-of-flight, the easiest to be classified seems to be position "A". A possible explanation for this might be related to the fact that the real wavefront, generated by the scattering on the wall, could be not entirely spherical and present some preferred direction. Thus, since "A" is the closest position to the wall, this might introduce a further element of distinction for the relative histograms by influencing the way in which this position is illuminated.

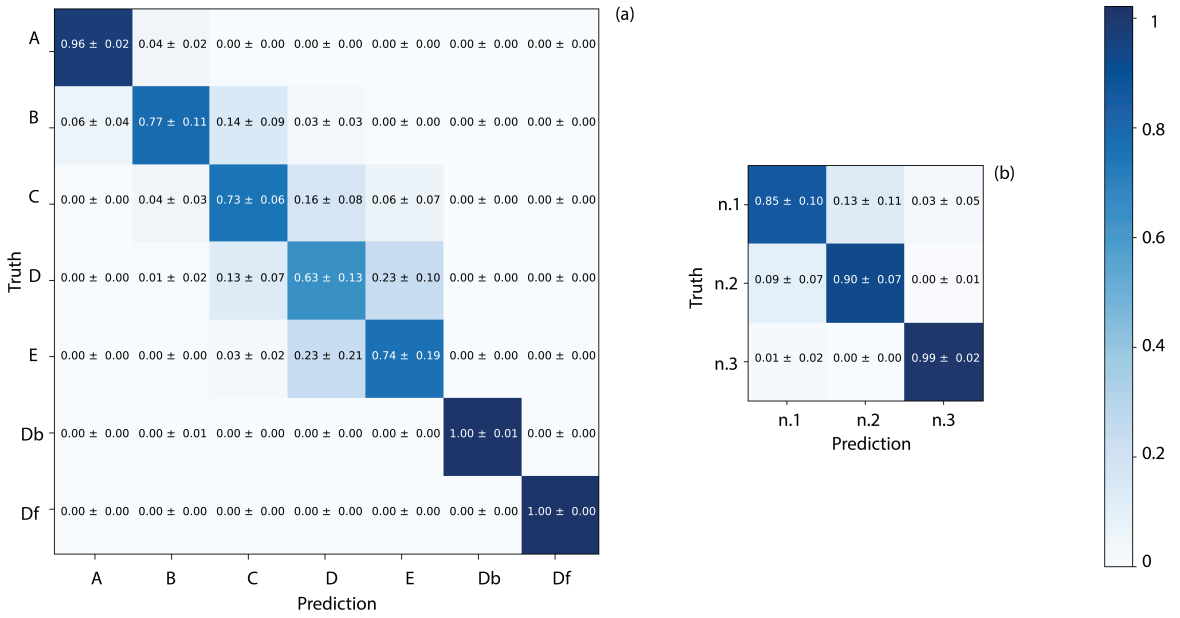

Figure 2.9: Results for the "same-clothing" case. Data are reported with the same approach employed for Fig. 2.8. Figure reprinted from [1].

Considering Fig. 2.8 (b), the classification for the individuals is almost perfect. Indeed, the different reflectivity properties of the clothing that the individuals are wearing might play a key role. To verify whether the network is still able to correctly distinguish among the individuals when these are sharing a similar reflectivity, the same experiment is repeated with the targets wearing the same clothing (shown in Fig. 2.5(e)). The results for this second case are reported in Fig. 2.9. It can be observed that the predictions for the positions are basically unaltered and consistent with the ones from the different-clothing case. On the other hand, the results for the identification of the targets present now a small misclassification between individual "n.1" and individual "n.2", reaching about 10%. Because of the similar reflectivity, the neural network should be now focusing more on aspects related to the overall shape of the hidden targets. Seemingly, the network is spotting the difference in height between the first two individuals and individual "n.3". The classification performance for "n.1" and "n.2" is nonetheless surprising,

especially taking into account that the 120 ps temporal resolution of the SPAD detector is insufficient to provide 3D facial details in these conditions.

A different approach for visualising these results is proposed and applied to the same-clothing case. The idea is to report the classification percentages for each of the 5 measurements, and thus not averaging as in the previous case, in order to gain more insights about possible sources of misclassification. As shown for the example in Fig. 2.10, the results for the individuals are reported within a 7×3 grid. The columns represent the ground truth positions, whereas the rows represent the ground truth individuals. The percentages are obtained over the 800 pixels of a single measurement.
As expected from Fig. 2.9, most of the misclassifications happens between the individuals "n.1" and "n.2". However, it is noteworthy that the majority of the measurements have a classification percentage close to 100%. Instead, typically just one measurement out of the five has a largely incorrect prediction. This non-uniform distribution of results suggests that misclassification should be related to a limited capacity, of the ANN, to interpret the differences between the different measurements. In this sense, errors related to the repeatability of the camera (possible differences in target-poses or illumination, etc.) might introduce a variability that has not been fully captured in this dataset.

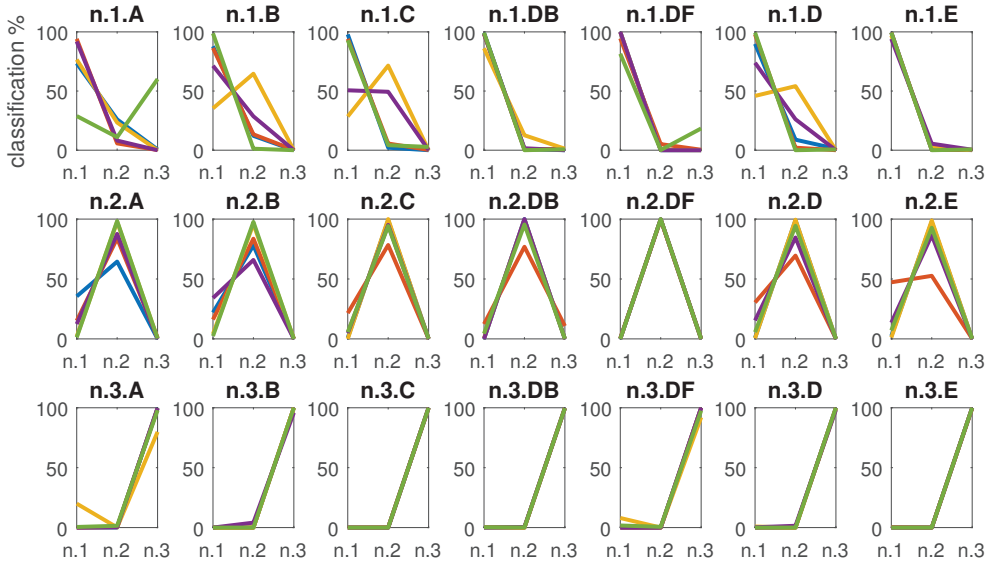

Figure 2.10: Classification percentage results for individuals visualised for each measurement-permutation. Columns and rows represent the different ground truth labels for individuals and positions, respectively. Each element shows the individual prediction for the 5 measurements. Figure reprinted from [1].

2.5.1 Alternative architectures

A number of different neural network architectures have been proposed and tested on the same-clothing data. In particular:

- a simplified version of the algorithm reported in Fig. 2.7 (with less convolutional filters);
- a network with only fully-connected layers (i.e. excluding the convolutional layers group);
- two separated classifiers (fully-connected networks) for position and identity.

Furthermore, these different architectures have been considered with either a Gaussian noise layer included or not. The idea of applying a Gaussian noise layer - directly to the input histograms - is to help the algorithm to overcome a possible broad variability among the pixels due to the different noise realisations among the detectors.
With respect to the model reported in Fig. 2.7, the convolutional simplified version presents a reduced number of filters (from 100 to 32) with size 5×1 for each layer. Moreover, the max-pooling is now replaced by average-pooling and batch-normalisation layers are removed.

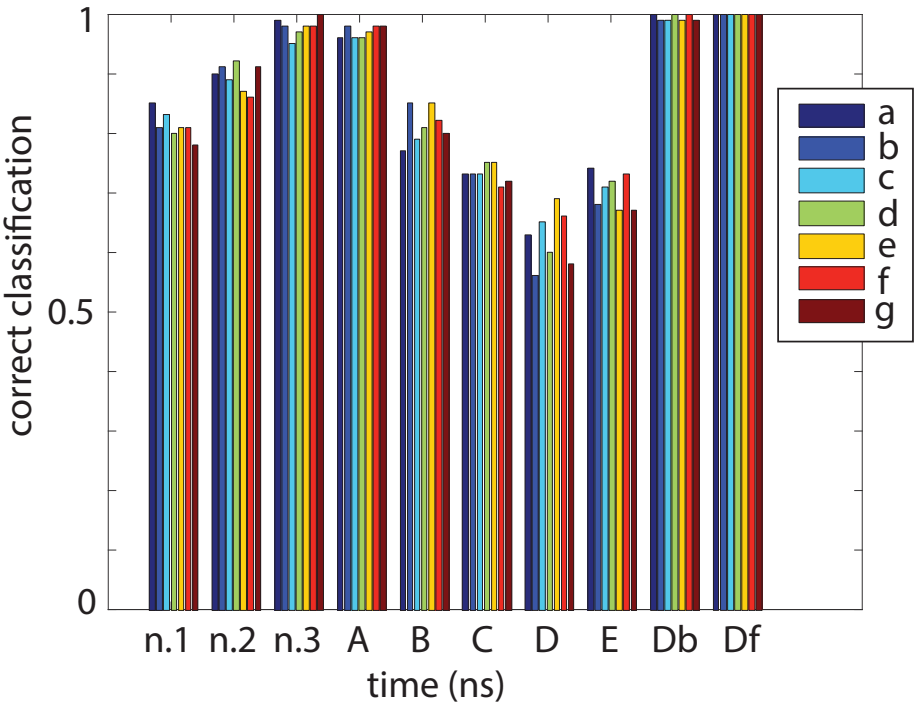

Figure 2.11: Comparison of the results for the different architectures. (a) the model presented in Fig. 2.7; (b) model (a) with added Gaussian noise layer; (c) simplified version of model (a) with less filters (32) of size 5×1 per layer; (d) model (c) with added Gaussian noise; (e) only fully-connected layer; (f) model (e) with Gaussian noise layer; (g) fully-connected models applied separately for identity and position. Figure reprinted from [1].

In Fig. 2.11, the results for these different approaches are shown. The models have been tested through the same leave-one-out cross-validation approach (i.e. averaging over the results of the 5 measurement-permutations) although only the correct classifications are now being reported. It is noteworthy that the performances of the different approaches are similar. Indeed, the fact that no main improvements are observed by changing the network model might be related once again to an insufficient representation of the variability of the data, thus indicating the need for a more controllable or larger dataset. On the other hand, Fig. 2.11(g) seems to indicate that learning simultaneously both location and identity might improve the overall classification. Possibly, this is related to a deeper capacity of the network to build a more realistic representation of the observed scene when dealing with both the location and identity classes.

In conclusion, the quest for an ultrafast optical detection system has paved the way for exciting and novel techniques related to the capacity of freezing light at the centimetre scale. In particular, SPAD detectors have demonstrated a single-photon sensibility in conjunction with short response functions. In the experiment proposed, a single-pixel SPAD detector has been employed for retrieving information about human targets hidden behind a corner by exploiting multiple light scattering. Since $3D$ spatial information is encoded in a single temporal histogram, the complexity of the problem does not allow to resort to a simple forward model. Nonetheless, a neural network algorithm has been used to provide correct classification of these histograms with respect to 7 designed positions and 3 individuals. Results have shown that the network predictions are in good agreement with the ground truth for both the cases of individuals sharing a similar reflectivity or not. Unfortunately, for this supervised network it is not possible to provide results for individuals or positions that do not belong to the training dataset. Furthermore, the idea of dealing with a more realistic scenario, thus having real human targets as opposed to small size objects commonly adopted for NLOS techniques, might introduce further elements of variability that have been not entirely captured by the proposed training dataset. Other sources of misclassification might be related to the sensor itself, which exploits multiple single-pixel detectors in parallel. In this sense, future work will be focused on the reproducibility of these results in a more controlled environment and the possibility to test this method on more challenging situations (for example with more individuals, different poses). Nonetheless, this work has served as a proof of concept with the possibility of extracting information about hidden targets in an ill-posed situation such as a 3D information collapsed into a 1D histogram. This might open up the possibility of further unsupervised methods with the ability to determine the 2D location of a hidden target on a not-prearranged set of position, for example exploiting 2D polar coordinates.

Chapter 3

Complex-valued network for transmission of natural scenes through multimode fibre

3.1 The need for optical fibres

The ability to transport light has always been of paramount interest in order to exchange information and interact with hardly accessible or distant places. Medicine, Neuroscience, Communication: the demand for reliable optical means to guide light is essential within this modern era, with applications ranging from channelling huge amounts of data to providing good quality in-vivo images.

Most commonly used within this context, the optical fibre is a type of waveguide which typically presents a cylindrical symmetry. The fibre internal core is encircled by the cladding that consists of a material with a different refractive index with respect to the core (the cladding refractive index is lower). It is exactly this gap in the refractive index that forces the light to propagate within the core, thanks to total internal reflection. The reason why these waveguides are so widely used is because of their flexibility, which makes them particularly easy to handle. Nonetheless, whenever a wave is confined, its propagation ceases to be free and, on the other hand, has to satisfy precise mathematical relationships, defined by the geometry of the confining medium and by the characteristics of the wave propagating in it, named "modes".

If the diameter of the core of the fibre is particularly small (typically around 8μm), there is the possibility that the fibre will allow just a single mode. Indeed, the term generally adopted in this case is "single-mode" fibre. This fibre is preferred for communication over very long distances because of the reduced power losses relative to an electric cable or a fibre with more modes. For example, considering the fibres that transmit high-speed internet inside cities, the information is encoded in the shape of a temporal signal, as a matter of fact, a single-mode fibre can only allow the transmission of a single mode (two if the polarization is considered as well) with an approximately Gaussian spatial profile [100]. As a consequence, it is not possible to transmit a more complex spatially-shaped signal, such as an image. The only alternative to temporally

encoding the information is to exploit a large number of single-mode fibres in parallel, each one carrying a different part of the original wavefront. This takes the name "fibre bundle" and it is the technology commonly adopted for endoscopes in Medicine or other fields. Nonetheless, this bundle presents strong limitations, in fact, whenever coherent light is used a strong coupling effect might arise if the single-mode fibres are too close, thus corrupting the image or signal. On the other hand, if the fibres are kept too far from each other, a pixelation effect will take place [101]. This poses a severe constraint on the use of fibre bundles for high-resolution imaging, where a large number of pixels is involved.

An alternative approach is to resort to "multimode" fibres. In this case, an image could be simultaneously coupled onto the multiplicity of modes allowed by the fibre. Furthermore, it can be demonstrated, through simple considerations, that the number of modes per unit of area could be up to a hundred times greater relative to fibre bundles [102]. Despite this evident advantage, the multimode fibre is characterised by a strong distortion that affects the signal at the output. This is the result of the interference between the different modes, excited inside the fibre, each one traveling with a different speed. The scrambled pattern takes the name of "speckle pattern", or simply speckle. In Fig. 3.1, an example of speckle patterns measured at the output of a fibre, with the respective coupled inputs (amplitude encoded patterns), is shown. This happens when the light is coherent, otherwise the modes are incoherently mixed (resulting in a uniform blob).

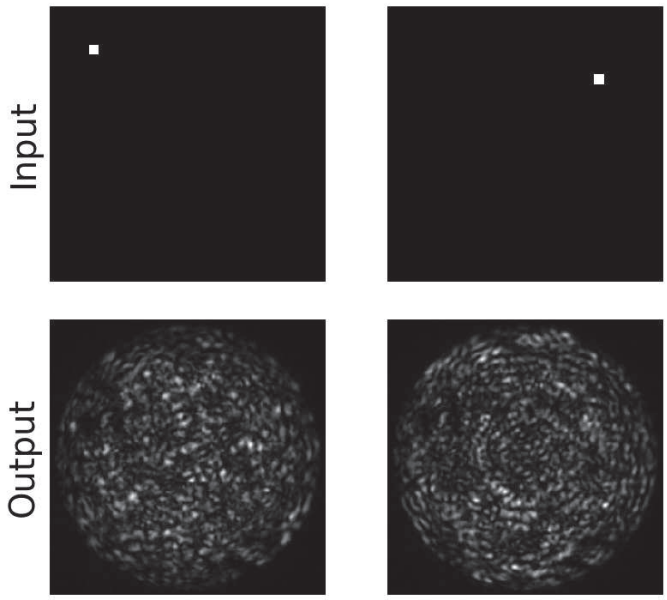

Figure 3.1: Example of distortion of an input pattern after being propagated through a multimode fibre. In this case a 28×28 amplitude image is coupled into the fibre. By changing the illuminated input pixel the output presents a completely different speckle pattern. Figure adapted from [3].

From a theoretical point of view, once the fibre is perfectly characterised, i.e. knowing length,

core diameter and refractive index profile, along with the radiation properties, the speckle field could be unscrambled in order to recover the original image or, similarly, the propagation of the light inside the fibre could be simulated. As a matter of fact, by assuming no losses, the system is linear and the propagation of the modes can be easily predicted. However, this is no longer true in practice. In a real fibre, imperfections, bending and temperature gradients can vary the refractive index locally and thus introduce coupling between modes. This represents a serious issue because, for example, internal imperfections cannot be characterised a-priori nor measured. As a consequence, for many years multimodes fibres were largely treated as random objects, i.e. as completely disordered media.

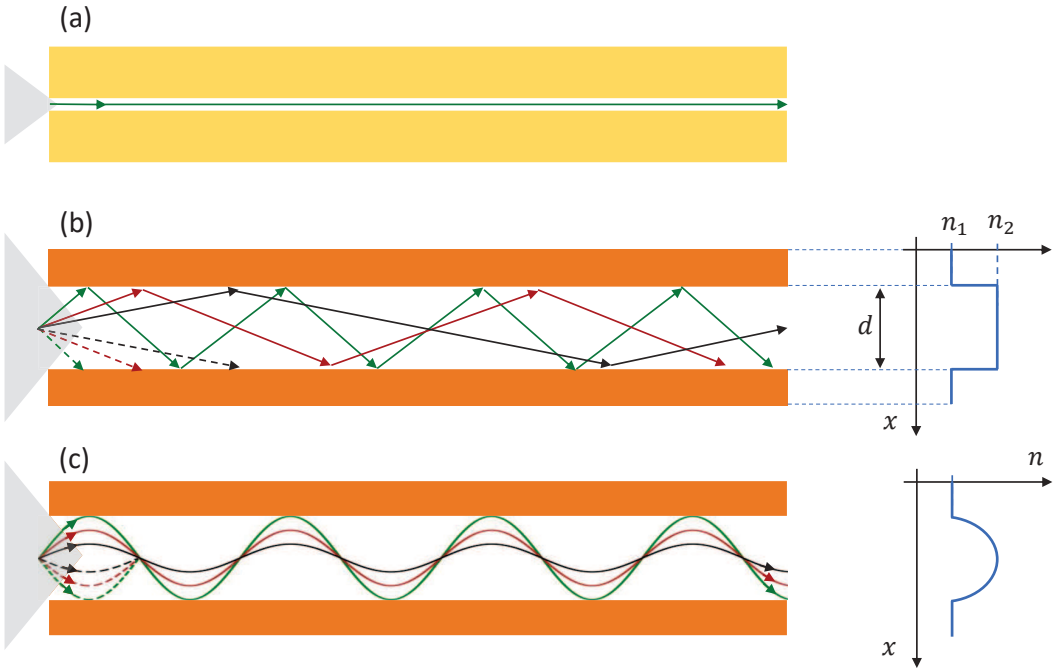

Figure 3.2: Schematic representation of three different optical fibres: (a) single-mode fibre, (b) step-index multimode fibre and (c) graded-index multimode fibre. The propagation is visualised in the form of light rays. In (a) just a single mode is allowed. In (b) different modes follow different paths thus leading to dispersion. The step-index profile is shown too. In (c) the parabolic-like refractive index profile minimises the dispersion with respect to (b), as the modes are periodically focused. The acceptance cone, indicating the possible angles of incidence allowed for the input light rays, is shown in gray. d indicates the diameter of the core. Typically, single mode fibres have a step-index profile, but they can be found also with graded one.

Referring to Fig. 3.2, commercially available multimode fibre can be distinguished into two main categories based on their refractive index profile: “step-index” or “graded-index” (GRIN). Considering a geometric-Optics approximation, modes can be associated to light rays. In the step-index case, the difference between the two refractive indices, core and cladding, forces total internal reflection of some light rays because of the Snell law, thus trapping the light. This defines a maximum angle of incidence above which the light is no longer reflected but

refracted. The numerical aperture (NA) is commonly used to indicate the sine of this maximum angle. Since different modes follow different paths they will be spread in time (apart from the symmetrical ones relative to the cylindrical geometry), as said before. This phenomenon takes the name of "modal dispersion". In the case of a graded-index profile, the light beams follow sinusoidal paths because of the particular continuous change of index of refraction along the radius. Ideally its profile, as represented in Fig. 3.2(c), is parabolic. A precise definition of the NA is debatable for a GRIN fibre; sometimes it is estimated in a similar way to the step-index case. Finally, the number of modes can be roughly calculated for both the types of multimode fibres.

3.2 Imaging through a multimode fibre

Early attempts to employ multimode fibres as imaging means include image encoding/decoding and optical phase conjugation. In the first case, transmission of images was demonstrated by means of spectral and spatial encoding in [103], a technique limited to a few meters range, and wavelength-time encoding [104]. Phase conjugation has been a long-standing research technique that aims to heal possible distortions that might affect a propagating wave by considering its phase-conjugated wave, i.e. a wave with the same complex field propagating in the opposite direction. In the context of multimode fibre, this was demonstrated firstly using different holographic approaches [105] [106]. During the past few years, an alternative was proposed in order to remove the holographic material, called "digital phase conjugation", that was applied to multimode fibres, both in order to focus and scan a spot at the output [107]. This was made possible thanks to the introduction of devices that could vary the phase of a laser beam, such as the spatial light modulator (SLM).

In parallel, another method to compensate the distortion was introduced under the name "transmission matrix" (TM). This approach, although it might be defined in different ways, cuts across different disciplines such as acoustic, seismology or communication, because it could be used for scattering problems due to propagation through disordered media or for interference phenomena for antennas communication. Within these contexts, basically it states that the propagation between two arrays, input and output, can be described by a single matrix T. This approximation of the propagation just makes sense if the system can be considered effectively linear and deterministic. Indeed, this corresponds to the present case as well, taking into account the description of multimode fibres previously proposed. Thus, the relationship that maps the input x of the output y could be simply written as $y = Tx$, or more explicitly as $y_i = \sum_j T_{ji} x_j$. In order to empirically measure T, the first step is to define a base for the input and a base for the output. In this way, by propagating each of the input-base elements and measuring the relative outputs, the transmission matrix T can be completely described. As can be noticed, supposing that the propagating field is complex, control of both amplitude and phase for both input and output is

required to obtain the complex TM. This request can be easily matched for acoustic and electromagnetic waves where commonly available point-like emitters and receivers work both in phase and amplitude. However, this is not the case in Optics, where it is more problematic to shape the illuminating field. Furthermore, obtaining the output phase requires dedicated techniques, usually interference with an external beam.
On the other hand, early attempts to control light at the input of a disordered medium in order to unscramble its output field can be found around ten years ago. With the developement of wavefront shaping devices some "brute-force" methodology, i.e. direct search of the best input-phase pixel-configuration in order to focus the beam in specific points, was applied to both complex media and multimode fibres [108][109]. The measurement of the empirical transmission matrix represents an extension of this technique as it aims to provide a complete or more general description of the input-output relation and, therefore, of the transmission matrix itself that in the previous approaches was treated as an unknown.
The transmission matrix was firstly measured by Popoff et al. for an opaque material and was used for both focusing [110] and retrieval of a transmitted image [111]. In both the cases, the idea was to collect the speckle patterns resulting from the self-interactions between an internal reference beam, in order to avoid stability issues, and the different elements of the Hadamard base [112]. In the same period, other approaches were proposed in order to obtain the transmission matrix, as scanning the angle of incidence of the input beam [113]. Instead, Cizmar et al. where the first applying their own optimisation method [114] also to multimode fibre [115][116] using input beamlets generated by an SLM as input and providing control over the full complex field, polarization included.
A number of papers have been exploring the possibility of enabling the utilization of multimode fibres as endoscopes, with techniques that involve measuring the transmission matrix [102] [117] or rely on digital phase conjugation [118].
Furthermore, a major effort was carried out by Cizmar's group [119] to demonstrate that not only the input-output relationship can be measured for multimode fibres, a deterministic system, but also the internal propagation can be predicted and controlled. This was made possible by considering the cylindrical symmetry of the system and thus developing a complex model taking into account possible practical-theoretical system differences and contribution from internal imperfections. This result strongly differentiates a multimode fibre from a complex fully-disordered medium. On the other hand, despite the surprising results achieved, such as allowing bending in a controlled environment, this technique could only work with fibres up to 30 cm long. Regarding bending, another work investigated an analogous but much simpler model made possible by switching from a (commonly used) step-index multimode fibre to a parabolic-index, or quasi-parabolic-index fibre as the GRIN one [120].

In conclusion, it is important to note that all the approaches seen so far were experimentally

enabled by the introduction of control over phase in the system. This might represent a problem in particular for the output of the system where phase-measurement techniques might involve interferometry approaches, leading to possible stability issues. Furthermore, it has been appreciated how the two most common techniques, digital phase conjugation and transmission matrix, differ in the amount of measurements required, the digital phase conjugation being a single-shot technique. On the other side, TM requires a minimum of measurements dictated by the number of elements of the input base (the output base is typically larger or has the same size) but grants a more complete description of the input-output relation. For practical use, a paper demonstrated that the resolution of a multimode fibre, exploited for imaging, is not limited to the number of input modes defined by the user, as is the case for TM, but instead could reach up to four times the number of modes propagating into the fibre [121]. Finally, the SLM refresh rate has been one of the main factors limiting the speed of these techniques, in particular if scanning is involved or a large amount of measurements need to be collected for characterising the system. Another slowing factor might be the presence of demanding calculations for encoding information to the SLM holograms. In order to speed up the experiments, one option is to move from a SLM to a digital micromirror device (DMD) [122] [123] [117], i.e. arrays of micromirrors that can quickly switch between two "on/off" positions, which, on the other side, complicates the control over the full complex field. Another option is to rely on GPUs whenever high-performance operations are involved [124] [125].

3.2.1 Neural networks for multimode fibre Imaging

A different route is to rely on artificial neural network algorithms. As previously discussed in section 1.3.1, this approach has been introduced by Tanida's group in 2016 for scattering media (in particular resin plates) in [49]. One year later, they applied different data-driven algorithm, including an ANN as well, to distinguish between two classes of images, containing a face or not, encoded on a laser beam by a phase-SLM and propagated through a multimode fibre [126]. Remarkably, no use is made of any knowledge about possible mathematical model describing the system. During the last two years, this paved the way to further neural network approaches, resorting to deep learning, that aimed to go beyond simple classification in order to unscramble the speckle patterns and thus retrieve the original images. Firstly, in [127] reconstruction of phase-encoded images from the relative output intensity speckle pattern was demonstrated for multimode fibre up to 1 km long. This result is remarkable because the transmission matrix of the system, for that fibre-length, is continuously varied by temperature variations and even small mechanical vibrations, even providing a good isolation. The same group explored the possibility to reconstruct, from intensity-only detection of the speckles (camera measurements), images encoded on amplitude or phase, in [128]. Interestingly, they proposed a method to measure the TM and verify its validity by projecting patterns at the output of the fibre. Moreover, Fan et al. [129] proposed a neural network approach to deal with a fibre subjected to bending. Similarly

to [127], also in this case the TM is dramatically modified because of changes in local refractive index along the fibre. However, in this case the system can be controlled better as the bending configuration of the fibre is known.

On the other hand, most of these works made use of a restricted class of images: the hand-written digits - taken from MNIST database (28×28 pixels images), which has been already introduced in chapter 1 -. As a consequence, the images that are meant to be reconstructed must belong to the same, or perhaps very similar, class of images. In this sense, a first step towards generalisation of the retrieval process was shown in [127] although limited to almost-binary and stylized simple images, such as an heart or the symbol "plus". However circumscribed, this result opened up the possibility for a broader application of ANN to generic imaging.
Finally, a paper by Turpin et al. [130] explored the possibility given by ANN algorithms for focusing through a complex medium and a multimode fibre. In this case, both single layer and deep neural network were tested. Moreover, they demonstrated that it is possible to relate transmitted and reflected patterns with an ANN. Once a first ANN has learnt the mapping between transmitted and reflected speckles, the appropriate holograms that control the illumination of the system, in order to freely focus a desired pattern at the output, can be predicted just by training a second ANN with the reflected speckles.

Within this context a novel approach is being proposed, in this thesis, which could take advantage of some physical insights related to the fibre in order to approximate the inverse of the transmission matrix. As a result, the present method has proved to be able to transmit images characterised by an unprecedented high-resolution and at high frame rates, even without involving any phase measurement. This was made possible thanks to a non-deep neural network by resorting, instead, on a single but complex-valued layer (because of the complex-valued nature of the light propagating within the fibre).
After having offered an introduction to the background of imaging through a multimode fibre, I will present in the next sections of this chapter the novel complex-valued approach starting from the experimental layout, then moving to the method and results.

3.3 Experimental layout and data

The layout with which the experiment is carried out is shown in Fig. 3.3. A laser beam, wavelength $\lambda = 532$ nm, is modulated by means of a phase-only SLM (refreshing rate ~ 20 fps) in conjunction with a polarizing beam splitter and a half-wave plate. In this way, grayscale intensity images (with values ranging from 0 to 100) are imprinted on the laser beam. The resulting images are coupled into a multimode fibre using an objective with NA $= 0.26$ and focal length $f = 34$ mm. The multimode fibre is characterised by a step refractive index profile and diameter

$d = 105$ μm. The number of modes allowed in these conditions is around 9000. Furthermore, two different fibres are employed with same characteristics but different lengths: 1 m and 10 m, both wrapped in a loose coil lying on the optical table. Another identical objective along with an imaging lens are used to image the resulting speckles at the output of the fibre on a CMOS camera. The images are collected at 350×350-pixel resolution with 255 grayscale levels.

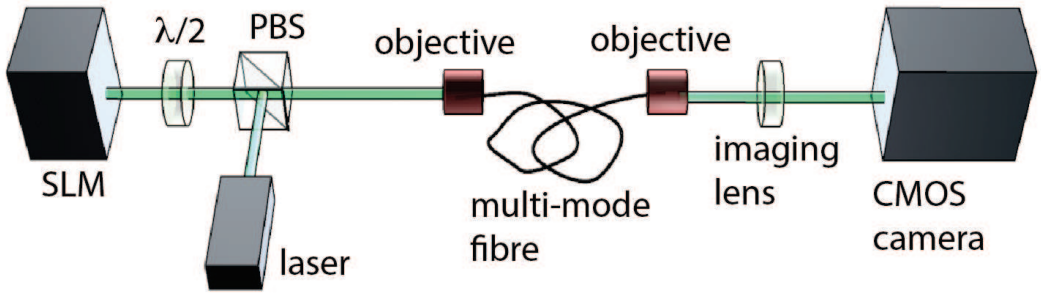

Figure 3.3: Scheme of the experimental layout. A phase only SLM is used to encode amplitude images on a laser beam thanks to the polarizing beam splitter (PBS) and the half-wave plate ($\lambda/2$). The laser is a CW laser with wavelength $\lambda = 532$ nm. The resulting images are coupled into a multimode fibre using an objective (NA $= 0.26$, focal length $f = 34$ mm). The speckle patterns are coupled out with an analogous objective and imaged on a camera by means of an imaging lens. The fibres employed are step-index, $d = 105$ μm, with lengths 1 m and 10 m. Figure adapted from [2].

In Fig. 3.4, different examples of input-output pairs are reported. The size of the input grayscale images is 28×28. The shown examples belong to four different categories of images: handwritten digits from the MNIST database, cloths from the Fashion-MNIST database [131], elements belonging to the Hadamard base and random patterns. Fashion-MNIST is a database developed by Zalando consisting of 60000 training images and 10000 testing images grouped into 10 different classes of cloths (shoes, jumpers, trousers etc...), with relative labels provided .

Along this chapter, other classes of employed images are: samples from the Muybridge collection, images from the ImageNet database and examples of natural scenes. Muybridge iconic works, that dates back to 1870s, pioneered modern cinematographic techniques by showing projections of pictures in motion [132]. He was particularly interested on representing the motion of animals, such as horses or birds, and humans. In the present case, the samples have been scaled to the required resolution. In the same way, the ImageNet database, previously introduced in chapter 1, along with other examples of natural scenes have been resized specifically for high-resolution image transmission. Since generic images from the ImageNet database are not square, a random selection of almost square images was considered and resized in order to obtain square images. All the data used can be found following the DOI link [133].

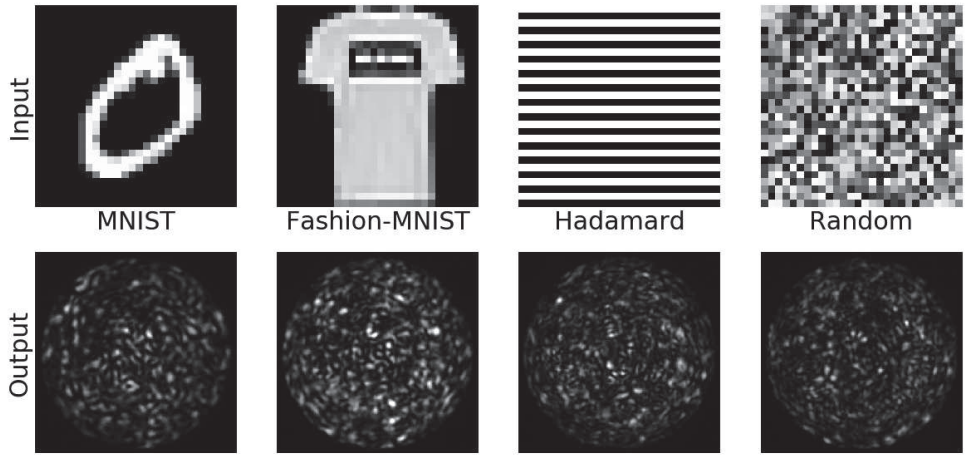

Figure 3.4: Examples of input patterns and output speckles relative to the introduced setup. Input images pixel-resolution is 28×28, with 100 grayscale levels. Output speckles are recorded at 350×350 pixels with 255 grayscale levels. Figure reprinted from [3].

3.4 Method

There are two possible ways to restore an image after being propagated through a multimode fibre, and thus distorted. On the one hand, an operator that maps the input image into the distorted one could be defined, measured and then inverted. On the other hand, the inverted operator could be directly determined. The approach introduced in this chapter considers this second route by inferring statistically, on the basis of the large amount of images employed, the relationship that maps output speckles into input patterns.

In order to achieve this, a single complex-valued fully-connected layer, mimicking the inverted transmission matrix itself, is employed. As a matter of fact, on one side the physical transmission matrix of a real fibre is complex-valued, on the other side the operation of the fully-connected layer, linking input to output, coincides with a matrix multiplication. Considering in particular Fig. 3.1, one could point out that local-approaches, as for example convolutional layers and max-pooling, and therefore any deep convolutional neural networks, should be avoided because of the system property of connecting a single input point to the whole output. For this reason, the choice felt on the fully-connected layer, opening up to possible challenges related to GPU RAM memory. On top of that, in the present experiment the phase of the speckle patterns is not collected. As a result, the addressed problem is severely underconstrained.

The idea of using neural network with complex values has been around for many years. Takeda and Kishigam, back in 1992, presented a pioneering neural network that could vary both amplitude and phase [134]. As pointed out by Hirose [135], complex-valued ANN are of paramount interest for real physical phenomena such as the ones involving wave propagation. In fact, the possibility to work directly with amplitude and phase, i.e. working with complex-related

operations, already imposes a constraint on the learning process related to the nature of the system; clearly, this is different from considering neural networks dealing with unrelated double-dimensional real numbers. Within the work of Hirose an overview of the results achieved by complex-valued ANN is offered as well. Furthermore, different works have explored the possibility to build deep neural network with complex values. In particular, Trabelsi et al. made a solid contribution to this context [136] providing the main components for a complex-valued DNN, such as complex activation function, complex convolutions and so on.

In Fig. 3.5 a representation of the complex-valued ANN method is proposed. As mentioned earlier, the output of the multimode fibre, i.e. the speckle field, is now the input of the neural network, as the inverse problem is directly addressed. Since just the intensity of the speckles is measured, no phase information is present in the input. On the other hand, the fully-connected layer is complex valued, therefore the input-phase will left blank. As reported in the figure, a model respecting faithfully the physical nature of the system should map the input complex field into the output complex field, and therefore should be considering the amplitude as input. This is effectively the approach that has been followed within this very chapter, when dealing with high-resolution images. However, at first it has been explored the possibility to work directly with intensity field, both for input and output. This work has been reported in Appendix B.1; in

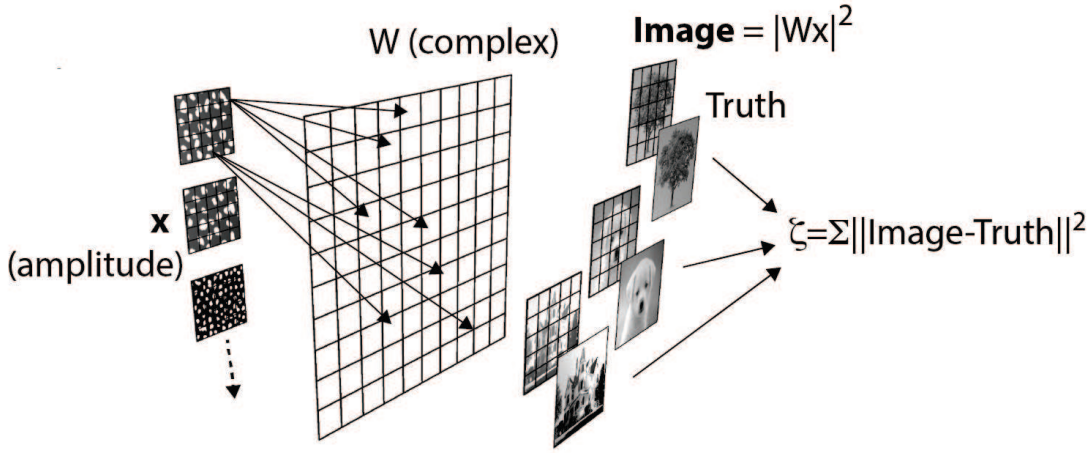

Figure 3.5: Schematic overview of the inverse problem method. A single complex-valued fully-connected layer W is used to map the input, the speckle field x, into the output image. In fact, fully-connected simply operates as Wx, which is used to approximate the inverse of the transmission matrix. The learning process consists in allowing the network to see several input-output pairs. In this way, the weights are modified in such a way that the distance between predicted images and ground truth image reaches a minimum. In other to achieve this, a cost function ζ is introduced with a metric (MSE) to quantify this distance. The partial derivatives of the cost function respect to the elements of W are calculated in order to apply stochastic gradient descent, which pushes the cost function to a minimum value. This iterative process is repeated for all the training images and for a certain number of epochs in order to assure convergence. Figure adapted from [2].

this case other measures has been employed to deal with possible non-linearity involved.

In contrast, the model reported in Fig. 3.5 is entirely linear. Here, the amplitude of the input speckle x, i.e. the square root of the intensity acquired on the camera, along with the left-blank phase, is passed through the fully-connected complex-valued layer W. This operation coincides with the matrix multiplication: Wx. Furthermore, to preserve linearity the activation function is just linear (since the layer is complex, the activation function works separately on real and imaginary part). Subsequently, the cost function ζ is built comparing the amplitude of the output $|Wx|$ to the amplitude of the original image (ground truth). Similarly, the square of the output $|Wx|^2$ could be compared directly to the intensity of the original image, since this was encoded on the beam as an intensity image, as reported in Fig. 3.5. The metric used to compare the prediction of the network to the ground truth is the mean square error (MSE). At this point, to calculate the corrections to be applied to the weights composing W, the derivatives $\partial\zeta/\partial w_{ij}$ with respect to the i, j elements of W are first calculated. Then, the elements of W are varied, applying small changes in order to minimize the cost function. This is realised using the stochastic gradient descent approach. This process is repeated for all the speckle and ground truth image pairs of the training dataset and iterated over a fixed number of epochs in order to allow ζ to converge to a minimum value. The regularisation applied to ζ is the l_2 norm. In Appendix B, it is discussed how this was implemented for the present complex-valued network. To ensure that the system has effectively learnt an approximation to the inverse problem, the network is tested with unseen speckle and ground truth image pairs, i.e. the testing dataset. Keras library and Tensorflow are employed to implement the network. In particular, the complex layer W is built as a custom layer. The code is provided and explained in Appendix C.

3.5 Results - Transmission of natural scenes

In this section, I will present the results for the linear model previously introduced. The aim is to reconstruct generic images such as pictures of everyday life, hence "natural scenes". In this way, this challenge poses itself on a different level of complexity relative to the ones addressed so far with ANN approaches. In fact, these techniques were demonstrated and could work exclusively on limited classes of images, such as hand-written digits. However, in these cases, the network could be working mainly as a classifier. If this is the case, it can be shown, for example considering hand-written digits, that the classification task can be easily solved using a common real-valued fully-connected network (see Fig. 3.6).

The training dataset, considered for the complex-valued method, is composed by a selection of 50000 images belonging to the ImageNet database. The size of the images is 92×92, which is comparable to the number of modes propagating into the fibre. The speckle field images

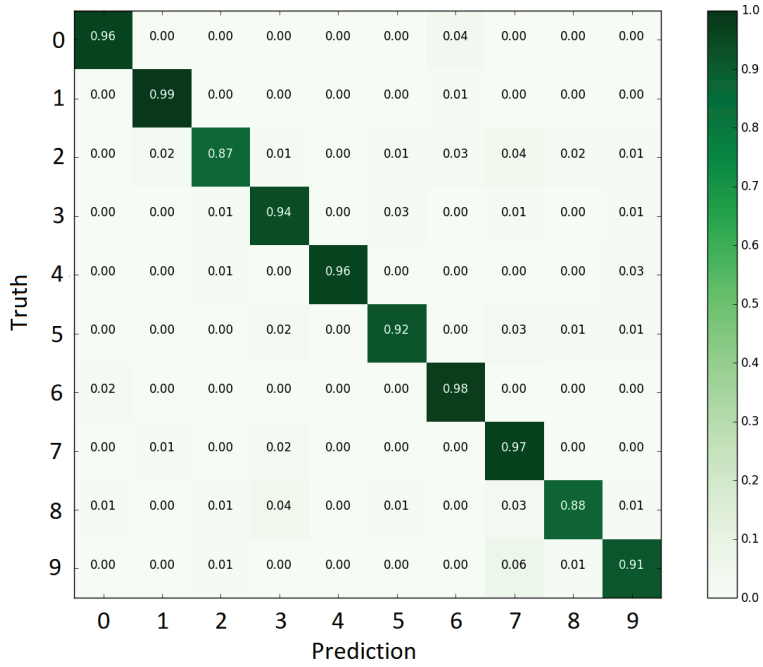

Figure 3.6: Confusion matrix representing the retrieval performance of a real-valued fully-connected network on images taken from the hand-written digits MNIST database (with pixel-resolution 28×28), encoded in amplitude (as shown in 3.3) and propagated through a 1m-long multimode fibre. The network is trained with 5000 speckle patterns (350×350-resolution) and test on 1000 speckles patterns. Figure reprinted from [4].

are collected at the resolution of 350×350 pixels but resized to 120×120 pixels in order to avoid GPU memory issues. As previously mentioned, both the speckle images, collected with the camera, and the fibre input patterns, encoded on the beam, are intensity fields. However, the linear complex transmission matrix operates on the input and output full complex fields (amplitude and phase). Thus, the square root of these intensity fields will be extracted and employed for the corresponding input and output amplitude fields of the network; whereas, the phase fields will be left blank.

Once the model is defined, the network is trained, using the training dataset, over 850 iterations. This process took around 48h on the employed GPU (Nvidia GeForce GTX 1080 Ti). After this operation, it can be verified whether the network has learnt a good approximation to the inverted transmission matrix using images belonging to a different database, different from that used for training, as a test set, by comparing the predictions to the ground truth images. Fig. 3.7 reports the performance of the complex-valued linear approach for different Muybridge videos for the 1 m-long fibre. In particular, a running horse, i.e. one of the most characteristic Muybridge's subjects, a jumping cat, a flying parrot and a "punching" man. The amplitude speckle field is shown along with the intensity ground truth image and prediction, i.e. the square of the amplitude reconstructed image.

In order to quantify the quality of the reconstructed images, two different metrics are employed: the Pearson Correlation Coefficient (PCC) and the Structural Similarity Index (SSIM) [137].

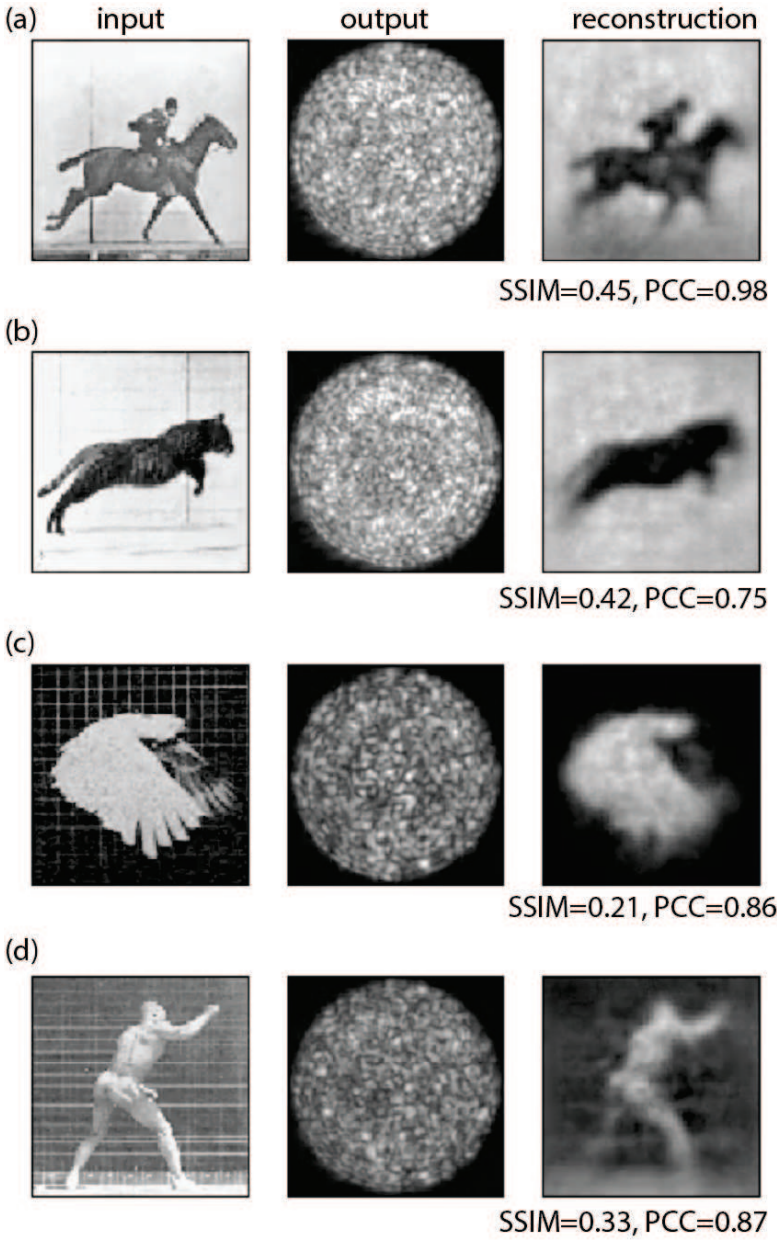

Figure 3.7: Muybridge videos reconstruction. The frames are transmitted independently through a 1-m long fibre at 4 fps. Image pixel resolution is 92×92. Speckle resolution is 120×120. Here, a single frame, from the full videos (available at [133]), is shown for, respectively: a) running horse, b) jumping cat, c) flying parrot, d) punching man. From left to right, the columns show the ground truth images, reported as intensity, the amplitude speckle field and respective intensity reconstruction (square of the amplitude prediction). Figure reprinted from [2].

Considering two images U and V, the PCC is defined as follows:

$$PCC(U,V) = \frac{\sum_i (u_i - \bar{U})(v_i - \bar{V})}{\sqrt{\sum_i (u_i - \bar{U})^2 \sum_i (v_i - \bar{V})^2}}, \tag{3.1}$$

having the sum operator $\sum$ iterating over the index i, which refers to the i-th pixel of the considered image: respectively, u_i for U and v_i for V. The operator $\bar{\cdot}$ indicates the average of the image.

The values of the PCC parameter ranges between $+1$, in case of a perfect positive correlation, and -1, in case of a negative correlation. For the 0-value, there is no correlation.
The formula describing the structural similarity index is

$$SSIM(U,V) = \frac{(2\bar{U}\bar{V} + C_1)(2\sigma_{UV} + C_2)}{(\bar{U}^2 + \bar{V}^2 + C_1)(\sigma_U^2 + \sigma_V^2 + C_2)}, \tag{3.2}$$

σ_{UV} being the covariance of U and Y, σ_U^2 the variance of U and σ_V^2 the variance of V. The parameters C_1 and C_2 are defined as follows: $C_1 = (K_1 L)^2$ and $C_1 = (K_2 L)^2$, where $K_1 = 0.01$ and $K_2 = 0.03$. L is the range defined by the difference between the maximum and minimum values that the image pixels can assume (also called dynamic range). The maximum value that the SSIM can reach is $+1$ as well, and its range is once again between $+1$ and -1. Typically, SSIM is adopted to focus the comparison between images more on visual similarities.

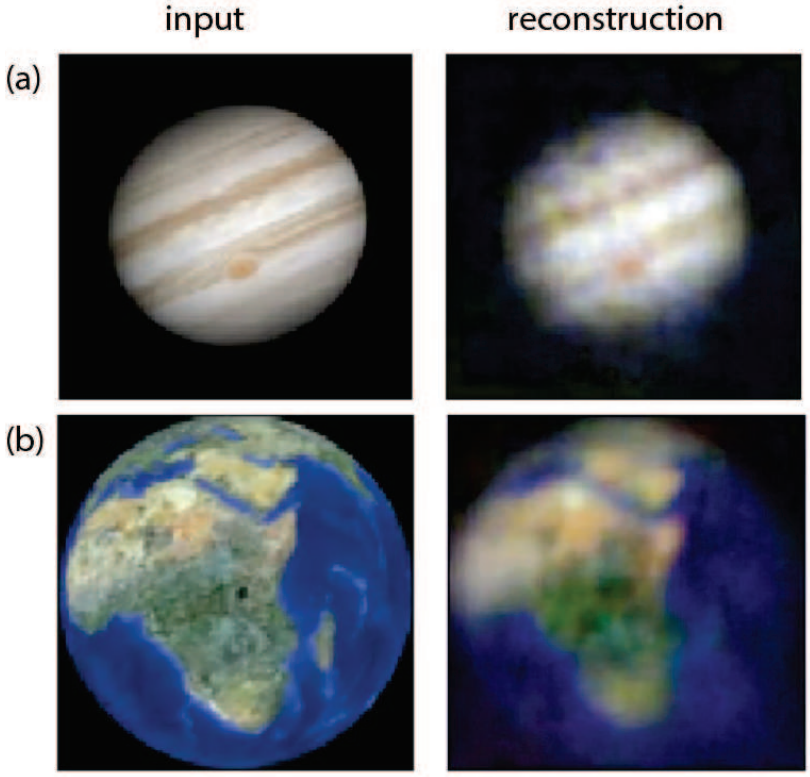

Figure 3.8: Full-color video reconstruction. Single frames are reported for RGB videos of (a) a rotating Jupiter and (b) Earth. The full videos are available at [133]. RGB channels are transmitted and restored separately, then recombined. The fibre employed is the 1m-long one. The inverted transmission matrix W is the same learnt on grayscale ImageNet samples. Figure reprinted from [2].

Fig. 3.8 shows the reconstruction of full-color images, transmitted by encoding separately the 3 "RGB" channels. Examples of a rotating Earth and Jupiter are presented. The training database and procedure are the same used for Fig. 3.7 (also in this case for the 1m-long fibre). Then, the estimated complex-valued W is employed to retrieve independently the three grayscale channels. The full-color image reconstruction is obtained simply by recombining the three channels. As it

can be observed, remarkable characteristics of the Earth and Jupyter can still be spotted, such as the african desert areas or the Red Spot.
Additional examples are shown in Fig. 3.9 and Fig. 3.10. The former figure presents examples of everyday-life natural scenes representing the University of Glasgow in full-color RGB format. A result for the 10m-long fibre is reported as well: the transmission of a panda picture. Other 10m-long reconstruction will be shown subsequently in section 3.5.2. The latter figure shows the performance of the present algorithm on selected images extracted from the ImageNet database but not employed for the training process itself.

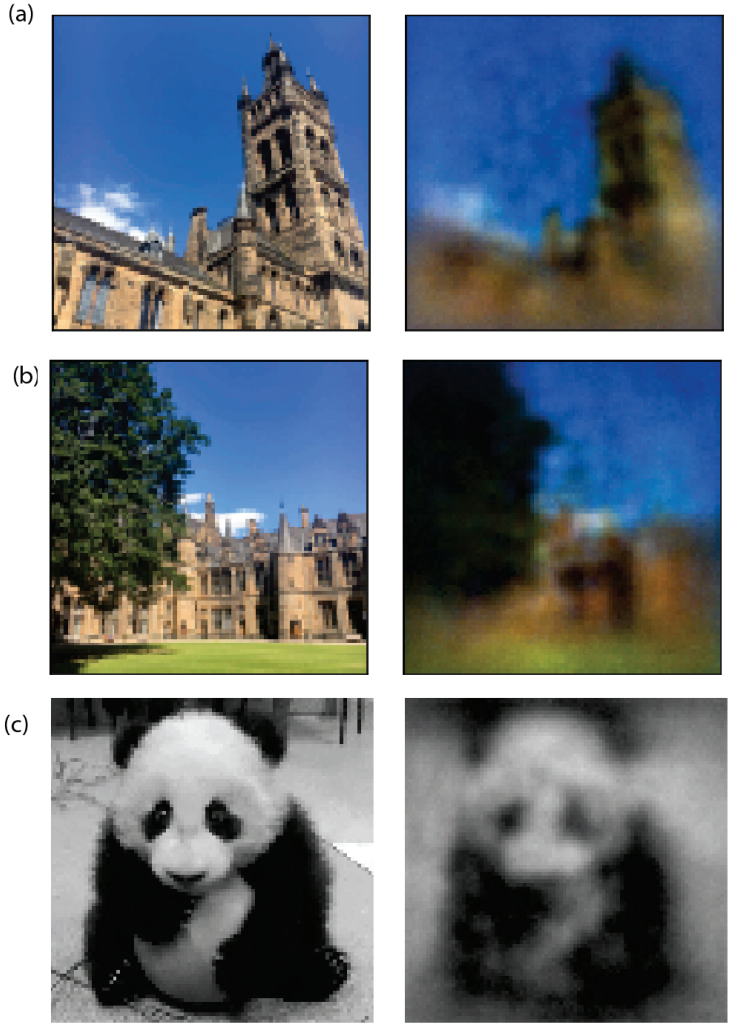

Figure 3.9: Further natural scenes examples. Full colour photographs of the University of Glasgow are presented in (a)-(b). In the left column, the original images are shown. The right column shows images reconstruction after propagation through a a 1m-long fibre. Whereas, for the case of a multimode fibre with length $= 10$ m the image of a grayscale panda is shown in (c). In both the cases, the respective networks have been trained with 50000 samples from the ImageNet database. Figure reprinted from [2].

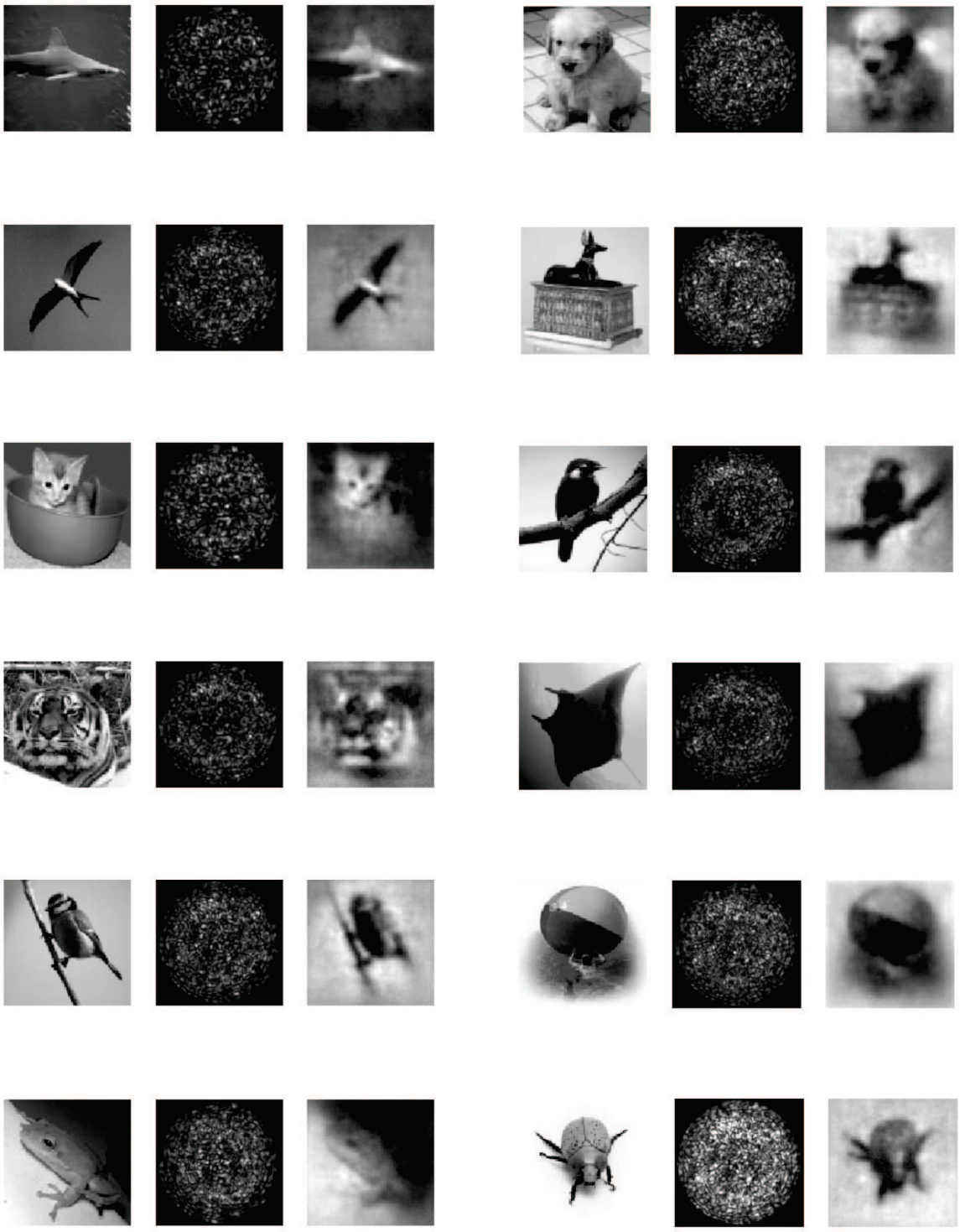

Figure 3.10: Reconstruction of images selected from the ImageNet database. This samples are not used for the training process. Once again, the left column shows the ground truth, the center the speckle field and the right the prediction. Figure reprinted from [2].

3.5.1 Generalisation

As has been shown, the complex-valued ANN approach has provided a solid reconstruction of transmitted images that do not belong to the same database as the one used for the training process. However, a further confirmation for the generalising capacity of the present method is here provided by relying on what I will indicate as a more "agnostic" approach. The basic idea is to train the network with a completely "not-informed" training dataset. Thus, it made use of 50000 random grayscale patterns, i.e. no assumption is made on the class of the transmitted images. On the other hand, in order to simplify the problem, the ground truth patterns are reduced to 28×28-pixels, thus avoiding possible memory issues (related to the RAM of the GPU). Nonetheless, the results in Fig. 3.11 show how the network is still able to restore the images, although with less quality. Moreover, reconstructions are provided for both the 1 m and 10 m fibres, demonstrating once again that this approach does not strictly depend on the length of the fibre.

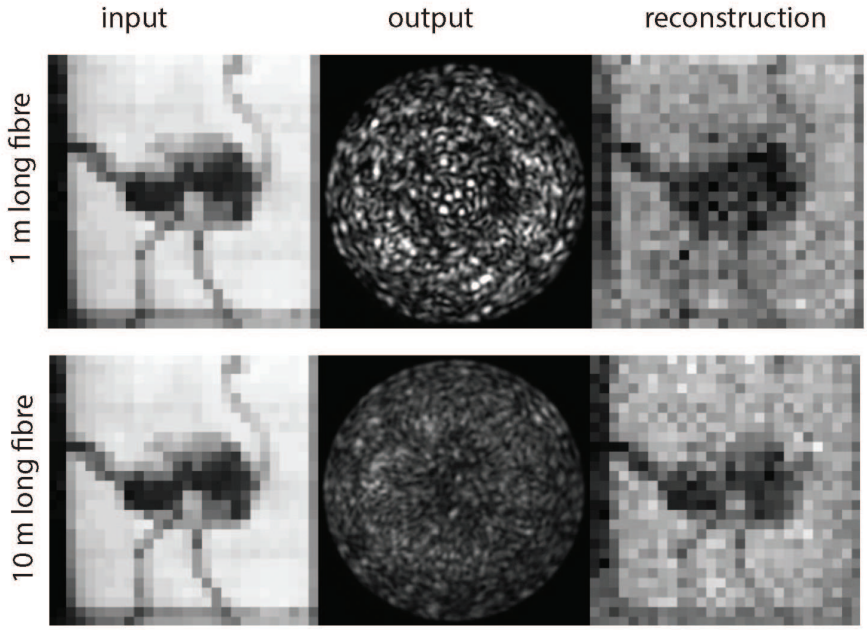

Figure 3.11: "Agnostic" approach results. Here, the network is trained with 50000 grayscale random patterns. Nonetheless, the algorithm is still able to restore an input pattern transmitted through both the 1-m and 10-m long fibres. On the left column is shown the fibre input pattern encoded on the beam, on the central one the speckle and on the right one the reconstruction. Figure reprinted from [2].

3.5.2 Impact of different focusing input configurations

In this section, the effect of varying the image size at the input of the fibre is studied. The same experiment proposed so far is repeated for three different focusing configuration. In particular, the input images are demagnified by factors $1\times$, $1.5\times$ and $4\times$. This is realised by means of a

telescope used after the SLM. Thus, three different approximations to the inverted transmission matrix W are estimated by training the ANN with the usual ImageNet database (the 1m-long fibre was used in this case) and employing these different configurations. The resulting performances are compared in Fig. 3.12.

As the image at the input of the objective is decreased in size, respectively 10 mm, 7.5 mm and 2.6 mm, the focusing angular spread is meant to decrease as well. As a consequence, the "resolving power" achieved by the objective will differ for the three distinct cases. This phenomena can be quantitatively addressed by considering the effective numerical aperture NA_{eff}, which can be calculated knowing the objective focal length ($f = 34$ mm). For the demagnifying factors $1\times$, $1.5\times$ and $4\times$, the respective effective NAs are $\mathrm{NA}_{eff} = 0.22$, $\mathrm{NA}_{eff} = 0.16$ and $\mathrm{NA}_{eff} = 0.05$. As it can be observed in Fig. 3.12, the experimental results indicate how the complexity of the speckle field increases for a larger effective NA, in fact the number of the speckles is increased and their size reduced. This is a signature of the amount of modes excited within the fibre. Since in the present experimental conditions the pixel-resolution of an input image (92×92) is comparable to the number of modes of the fibre, clearly the greater the number of modes to which the image is coupled the better the final reconstruction.

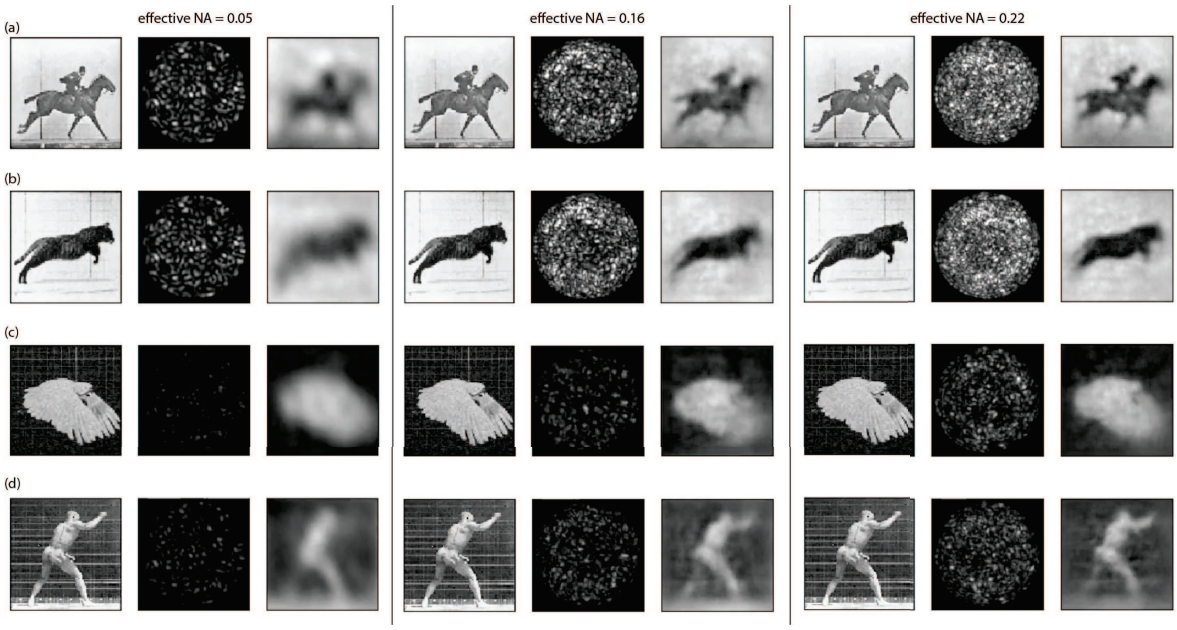

Figure 3.12: Effect of different focusing configurations. From left to right, the same experiment is repeated, on a 1 m-long fibre, varying the image size at the input of the objective, which couples the light into the fibre. As a result, by increasing NA_{eff} the speckle field presents a deeper complexity, signature of the excitement of more modes (when retaining similar experimental conditions). Evidently, encoding the information on many modes leads to a better reconstruction. In fact, in this particular experiment the pixel resolution is comparable with the number of modes allowed by the fibre. Figure reprinted from [2].

It can be observed that there are two other different aspects that can add complexity to the output speckle field. On one side, by considering Fig. 3.11, it can be noticed how the increase in length

typically generates thinner speckles. In fact, for longer fibres the number of manufactoring imperfections are increased. Moreover, there could be more bent segments for the fibre (in the present case, the fibres are just loosely rolled up). Consequently, the starting information could be coupled on more modes while propagating. Therefore, in this case thinner speckles do not immediately indicate an increase in the amount of modes excited at the fibre input. However, as it can be appreciated in Appendix B.2.1, this could potentially lead to issues related to the resolution of the speckle images. For example, it might happen that in a particular experiment the speckles are so small that the required resolution will not fit the GPU RAM memory.
On the other side, it has been experimentally observed that the placement of the focused beam onto the fibre input has an important role regarding the modes on which an image is coupled. As a matter of fact, if the focused beam is placed on a very central position the image tends to couple to relatively few modes with circular symmetry. This phenomenon is highlighted by the fact that the speckles tend to accumulate in the centre of the output and ring-like structures start to appear (signature of few high modes), as stated also in [107]. However, to address the need for the excitation of many modes, the beam is displaced slightly with respect to the center. In this way, it is possible to find a configuration that allows high-resolution image reconstruction.

3.5.3 Time degradation

Finally, another important aspect is explored: the degradation of the reconstruction over time. As is well known, a multimode fibre can be subjected to different situations, such as bending or temperature gradients, that can alter locally the refractive index. Whenever this happens, the transmission matrix will be dramatically modified as a result. In this way, a good question is whether the system is able to restore correctly the transmitted ground truth images even after some time has passed after the training process. This question is particularly important respect to the fact the neural network takes a certain amount of time to learn the input-to-output mapping function. In the present experiment the neural network is trained for 48h.

As a proof of the robustness of the present method with time, the performances of the neural network at different times are compared. The first step is to train the complex-valued neural network with the ImageNet database, thus obtaining W (i.e. the approximation to the transmission matrix). Then, the operator W is employed to restore the video of the rotating earth, shown in the example reported in Fig. 3.13, transmitted at 4 different times after the transmission of the training dataset: 1h, 16h, 40h and 52h. The multimode fibre used here is the 10 m-long one. Quantitative analysis is reported, using the SSIM metric. As a result, it can be observed how the correlation between the speckle fields and speckle field used as a reference, namely the 1h one, still remains high; the same applies to the single-channel retrieval reported in Fig. 3.13 (b). It is interesting to note that the experiment has been carried out with the multimode fibre wrapped in a loose coil and placed on the optical table. Therefore, the present method proved to be robust

even in absence of a specific controlled environment.

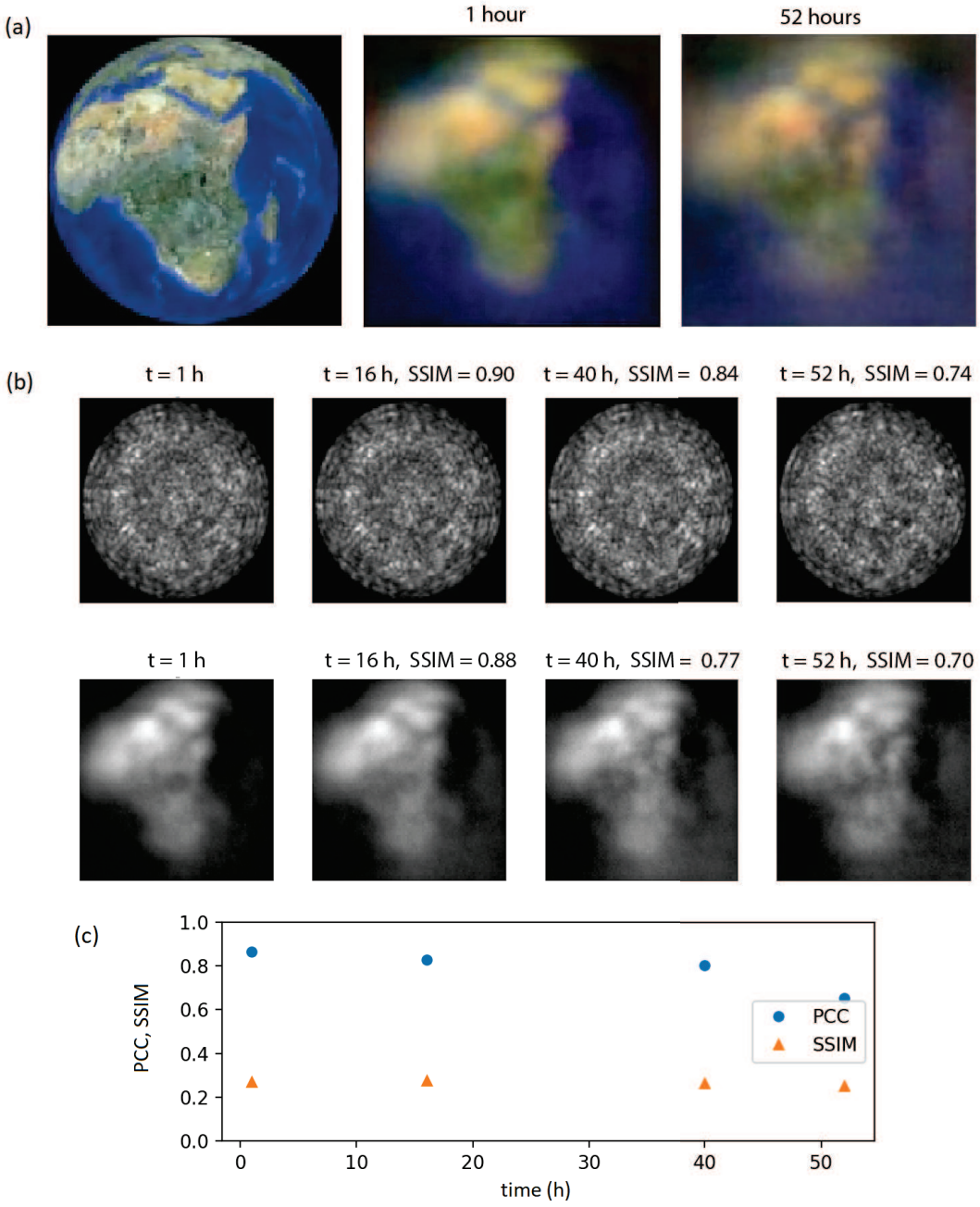

Figure 3.13: Time degradation. Individual frames extracted from the video of a full-color rotating Earth transmitted through a 10 m-long fibre. The reconstruction provided by the matrix W, calculated through the ANN algorithm, is tested for the same ground-truth image transmitted at different times (1 hour, 16 hours, 40 hours and 52 hours after the transmission of the training dataset). In (a), the full-color reconstruction after 1h and 52h is shown. In (b), it is provided a quantitative analysis (using the SSIM parameter) of the temporal degradation of the speckle field and the restored images with respect to the 1h ones (green channel). In (c), the same image predictions shown in (b) are compared to the relative ground truth. This result demonstrates a certain robustness to time degradation even if the fibre is not placed in a controlled environment. Figure reprinted from [2].

In conclusion, an alternative approach for restoring images transmitted through a multimode fibre has been proposed and demonstrated. The present method is based on solving directly the inverse problem even without any phase control of the output of the system. A complex-valued single layer neural network is employed for approximating the inverse of the matrix that maps the input field into the output field. In this way, it has been possible to demonstrate transmission of high-resolution images, with the number of pixels comparable to the number of modes propagating in the fibre, at high speed. As a matter of fact, this is a single shot technique and, thus, is basically limited just by the refresh rate of the device used to imprint the images on the beam (in the present layout an SLM). The speed of the present approach allows the transmission of full-color (RGB) images, possibly at video-rate, by separately reconstructing the single-shot-transmitted channels. Not only has this method has provided good reconstructions for fibres with different lengths (1 m and 10 m), but it has also shown a solid robustness with respect to time degradation, which typically affects imaging through a multimode fibre.

For future work, an interesting point to investigate could be the possibility to physically inform ANN algorithms in order to deal specifically with those factors that deteriorate the reconstructions in time, such as possible bending or temperature gradients. In contrast to other deep neural network approaches, the present method has shown important results related to the problem of generalisation. As a matter of fact, the ANN algorithms employed so far have been limited to the retrieval of images belonging to the same, or very similar, classes of the training dataset samples. Instead, considering the results reported in chapter, the ANN could reconstruct the ground-truth images even when trained with uninformed patterns (random patterns). Finally, the impact of different input-focusing configurations has been studied in relation to the ability to retrieve the transmitted information.

Main limitations are related to the commercially-available GPU RAM memories, as a matter of fact a fully-connected layer, as the one here employed, scales with the size of input and output images, as $O(R_i^2 R_o^2)$, where R_i is the input pixel resolution and R_o the output pixel resolution.

Chapter 4

A comparison of techniques: transmission matrix and complex-valued network

In the previous chapter, the problem of imaging through multiple scattering media has been introduced and addressed through a novel approach based on a complex-valued neural network algorithm. In particular, this method has been applied to a multimode fibre but, in principle, it might be extended to every scenario in which the information is propagated through a number of transmission channels. Indeed, the idea behind this technique is simply to identify the operator that maps an input field to the output. Moreover, as opposed to most of the ANN approaches that typically rely on a number of real-valued hidden layers, in this case the algorithm consists of a single, but complex-valued, fully-connected layer. As explained previously, this solution is driven by the physics involved in the problem. In fact, the operator is expected to be linear - the fully-connected layer corresponds to a matrix multiplication - and complex-valued.

Formulated in this way, the definition of the problem seems analogous to the one of the transmission matrix approach. As a matter of fact, recalling the review offered in Chapter 3, the aim of this technique is exactly to provide an empirical matrix describing the input-output mathematical relationship. Of course, the problem is once again linear and complex valued, thus requiring a method to acquire the output phase - in an optical context, most of the approaches that provide phase-measurements are based on interferometry techniques, which, in turn, might lead to stability issues.
The procedure for measuring the transmission matrix consists simply in defining both an input and an output base and measure, for each element of the input base (N pixels-resolution), the relative output (M pixels-resolution). Thus, the transmission matrix, expressed in the previously-defined input and output basis, is built simply arranging the measured outputs in columns, so that a $N \times M$ matrix is obtained. Given an input pattern x, the output distorted pattern y and the transmission matrix T, the mathematical expression will be

$$y = Tx. \tag{4.1}$$

I will refer to this as the forward problem.

In contrast, the complex-valued ANN approach aims at restoring the input pattern from the relative output, finding the matrix V that solves directly the inverse problem:

$$x = Vy. \tag{4.2}$$

As opposed to the transmission matrix approach, here the matrix of interest is inferred statistically by letting the algorithm learn the transformation by analysing a number of input-output training samples. Moreover, the measurement of the transmission matrix provides a solution for the forward problem, thus requiring a matrix inversion in order to solve the inverse problem. As is well known, in the presence of noise this matrix inversion might be non-trivial. Nonetheless, a solution was provided, in Optics, by Popoff et al. in [110]. Their method was tested on an opaque scattering material and provided the inverted matrix T^{inv} for solving

$$x = T^{\text{inv}}y. \tag{4.3}$$

Given the analogies between these two techniques, one might ask how they compare to each other. In fact, the complex-valued ANN approach is the only method, among the ANN methods introduced in the previous chapter, that identifies a single complex-valued matrix, V in (4.2), that solves the exact same mathematical complex-valued expression reported in (4.3). On the other hand, how could the two matrices V and T^{inv} be compared? Since there is not a reference matrix, no metric can provide a sense of the distance between these two matrices. In this sense, the idea of this chapter is to perform this comparison in two different ways:

- comparing the performance of the two matrices on the same measured data y;
- comparing the effect of each matrix to the expected physical behaviour.

This will offer the possibility to explore the transmission matrix approach proposed by Popoff et al. and its physical implications, perform this method on a multimode fibre - instead of a opaque scattering material -, test the complex-valued ANN on a complete complex-valued input field - up to now this has been applied just on amplitude-only data - and finally study the physical relevance of the matrix obtained with this neural network algorithm.

4.1 A transmission matrix inversion approach for complex optical media

As reported in Chapter 3, the transmission matrix technique can be broadly applied to many fields of Physics and Engeneering. In Optics, compared to other areas of research, the difficulties related to the ability to control and detect phase delayed the introduction of this approach. As has been said, in 2010 Popoff et al. [111] successfully demonstrated the transmission matrix technique to solve the forward problem for an opaque material. Their method involved an internal reference beam to determine the output phase. Shortly thereafter, they provided the solution for the inverse problem [110]. Furthermore, in 2015, the same group proposed an alternative approach in order to allow the characterisation of the full complex-valued transmission matrix by means of intensity-only measurements [138].
In this section, I will shortly discuss the methodologies and the results related to the first two papers. This will serve to build the comparison with the complex-valued ANN approach by introducing the procedures that have been employed to acquire data, calculate the transmission matrix and its inverse and analyse the transmission properties of the matrix. This approach has been considered for the possibility to operate also with the full-complex output field, i.e. full-complex ANN input.

In order to study the physical properties of a transmission matrix, the "singular value decomposition" is employed. This states that a generic $p \times q$ matrix G can be decomposed into

$$G = U\Sigma V^{\dagger}, \tag{4.4}$$

where U and V are two unitary matrices that represent, respectively, the set of left and right "singular eigenvectors", i.e. the eigenvectors for $GG^{\dagger}$ and $G^{\dagger}G$. The relative eigenvalues, for both the matrices, are the squares of the elements on the diagonal of Σ that are also called "singular values". These elements are real and non-negative, whereas the matrix is zero everywhere else. The number of singular values is dictated by the minimum between p and q. Basically, the singular values might be visualised as coefficients for the transmission channels (or eigenchannels) that map each input mode into one and only one output mode. In Optics, considering a transmission matrix that links the input and output electric fields, each singular value indicates the intensity transmission coefficient related to a particular channel.

A schematic representation of the layout of the experiments is reported in Fig. 4.1. Here, a SLM is employed to imprint patterns on the phase of a laser beam. As it can be observed from the cross-section of the beam as it entres the objective "Obj1", part of the beam is fixed and used as an internal reference, whereas the other part is dedicated to controlling the input pattern x. A set

of two objectives, “Obj1” and “Obj2”, are employed to focus the beam on the scattering opaque material and to image the distorted output pattern on a camera.

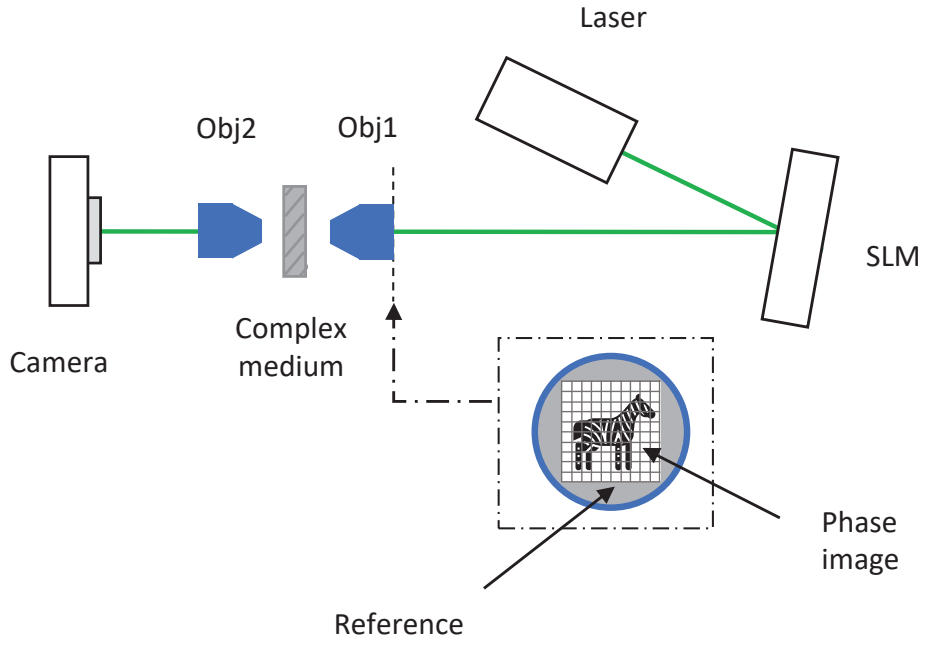

Figure 4.1: Scheme of the layout used for measuring the transmission matrix for a multiple-scattering medium. A laser beam is modulated in phase by means of an SLM, leaving part of the modulated beam as a reference. The phase pattern is focused on the material and the resulting speckle is imaged on a camera thanks to the two objectives. Beam cross-section as light entres Obj1 is also shown.

The full complex electric output field is calculated by means of the “four phase method” [139]. This is an interferometry technique which states that shifting in phase an input pattern x by $e^{i\gamma}$ with $\gamma = 0$, $\pi/2$, π and $3\pi/2$, and measuring the relative intensity outputs I_γ, the electric field y at the output is described by the equation:

$$y = \frac{1}{4}\left((I_0 - I_\pi) + i(I_{3\pi/2} - I_{\pi/2})\right). \tag{4.5}$$

The transmission matrix T provides the transformation that maps the N input modes into the M output modes. Considering the setup in Fig. 4.1, the number of elements of the input and output basis are fixed by the choice of the relative pixel-resolutions. In this sense, regardless of the actual optical propagation inside the medium, the transmission matrix is basically a pixel-to-pixel transformation. Both in [111] and [110], the choice for the input base fell on the Hadamard patterns, which is easy to implement on a SLM and which, compared to the canonical base, achieves a better signal-to-noise ratio - in the canonical base each element has only one non-zero pixel. For each element of the Hadamard base, x_n, four measurements are collected to

obtain the output y_m. In this way, the transmission matrix is defined through the equation

$$y_m = \sum_{n=1}^{N} T_{mn} x_n. \tag{4.6}$$

It must be noticed that the two vectors x_n and y_n are expressed in two different bases, i.e. respectively Hadamard base and canonical base (since the output has been measured with the camera). Thus, a change of base is required if one is willing to have both the vectors expressed in the canonical base. The advantage of this system lies in the fact that it makes use of a single camera and does not rely on an external reference beam, which can reduce the stability of the system. Because of the interference with the reference beam, it can be easily shown that the measured matrix T does not coincide with the real matrix, which I will denote as $\mathscr{T}$, but presents an added contribution from the reference beam, that I will indicate with R, so that $T = \mathscr{T} \cdot R$.

Using the time reversal operator $T^\dagger$, the authors of [111] and [110] were able to calculate the input patterns to be applied to the SLM in order to achieve focusing on one or more spots at the output, showing the dependence of the ratio between the single intensity focused spot and the averaged intensity elsewhere, from the input degrees of freedom (upper boundary for the ideal case). Moreover, they demonstrated that the time reversal operator can be employed for the reconstruction of simple input patterns, e.g. single or multiple dots.

4.1.1 Inverse operator

As mentioned earlier, the retrieval of an arbitrary image was achieved shortly afterwards, in [110], by proposing a different operator for the inverse problem. From an experimental point of view, the main issue with matrix inversion is related to the presence of noise. Considering the inverse operator T^{-1}, defined through $T^{-1}T = I$ with I indicating the identity, or equivalently the pseudo-inverse operator $[T^\dagger T]^{-1}T^\dagger$, in case $N \neq M$, since the singular values of T^{-1} correspond to the inverse of the singular values of T, the strongest contributions will be provided by the lowest singular values. Those below the noise level threshold will thus corrupt the retrieval process.
An alternative is the time reversal operator previously used for focusing. This operator has been extensively studied in Acoustics [140] (this thesis considers the monochromatic case). Time reversal tends to favour the strongest channels, thus is more robust with respect to noise. However, the relative reconstructions are degraded when the elements outside the diagonal of $T^\dagger T$ are not null.

A possible solution is to resort to "zeroth-order Tikhonov regularisation" for the problem in

(4.1). It defines the operator:

$$K = [T^{\dagger}T + \lambda I]^{-1}T^{\dagger}. \tag{4.7}$$

Interestingly, this solution lies between the time reversal and the inverse operator. In fact, looking at the two limit behaviours of the parameter λ in (4.7): for $\lambda = 0$ the operator collapses into the matrix inverse, whereas for $\lambda \to \infty$ into the time reversal. In this sense, the application of this operator can be viewed as a way to balance the contributions from strong and weak transmission channels for the inverse problem. In particular, it can be shown [141] that, indicating with σ_i the singular value of T, the effect on the inverse solution is to apply "filter factors":

$$f_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda} \tag{4.8}$$

that modulate the contributions from each singular value. As it can be appreciated, for $\sigma_i^2 \ll \lambda$ $f_i \sim 0$, thus avoiding aberrated strong contributions.

By averaging over a certain number of disorder realisations, the authors of [111] and [110] reported the restoration of an input image with $N = 32 \times 32$ pixels, distorted after being propagated through the opaque multi-scattering material, in good accordance with the ground truth. Moreover, they compared the results for the operator K and the pseudo-inverse varying the output pixel resolution $M = \gamma N$, with $\gamma \geq 1$. Increasing the ratio γ, it can be observed that the distribution of the singular values - which will be discussed extensively in the next section - tends to be concentrated around a certain (non-zero) value. In doing so, the value of the lowest non-null singular value must increase as well. Remarkably, the authors demonstrated that when there are no singular values below the noise level anymore, the pseudo-inverse becomes equivalent to the operator K.

4.1.2 Random matrix theory

The statistical properties of the transmission matrix can be analysed in order to extract possible relevant information about the physical nature of the material and the propagation within it. In Acoustics, different works have exploited the singular value decomposition to study the wave propagation through media composed of point-like scatterers. For a small number of these scatterers ($< N$) and neglecting multiple scattering, it can be demonstrated that there is an almost one-to-one relationship between the scatterers and singular values of the transmission matrix [142]. In the opposite case, a large number of scatterers randomly disposed might lead to an overall propagation that can be considered random [143]. For these chaotic media, the distribution of the singular values of the transmission matrix can be associated to the statistics of some random matrices. This is the basic assumption of the so-called "Random Matrix Theory" (RMT) [144], which has been applied to different fields in Physics [145][146].

In particular, considering the normalised singular values

$$\widetilde{\sigma}_i = \frac{\sigma_i}{\sqrt{\sum_{j=1}^{N} \sigma_j^2}}, \tag{4.9}$$

the Marcenko–Pastur law [147] affirms that, under certain conditions (such as elements uniformly distributed in $[-1,1]$ and mean equal to zero), the distribution $\rho(\widetilde{\sigma})$ of the normalised singular values of a $N \times M$ (with $M = \gamma N$) random matrix can be described by the expression

$$\rho(\widetilde{\sigma}) = \frac{\gamma}{2\pi\widetilde{\sigma}} \sqrt{(\widetilde{\sigma}_{max}^2 - \widetilde{\sigma}^2)(\widetilde{\sigma}^2 - \widetilde{\sigma}_{min}^2)}, \tag{4.10}$$

for all the $\widetilde{\sigma}$ in the interval $[\widetilde{\sigma}_{min}, \widetilde{\sigma}_{max}]$, having $\widetilde{\sigma}_{min} = (1 - \sqrt{1/\gamma})$ and $\widetilde{\sigma}_{max} = (1 + \sqrt{1/\gamma})$. In Fig. 4.2, the distributions of the singular values for a random matrix, for varying γ, are reported along with the theoretical curve. The particular case of $\gamma = 1$ (Fig. 4.2(a)) takes the name of "quarter-cycle law". As mentioned earlier, the effect of increasing γ leaves the system the possibility to explore and exploit transmission channels with similar normalised singular values (the distribution tends to converge toward $\widetilde{\sigma} = 1$).

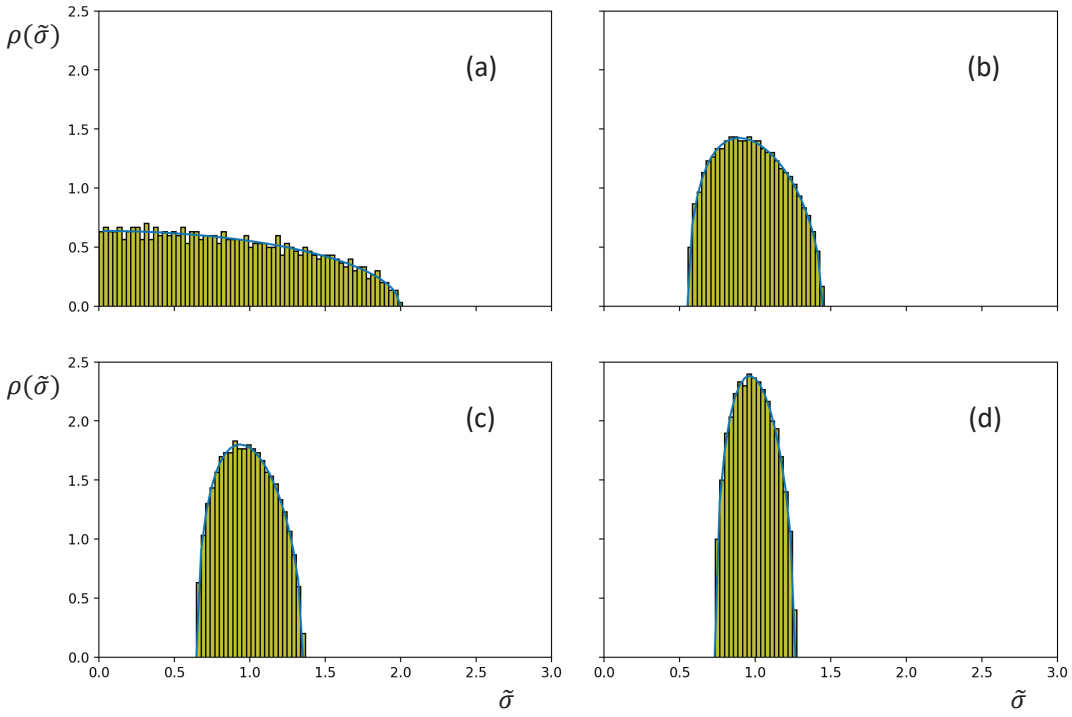

Figure 4.2: Examples of distribution of singular values for a $N \times M$ random matrix, with $N = 1000$ and $M = \gamma N$. Histograms for the density of singular values and the respective curves predicted by the Marcenko–Pastur law are reported for the following different cases: (a) $\gamma = 1$, (b) $\gamma = 5$, (b) $\gamma = 9$, (b) $\gamma = 14$.

Popoff et al. were able to demonstrate experimentally singular value distributions in good agreement with the expected behaviour indicated by the equation (4.10). However, in order to avoid

spurious correlations, they had to follow essentially two main steps: removing the reference beam contribution and avoiding neighbouring pixel correlations. For the former, the effect of the reference R on the transmission matrix can be expressed explicitly as

$$T_{mn} = \mathscr{T}_{mn} R_{mm}, \tag{4.11}$$

with $\mathscr{T}$ indicating the real transmission matrix and T the one measured. Since the reference is the same for every input, the contributions will be the same along each one of the M lines (every output pixel receive the same reference contribution from each input element). In this sense, the amplitude of the R_{mm} can be determined by calculating the standard deviation of T on every line:

$$\sqrt{\langle |T_{mn}|^2 \rangle_n} = \sqrt{\langle |\mathscr{T}_{mn}|^2 \rangle_n} |R_{mm}|. \tag{4.12}$$

Dividing the equation (4.11) by the equation (4.12), a new filtered transmission matrix T^{fil} is obtained where the contributions from the reference have amplitude equal to 1. Thus, it can be demonstrated that the singular values for T^{fil} and $\mathscr{T}$ are the same.
Considering the neighbouring-pixel correlations, the authors of [111] and [110] removed one out of two consecutive elements for each column and row of the matrix, creating a $N/2 \times M/2$ truncated matrix, in order to avoid possible short range correlations (possibly due to the instrumentation itself). However, increasing γ, the authors noticed that the singular values distribution tends to show a relatively small deviation from the behaviour predicted by the RMT theory. This is probably caused by short range pixel correlations whose results are more difficult to suppress.

4.2 Technique comparison: experimental layout

The experimental layout, employed to compare the transmission matrix and the complex-valued neural network techniques, is illustrated in Fig. 4.3. The system is analogous to the one used in Chapter 3, however, the SLM is now replaced by a DMD for an overall quicker data acquisition, input patterns are phase-encoded and data are recorded according to the four-phase method.
Phase-only patterns are imprinted on a CW laser beam (wavelength = 532 nm) by means of the DMD. Since the DMD consists of an array of micromirrors that can only work in a binary "on-off" configuration, thus reflecting the light in two possible directions, phase modulation is achieved by means of a particular approach named "Lee Holography" [148] [122]. This approach involves the generation of appropriate amplitude patterns to be filtered in the Fourier plane (i.e. a grating is programmed ensuring that the light pattern at the first diffraction order contains the desired phase and amplitude distribution). In the proposed layout this filtering operation is realised by means of a telescope (with lenses L1 and L2) and a pin-hole (PN). Moreover, the binarised nature of the DMD imposes a level of approximation that must be compensated

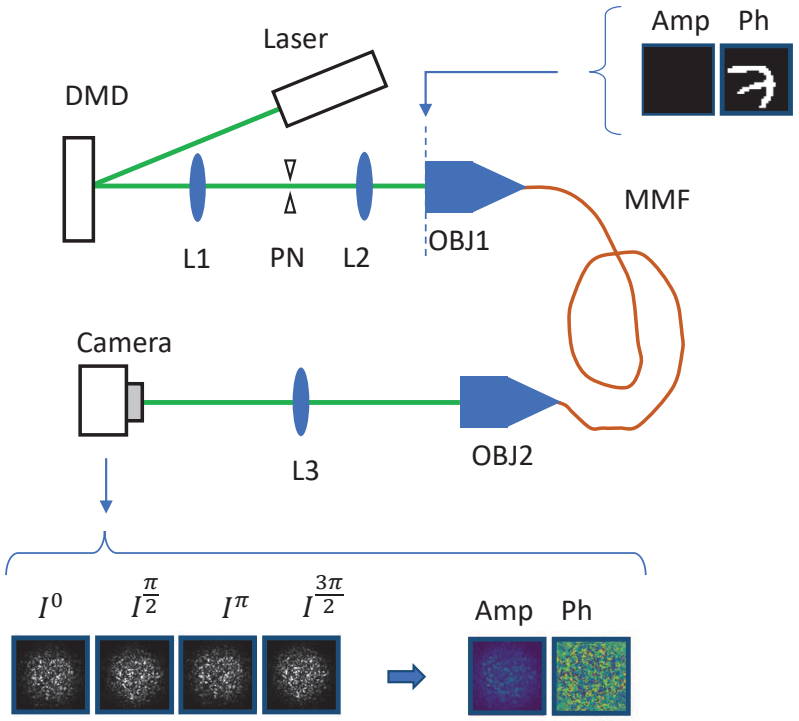

Figure 4.3: Experimental setup for the technique comparison. A laser beam (wavelength = 532 nm) is modulated in phase by means of a DMD. The field needs to be filtered with a pin hole (PN) placed in the Fourier plane of a telescope realised with the two lenses L1 and L2. Then, a first objective (focal length = 34 mm, NA = 0.26) is used to couple the phase-only pattern, like the "7" image (whose expected field has been reported in the figure), in the multimode fibre (step-index, length = 1 m, $d = 105$ μm). A second objective (identical to the first) and a lens (L3) are used to image the output onto the camera. The reference, part of the image 4.1, has been omitted for simplicity. Examples for the phase-shifting technique are reported: 4 speckle images acquired to obtain the full-complex output field.

by resorting to macro-pixels, i.e. mapping each single pixel of the amplitude pattern on a larger square composed of many DMD-pixels. In this way, phase-only binary input patterns can be easily generated, as in the example shown in Fig. 4.3, and then coupled into the multimode fibre (MMF) thanks to an objective (OBJ1: NA = 0.26, focal length $f = 34$ mm). In particular, the pixel resolution of the input patterns is 28×28; the multimode fibre has a step-index profile with an inner-core diameter $d = 105$ μm, and is 1 m-long. The resulting distorted pattern at the output of the fibre is imaged onto a camera (recorded at a 192×192 pixel-resolution) thanks to a second objective (OBJ2 with same characteristics as OBJ1) and a lens (L3).

The need for input phase-encoding is dictated by the way the interferometry method, employed to record the full-complex output field, has been formulated in the previous section. In fact, the four-phase method relies on recording four different intensity speckle images for the same input pattern, each one corresponding to a different phase-shift, stepping $\pi/2$, applied to the input patterns. As an example, in Fig. 4.3 the four intensity images relative to the input "7" phase-image are reported along with the full complex field calculated through equation (4.5). Once the patterns are loaded into the memory of the DMD, limited to about 80K patterns, the

images are displayed at a frequency of 1 KHz. As for the setup shown in Fig. 4.1, part of the DMD (surrounding the input pattern) is used to encode the internal reference beam, but is not shown for simplicity.

The datasets used for the input images are: hand-written digits (MNIST), clothing (Fashion-MNIST), Hadamard base and random patterns. All these patterns have been binarised (having for 0, phase $= \pi/2$, and for 1, phase $= 3\pi/2$) and share the same 28×28-pixels resolution. As a consequence, the Hadamard base has 28×28 elements. The output speckle images are recorded as grayscale images with 256 levels.

4.3 Method

In this section, the two methods considered for the solution of the inverse problem are presented. Since both the approaches have been previously introduced, in Chapter 3 and in the present chapter, here I will just highlight the most relevant aspects and differences related to the analysis reported in this work.

The complex-valued neural network algorithm is an approach able to directly solve the inverse problem, thus without any need for a matrix inversion technique. The main element of this algorithm is the single but complex-valued, fully-connected layer. Formally, the application of a fully-connected layer corresponds to a matrix multiplication reproducing the problem described in equation (4.2). As illustrated in Fig. 4.4, the input of the neural network is the speckle field recorded at the output of the multimode fibre. However, in contrast to Chapter 3, where only the amplitude was recorded, in this case the full-complex field can be accessed (thanks to the four-phase method). Furthermore, the patterns that are coupled at the input of the fibre are now encoded on the phase of the beam, therefore not on the amplitude as done in the experiment presented in Chapter 3. Thus, the final output layer, this time, must extract the phase from the result of the complex valued multiplication, as Fig. 4.4 shows. In this way, the cost function - that will be minimised - compares this phase pattern to the respective ground truth pattern. I will refer to this method as the "full-complex" network approach.
The code is essentially equivalent to the one presented in Appendix C, with a small number of modifications related to the changes explained within this section.
Alternatively, the same network can be employed with amplitude-only input speckle patterns, as for the experiment reported in the previous chapter. This can be achieved by selecting just one out of the four speckle patterns recorded by means of the four-phase method (since these are intensity images, the square root is extracted), i.e. fixing a particular input phase-shift. I will refer to this approach as the "single-shot" network, as it relies on a single camera acquisition.

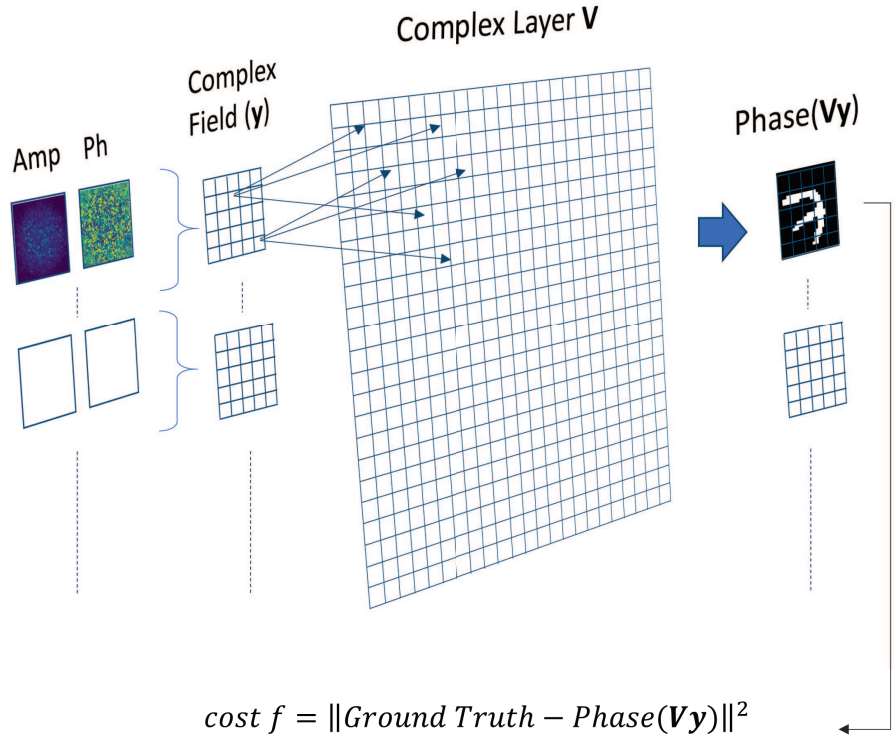

Figure 4.4: Schematic representation for the complex-valued neural network. The algorithm is analogous to the one in Fig. 3.5, however, in this case full-complex input field is used. Moreover, the phase is extracted at the output and compared to the ground truth image to be retrieved.

Finally, the transmission matrix approach is considered. As has been described, the transmission matrix T, which characterises the forward problem, is acquired relying on the Hadamard base. Then, the matrix is inverted exploiting the Tikhonov regularisation, i.e. the operator K is defined as expressed in the equation (4.7). This leaves the freedom to determine the parameter λ which balances the operator K between the pseudo-inverse operator and the time reversal operator. A possible estimation for λ can be obtained, for example, just by looking at the singular values of T. In fact, referring to equation (4.7), to make the two terms $T^\dagger T$ and λI comparable, λ must be comparable to the singular values of $T^\dagger T$, i.e. to the square of the singular values of T.
Another approach to determine the parameter λ is through the experiment itself, i.e. employing a correlation coefficient to compare the retrieval performances of K to the parameter λ, for different values of λ. Once the transmission matrix has been acquired, the operator K can be applied to different testing images in order to build a statistic. In Fig. 4.5, the Pearson correlation coefficient, already introduced in section 3.5, is employed to evaluate the reconstructions of 800 MNIST and 800 Fashion-MNIST phase-only patterns propagated through the multimode fibre and recorded both in amplitude and phase. The plot shows the behaviour of the averaged results for the Pearson correlation coefficient to varying the square root of λ, for the two test sets. Thus, the value of λ for best retrieval performance is estimated by averaging the two λs relative to the maxima of the two respective test datasets.

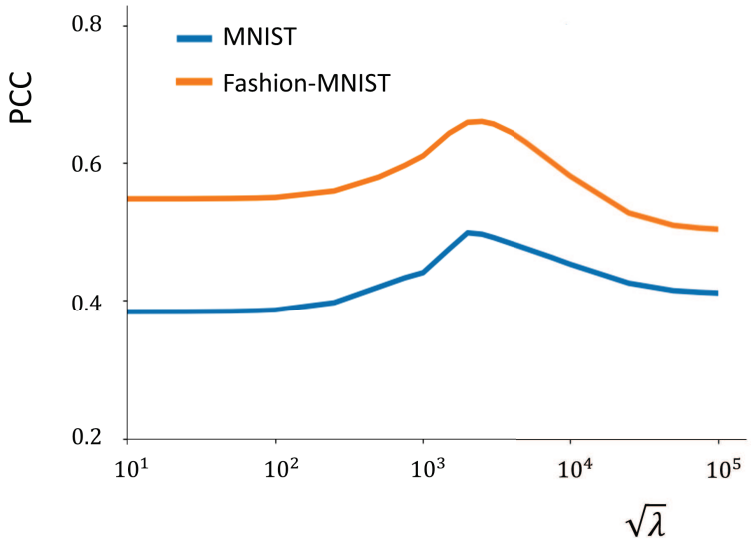

Figure 4.5: Evaluation of the Tikhonov-regularisation inverse operator K to varying the parameter $\sqrt{\lambda}$. The operator K is tested on 800 MNIST and 800 Fashion-MNIST patterns propagated into the fibre in the same experimental conditions. In this way, the parameter λ for best retrieval performances can be obtained.

Moreover, Fig. 4.5 can help visualising the performances for the pseudo-inverse operator (in the limit $\lambda \to 0$) and the time reversal operator (in the limit $\lambda \to \infty$) as well. As has been discussed in section 4.1.1, by increasing the ratio between the controllable output modes and the input modes, the system starts to rely on less small-valued singular values. Consequently, noise has less influence on the final reconstruction (in fact, as it can be observed in Fig. 4.5 the inverse operator leads to good reconstructions). A direct comparison with Popoff et al.'s results for a random medium is not considered because, as I will show later, the measured singular values distribution differs from the RMT expected behaviour. Regarding possible sources of noise, instabilities might arise in this experiment from the laser, external perturbation applied to the multimode fibre, or amplitude and phase distortions relative to the DMD generated input patterns and imperfect coupling at the input of the fibre.
To summarise, both the approaches leads to a matrix with size $M \times N$, complex-valued, that solves the same inverse problem. In case the network input data y are the same as those used for the inverse transmission matrix K, the two problems are formally identical.

4.4 Results

In Fig. 4.6, a comparison between the performances of different neural network models and the transmission matrix inversion operator K is shown. As for Fig. 4.5, the different methods are tested on two datasets: 800 images of handwritten digits from the MNIST database and 800 images of clothing from the Fashion-MNIST database. Referring to the neural network algorithms,

these testing data are not included in the training process. Moreover, once again the Pearson correlation coefficient is employed to evaluate the goodness of each retrieval process. In this way, for each method average and standard deviation of the resulting coefficients are calculated, over the two testing datasets, and reported in the bar plot.

The presented methods can be subdivided into three main categories:

- single-shot neural network: single-shot amplitude-only speckle measurements;
- full-complex neural network: full complex field speckle measurements;
- transmission matrix inverse operator K.

The neural network approaches are thus differentiated by the input y of the inverse problem. Instead, the full-complex ANN and the transmission matrix (TM) inversion approaches share the same input which is obtained through the phase-shifting method presented. For the input of the single-shot ANN, the first out of the four intensity images, obtained through the four-phase method, is selected and the square root is extracted. Inside each category, the datasets and the amount of data employed for calculating the respective matrices are reported for each method.

As it can be observed, in Fig. 4.6, the best results are achieved by the network trained with full complex field samples consisting of 5000 random patterns, 5000 MNIST images and 5000 Fashion-MNIST. Moreover, to demonstrate that the network can generalise and properly reconstruct images from datasets that are not present in the training process, the results for the full-complex ANNs trained with random patterns are reported. In this case, the networks are trained with a different amount of samples - 5000, 2500 and 1000 - in order to observe the impact of the amount of training data on the retrieval process. Interestingly, when considering the full-complex samples, the network seems unable to retrieve the images if the phase is excluded, i.e. with amplitude-only samples (training with 5000 random patterns). On the contrary, the respective phase-only case, in the same conditions, leads to good reconstructions. For the single-shot neural network, results are reported when training the network with MNIST and Fashion-MNIST databases, with a total of 1000 samples or 3000 samples.

The reported bar plot allows also a direct comparison of the performances of the transmission matrix inversion operator K, already observed in Fig. 4.5, relative to the full-complex neural network trained with the same data, i.e. the Hadamard base. In this case the TM inversion method seems to outperform the relative neural network. Thus, one might wonder if there is any case in which the neural network could provide better reconstructions relative to the TM having at its disposal an equal or smaller number of training data. Referring to Fig. 4.6, this happens for the single-shot network trained with a total of 3000 samples (MNIST and Fashion-MNIST), respect to the 784×4 measurements (Hadamard) required for the TM (4 acquisitions are needed

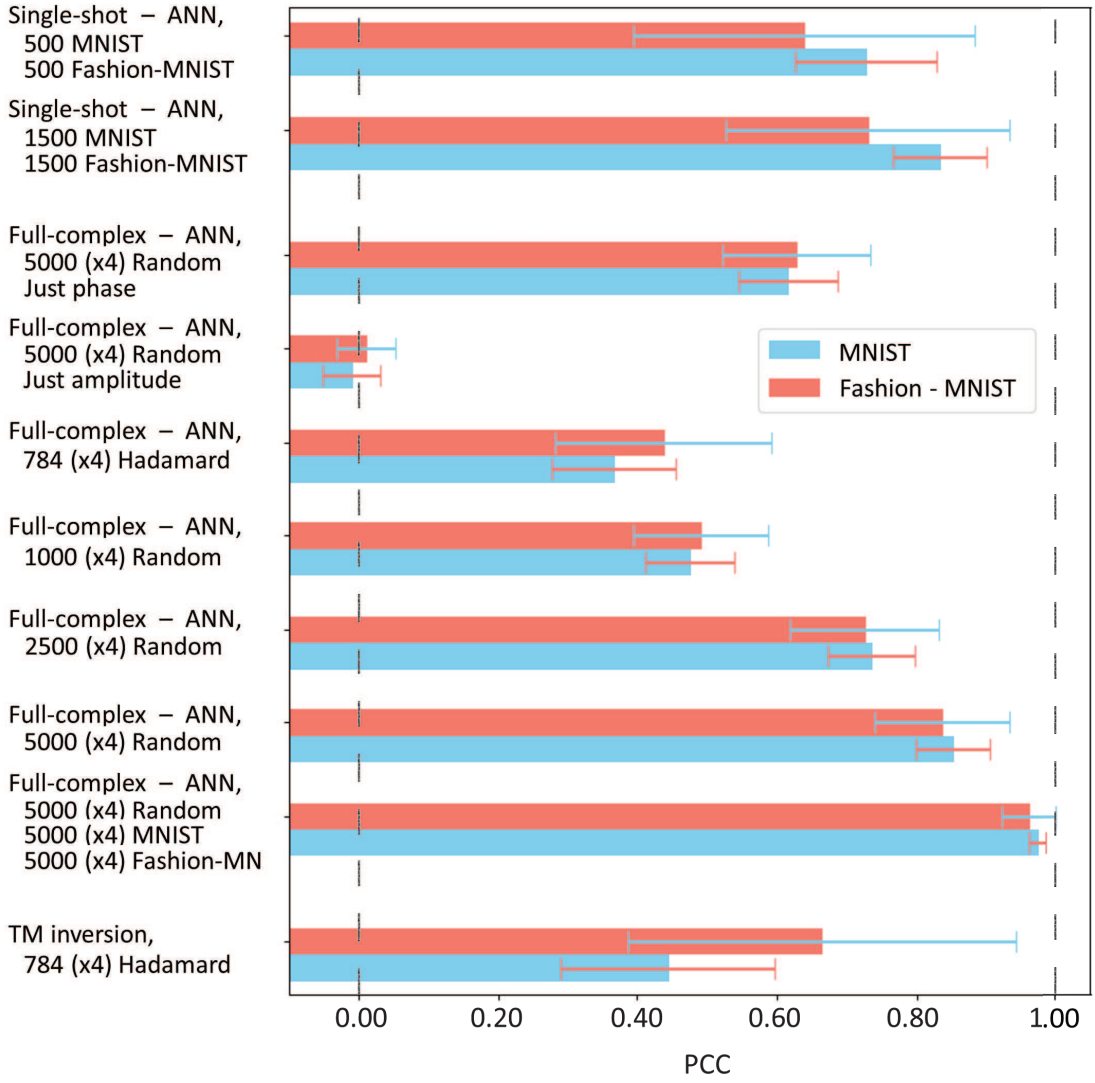

Figure 4.6: Comparison of the performances of different methods for the reconstruction of the input pattern. The performances are calculate in terms of Pearson correlation coefficient (PCC). As for Fig. 4.5 (c), the test images are 800 MNIST and 800 Fashion-MNIST and are not part of the training process. Three different categories of approaches are presented, from the bottom: TM inversion technique, ANN with complex speckle input field and ANN with “single-shot" data as input (therefore just amplitude). The datasets used for training the networks and calculating the TM inverse are kisted. The symbol (x4) indicates that the data are acquired by means of the four-phase method (therefore there is a x4 factor respect to the single-shot acquisition).

for a single full-complex sample).

Examples of the restored images are shown in Fig. 4.7. Overall, the quality of the images provided by the two full-complex ANN methods is better better that that of the images obtained through the TM inversion and the single-shot ANN approaches. Once again, the best results are achieved by the full-complex network with the larger training dataset (5000 samples for each one of the MNIST, fashion-MNIST and random pattern datasets). In particular, this network

seems to have removed the noisy parts that appear on the image reconstructions of the TM inversion approach and on the network trained with random patterns, and that could indicate the presence of reflections or possible interferences with other orders produced by the Lee hologram technique in the Fourier plane. To allow a fairer comparison, these parts have been not taken in considerations when calculating the Pearson correlation coefficient for Fig. 4.6.
Another important aspect is generalisation. As has been discussed, the single-shot method presents an overall better performance respect to the TM inversion approach, having at its disposal the same amount of experimental measurements. However, when reconstructing patterns from different databases, with respect to the ones used for the training process, such as the "walking man" and the "running horse" (Fig. 4.7), the single-shot method achieves poor results. Therefore, the application of the single-shot ANN must remain restricted to images that are similar to the ones used for training the network, whenever considering a small training dataset. On the other hand, as already demonstrated in Chapter 3, this restrictions could be avoided by training a network with a larger amount of random patterns (20000 for the case reported in Chapter 3).

4.4.1 Singular value decomposition analysis

Within the context of transmission of information through complex media, since the wave propagation is generally too complicated to be described by a theoretical model, the verification of the validity of an empiric model is typically based on the evaluation of its experimental performances and on the analysis of its physical behaviour. Thus, after having examined the reconstruction ability of the two proposed methods, I will also investigate their transmission properties by means of the singular value decomposition. As mentioned earlier, this technique allows access to the transmission eigenchannels and their respective intensity transmission coefficients (also referred as transmittance), i.e. the singular values.

Firstly, this analysis is applied to the experimental $N \times M$ transmission matrix T, having $M = 96 \times 96$ and $N = 28 \times 28$. Following the procedure described in section 4.1.2, the transmission matrix is normalised - to remove the reference contribution - and then reduced in size by skipping one out of two elements for each row and column - in order to avoid neighbouring pixel correlations. The normalisation is performed in the Hadamard (input) base before converting the transmission matrix to the canonical base. In Fig. 4.8(a), the square of the singular values σ of T, i.e. the singular values of $T^{\dagger}T$, normalised by σ^2_{max} are reported in descending order. Whereas, the distribution of the singular values, normalised according to the equation (4.9), is shown in Fig. 4.8(b).
The plot presented in Fig. 4.8(a) is in line with the experimental results obtained in [149] [150]. In particular, a comparison between the expected behaviour of an ideal fibre and the experimental curve is shown in [149]. In this case, the authors demonstrated that the rapid decline

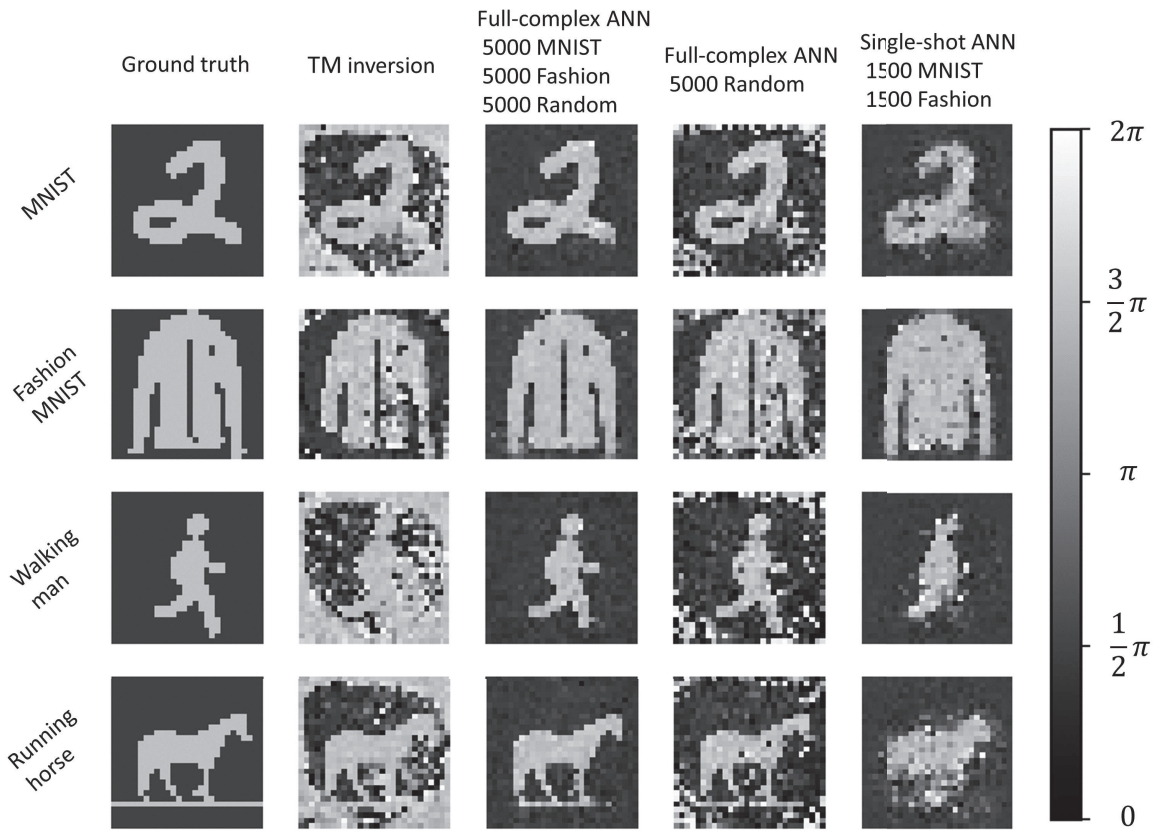

Figure 4.7: Examples of reconstructed images. The results from different approaches are reported for 4 different test sets: MNIST, Fashion-MNIST, “walking man" and “running horse”. As it can be observed, the best results are achieved by the full-complex ANN, similarly to Fig. 4.6. Moreover, the ANN trained with 5000 random patterns shows a good generalisation ability by reconstructing images from databases on which has not been trained. In contrast, this seems to be no longer true for the “single-shot" approach, as it can be noticed from the “walking man" and “running horse” results.

of the measured singular values, visible also in Fig. 4.8(a), is due to the losses experienced by the higher-order modes. On the other hand, Fig. 4.8(b) reveals a distribution that, in addition to the presence of strong isolated singular values, evidently differs from the Marcenko–Pastur law (solid blue line). Indeed, despite the possible presence of spurious correlations, the nature of the propagation of the light inside a multimode fibre differs from the propagation through a complex medium, such as an opaque material. As has been demonstrated by Cizmar’s group in [119], the multimode fibre must not be considered a chaotic medium. On the contrary, they were able to experimentally address the deviations from the ideal model for lengths up to 300 mm (they also declared to expect no main randomisation process for fibres with lengths in the order of several meters). Thus, it is reasonable to observe a single value distribution that differs from the behaviour expected by a chaotic medium. However, it must be underlined once more that a deviation from the random matrix theory simply indicates the presence of some sort of correlation. In this sense, a proper theoretical model or simulation is necessary to validate any measured singular value distribution related to a non-chaotic medium. Interestingly, Chiara-

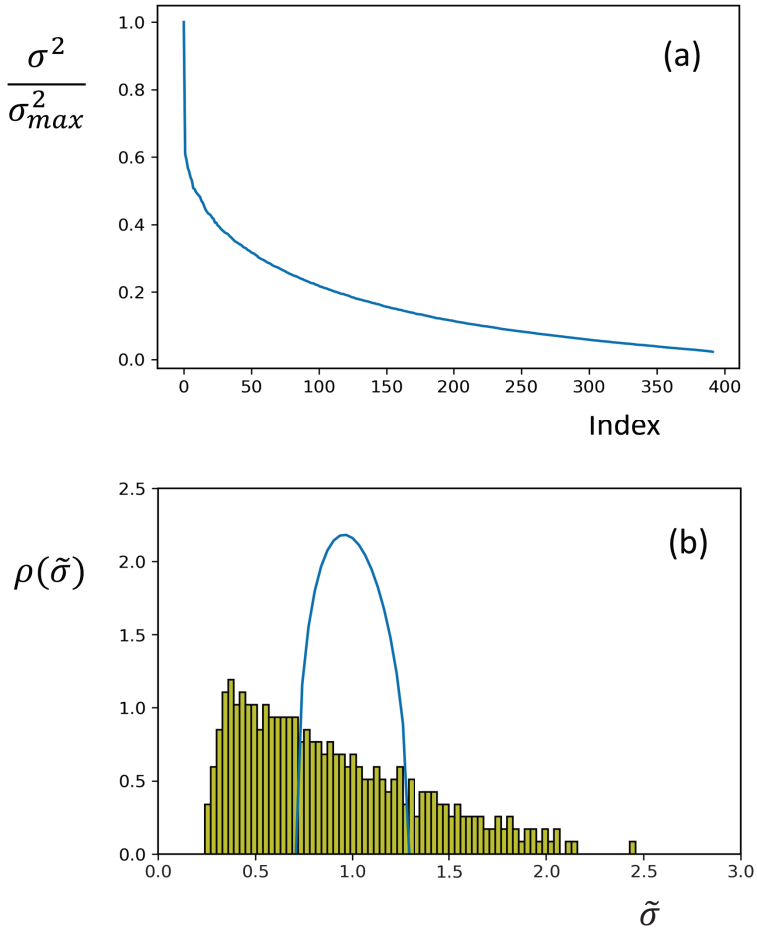

Figure 4.8: Analysis of the singular values for the experimental transmission matrix (normalised and truncated). In (a), the normalised squared singular values are reported, showing the typical drop due to the losses of higher-modes losses. In (b), the distribution of the singular values (normalised according to the equation (4.9)) indicates the presence of correlations, by differing from the RMT prediction.

wongse et al. proposed a model for the scattering process in a multimode fibre in [151]. This involves an accurate description for the mode-dependent losses and polarisation mixing that are not controlled in the setup reported in this chapter. Thus, a dedicated experiment and analysis are required (the authors validated their model through simulations and experiments on a 52 modes-fibre) in order to proceed in this direction. However, these have been judged beyond the scope of the present chapter.

The singular value decomposition analysis for the Tikhonov regularisation inverse operator K is not taken into consideration for this work. As a matter of fact, this operator is determined by the arbitrary choice of the parameter λ that, as has been discussed in section 4.1.1, deeply modifies the nature of the singular values itself. Unfortunately, in the absence of an accurate reference model, no physical significance can be ascribed to the singular value distribution obtained from this operator.

In Fig. 4.9 the singular values distributions of the $M \times N$ matrices obtained from different complex-valued ANN models (but sharing the same training dataset, i.e. 5000 random patterns), with $M = 96 \times 96$ and $N = 28 \times 28$, are reported. The idea is to study how the distribution is affected by the way in which the network is modelled and, since the analysis of the statistics of the inverse model has been excluded, build a loose comparison with the result from the transmission matrix T. However, it must be stated that the statistics of the inverse and forward operators are equivalent just for the unitary case, i.e. when the two matrices share the same singular values distribution.

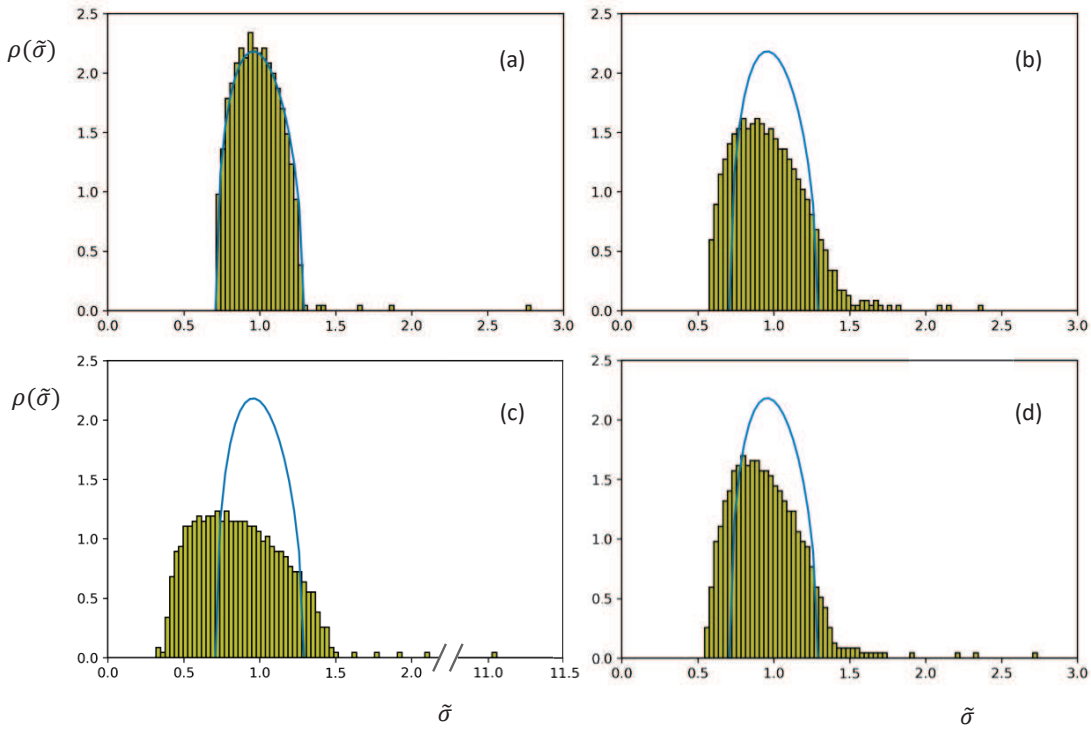

Figure 4.9: Distributions of singular values for three different ANN models. Data used for the training process are the same for (a),(b) and (c), i.e. 5000 random patterns (full-complex case). In (a), the network relies on a the l_2 regularisation; in (b), it relics on the unitary regularisation. In (c) the network is equivalent to (a) but it is initialised with a constant for both real and imaginary part (therefore, not with randomly distributed numbers). Finally, (d) shares the initialisation of (c), but uses the unitary regularisation

Firstly, the case of the l_2 norm regularisation is considered (see Fig. 4.9(a)). This norm is the same one that has been applied to all the results shown previously. It is interesting to note that in this case the histogram shows, apart for the usual isolated high singular values, a perfectly random statistics (regardless on the training dataset). Instead, by employing a unitary regularisation (see Fig. 4.9(b)), as the one introduced in section B.1, the distribution starts assuming a shape that recalls more the one in Fig. 4.9(b). Indeed, a similar result is obtained considering two networks with a non-random weight initialisation (a flat constant for both real and imaginary

part: the former using the l_2 regularisation (see Fig. 4.9(c)) and the latter relying the unitary regularisation (see Fig. 4.9(c)). In fact, the neural networks are typically initialised with random numbers, implying that the algorithm is starting from a specific suggestion, i.e. no correlations. In contrast, when a physical constraint (unitarity) or a non-random starting guess are taken into consideration the system starts to show evidence of correlations. In this sense, apart from identifying an accessible simulation method or theoretical model, an interesting future step would be to investigate if this physical constraint leads to a better representation of the experimental phenomenon: as an example, a perfect unitary transmission matrix implies a better focusing ability. Regarding the input retrieval performance, instead, no main qualitative improvement is found for this experiment.

To summarise, in the present chapter a comparison has been proposed between two different approaches that aim at restoring patterns (encoded on a coherent laser beam) that have been distorted after being propagated through a multiple scattering medium. On the one hand, the transmission matrix approach describes empirically the forward model, thus requiring a further step to provide a solution for the inverse problem. On the other hand, the complex-valued neural network approach, introduced in Chapter 3, is implemented for solving directly the inverse problem. Both the techniques identify a single complex-valued matrix - since the physical phenomenon can be considered linear - with same dimensions. Interestingly, the proposed ANN method is the only one, built so far, that allows a formulation of the problem analogous to the one defined by the transmission matrix approach. In this sense, I have investigated the possibility to compare the two obtained matrices in terms of performances and statistical behaviour, performing an experiment with a multimode fibre. Overall, the ANN models have provided better retrieval performances with respect to the transmission matrix approach. However, considering a relatively small amount of measurements, the latter tends to prevail in terms of generalisation capacity. Analysing the statistical properties of the transmission matrix, a deviation from the expected random matrix theory is observed. Moreover, the roles of regularisation and weights initialisation, for the ANN, have been studied respect to their influence on the statistics of the singular values. In a sense, the distributions obtained from the two methods can be put in comparison. However, the possibility to extract accurate physical contents remains related to the validation of these results through a theoretical model or a simulation approach (which require an accurate in-depth study). Nonetheless, the results of the reported analysis indicates the complex-valued ANN as an interesting candidate for the study of imaging in complex media and its possible applications within this context. In this sense, two interesting future directions might be related to the ability to improve focusing at the output of the media, through a unitary regularisation, or a direct comparison with the forward model obtained through the transmission matrix technique.

Chapter 5

Conclusion

The research presented in this thesis aims to implement and demonstrate novel paradigm models for the study of imaging problems in the presence of light scattering. The nature of these scenarios is such that specific techniques are required in order to address complex and possibly underconstrained challenges.
Over the past few years, more and more attention has been drawn towards artificial neural networks and their peculiar capacity to find solutions to specific problems even in the absence of an explicit mathematical model. In fact, this approach is based on inferring statistically the relationship that links the defined input and output (at least considering a supervised approach) of a certain system, by analysing a large amount of data. As a downside, it is not usually straightforward to gain insights on the internal representation found by the network and its procedure to arrive at the output.
The general purpose of this thesis is to explore the possibilities given by these algorithms, for the solution of specific imaging problems involving scattered light, by taking advantage of possible physical insights related to the problems addressed. Overall, the experimental results indicate that the implemented networks are particularly effective for the proposed challenges. Moreover, whenever possible, the possibility to extract relevant physical information from the networks has been investigated.

The first problem addressed (Chapter 2) is the simultaneous identification and location of human targets hidden behind a corner. In the context of non-line-of-sight-imaging, or sensing, multiply scattered light is employed to obtain information about an hidden target. However, 3D reconstruction typically requires long-time acquisition and calculation, large amounts of data and high temporal resolutions. In an attempt to relax these constraints, a single-pixel temporal histogram, built collecting the light backscattered from the hidden 3D target, is considered. With this poor representation, the idea is to treat identity and position as classes to be distinguished by a neural network supervised classification algorithm. A SPAD array is used to obtain simultaneously 800 histograms (IRF = 120 ps), in order to train the network. In this way, the system was able

to correctly classify the single-pixel histograms among 3 different individuals and 7 different positions, for two different cases: same clothing and different clothing. The main limitations are related to the presence of variability among the data that is not easily representable in these experimental conditions. As a consequence, different neural network architectures lead to similar results. Nonetheless, this proof of principle opens up interesting future directions, for example the use of complex-valued networks for classification-free 2D location or the implementation of 3D reconstructions from similar setup configurations. In this direction, Turpin et al. have considered the possibility to retrieve, in a lidar configuration, the 3D image of a target from a single-pixel histogram [152].

The main part of this thesis is Chapter 3. Here, the problem of transmission of images through a multimode fibre is studied. Considering real fibres, the presence of impurities, bending and temperature gradient can dramatically alter the ideal propagation of the light within these waveguides. Notwithstanding the complexity of the propagation, a complex-valued network with a single fully-connected layer is implemented in order to represent the physical linear (and complex-valued) relationship that connects the input field to the output field. As a result, even without any phase measurements, thus facing an ill-posed problem, the network is still able to provide image restoration at relatively high-resolution (96×96 pixels) and high-speed (limited just by the refreshing rate of the device used to imprint the images on a coherent laser beam). Remarkably, compared to other artificial neural network methods implemented to perform an analogous task, the complex-valued architecture shows a strong generalisation capacity - meaning that the image retrieval process is not limited to the specific classes represented during the network training. This method is fundamentally limited by the GPU RAM memory. For example, the complex-valued matrix created with 96×96 input pixel-resolution and 120×120 output pixel-resolution is fitting basically the whole 11 GB memory of the graphic card used for this experiment. With these settings, the training process takes around 2 days. On the other hand, it is shown that the network is able to address temporal degradation. This work paves the way for exciting future prospectives, such as the possibility to include other physical insights in order to address alterations due to bending or temperature gradients. Another important aspect is the fact that this network has a mathematical formulation analogous to the one of the transmission matrix, i.e. a technique commonly adopted to study and exploit the propagation through random media and multimode fibres. In this way, in Chapter 4, the possibility to compare these two techniques is investigated.

The transmission matrix method is a technique that, under the assumption of a linear and deterministic system, expresses the propagation inside a complex medium through a single matrix which can be measured empirically. This matrix relates the complex field at the input of the fibre to the complex field at the output, thus describing the forward model. In order to com-

pare this method to the complex-valued network proposed in Chapter 3, that solves directly the inverse problem, the matrix requires an inversion, which in presence of noise is not always a trivial operation. Furthermore, following the four-phase method, 4 measurements are required to obtain the full-complex field. Having at one's disposal the full-complex speckle field allows to test the complex-valued network with a complete (amplitude and phase) input. The matrices obtained from the two different methods are compared both in terms of performance and statistical behaviour, i.e. analysing their singular value distribution. As a result, the performances of the neural network models typically overcome the one of the inverted transmission matrix. However, by limiting the amount of measurements to the ones required by the transmission matrix approach, the network leads to better performances, compared to the transmission matrix, just by decreasing the capacity to generalise the result. The statistical behaviour of the measured transmission matrix differs from what expected for a random medium. Unfortunately, in the absence of a proper simulation or theoretical model it is not possible to verify if this is a sign of the fact that the multimode fibre is not a chaotic medium or if indicates the presence of other sources of correlations. The same applies to the singular value distributions observed for the neural networks, thus indicating the need for an external reliable model as a possible future direction. Interestingly, the weight initialisation and the regularisation have an impact on the final distribution. Thus, another possible future direction is to verify experimentally if a certain Physics-inspired assumption, such as applying the unitary regularisation, can lead to a better capacity to control the setup (for example improving the ability to focus at the output of the fibre).

Personally, I think that neural networks are a tool as exciting as they are bizarre. Evidently their performance capacity is considerable. On the other hand, Physics is probably one of the few fields in which a realistic interpretation for the obtained results can be proposed and verified. In this sense, to provide inputs and constraints related to the physics of a system of interest could represent a first step towards this direction, helping to gain a better control of the system, improve the performance and facilitate the comparison with an external physical model. In other words, helping to build a constructive dialogue with these algorithms.
In the end, science is made by women and men. No machine will ever spare us from the task of acknowledging the meaning of the results; in the same way as the VAR, or any technological advancement, will never ultimately substitute a football referee. This is the beautiful drama of being a scientist.

Appendix A

Neural network Python Code for identification and location of hidden targets

This code has been implemented to correctly distinguish location and identity of hidden human targets through a neural network algorithm. In particular, the reported model represents a non linear classifier that processes the input temporal histograms in parallel through a series of convolutional layers and a fully connected layer. The function, here presented, allows to use this classifier in a "leave-one-out" cross-validation scheme over 5 measurement-permutations. Keras library [23] and Tensorflow [25] has been used to built this code on Python 3.6.5.

Listing A.1: Looking-around-the-corner neural network model

```
def setup_model_default(units, X_TR, l2_value, useConv, onlyClass,
    Ntr, ep_value, y_TRonehot, X_TE, y_TEonehot, y_TRLonehot,
    y_TELonehot):

# X_TR and X_TE are the training and testing temporal histograms
    respectively.
# y_TRonehot and y_TEonehot are the person labels for each histogram
    respectively.
# y_TRLonehot and y_TELonehot are the location labels for each
    histogram respectively.
# l2_value is the l2 weight decay constant.
# Ntr is the number of training inputs.
# ep_value is the number of training epochs.

     print('Model: Default')
```

```
regConst   = l2_value
input_vec  = Input(shape=(X_TR.shape[1],))
input_vec2 = ExpandInput()(input_vec)

xd = Dense(units, kernel_regularizer=l2(regConst),
   input_shape=(X_TR.shape[1],))(input_vec)
xd = BatchNormalization()(xd)
x  = Dropout(0.5)(xd)
xd = Activation('relu')(xd)

if useConv:
    x = Convolution1D(units, 10, padding='same',
       kernel_regularizer=l2(regConst), input_shape =
       (X_TR.shape[0],X_TR.shape[1],1))(input_vec2)
    x = BatchNormalization()(x)
    x = Activation('relu')(x)
    x = MaxPooling1D()(x)

    x = Convolution1D(units, 10, padding='same',
       kernel_regularizer=l2(regConst))(x)
    x = BatchNormalization()(x)
    x = Activation('relu')(x)
    x = MaxPooling1D()(x)

    x = Convolution1D(units, 5,
       padding='same',activation='relu',
       kernel_regularizer=l2(regConst))(x)
    x = BatchNormalization()(x)
    x = Activation('relu')(x)
    x = MaxPooling1D()(x)

    x = Convolution1D(units, 5,
       padding='same',activation='relu',
       kernel_regularizer=l2(regConst))(x)
    x = BatchNormalization()(x)
    x = Activation('relu')(x)

    x = Flatten()(x)
    x = layers.concatenate([x, xd], axis=1)
else:
    x = xd
```

```
x = Dense(units, kernel_regularizer=l2(regConst))(x)
x = Dropout(0.3)(x)
x = BatchNormalization()(x)
x = Activation('relu')(x)

x = Dense(units, kernel_regularizer=l2(regConst))(x)
x = Dropout(0.2)(x)
x = BatchNormalization()(x)
x1 = Activation('relu')(x)

xC = Dense(3,activation='softmax',
   kernel_regularizer=l2(regConst))(x1)
xL = Dense(7,activation='softmax',
   kernel_regularizer=l2(regConst))(x1)

if onlyClass:
   model = Model(input_vec, outputs=xC)
else:
   model = Model(input_vec, outputs=[xC,xL])

# randomly change order of inputs.
order= np.random.permutation(Ntr)
# train the model using SGD + momentum.
batch_size    = 32
nb_epoch      = ep_value
learning_rate = 0.001
decay_rate    = 0.0
sgd = SGD(lr=learning_rate, decay=decay_rate , momentum=0.9,
   nesterov=True)
if onlyClass:
   model.compile(loss='categorical_crossentropy',
      metrics=['accuracy'], optimizer=sgd)
   hist_conv = model.fit(x=X_TR[order,:],
                         y=y_TRonehot[order,:],
                         batch_size=batch_size, epochs=nb_epoch,
                         validation_data=(X_TE, y_TEonehot),
                            verbose=2)
else:
   model.compile(loss=['categorical_crossentropy',
      'categorical_crossentropy'], metrics=['accuracy'],
```

```
        optimizer=sgd)
    hist_conv = model.fit(x=X_TR[order,:],
                          y=[y_TRonehot[order,:],y_TRLonehot[order,:]],
                          batch_size=batch_size, epochs=nb_epoch,
                          validation_data=(X_TE,
                              [y_TEonehot,y_TELonehot]), verbose=2)

return model, hist_conv
```

Appendix B

Intensity-to-intensity approach for restoring images propagated through a multimode fibre

In Chapter 3, the problem of imaging through a multimode fibre has been introduced. In order to address this problem, a neural network with a single complex-valued fully-connected layer approach has been proposed. This choice reflects the physical nature of the problem that can be described by a single complex-valued matrix that connects an input base to an output base. In other words, the system is linear and puts in relationship the input complex field (amplitude and phase) to the output one. However, in a first instance, the attempt was to find the transformation that could map directly input intensity into output intensity. In this appendix, the approach proposed to address this particular case will be discussed, along with the results obtained. Furthermore, the regularisations employed for the cost function ζ will be introduced.

The intensity-to-intensity relationship introduces a non-linearity that is here addressed by adopting a non linear activation function, after having extracted the amplitude $|Wx|$ (referring to notation adopted in section 3.4), and an Hadamard multiplication layer. In other words, in this case the only assumption considered is that the input-output intensity relationship implies a com-

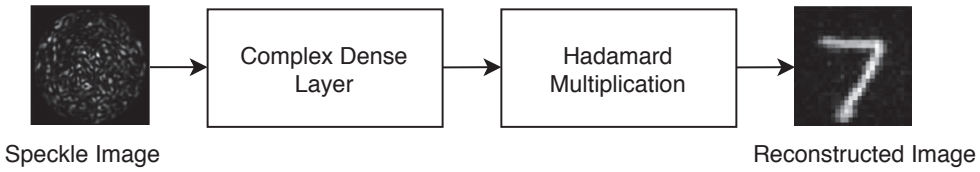

Figure B.1: Schematic overview of the intensity input and intensity output model. The complex-valued fully-connected layer is followed by PReLU activation function, applied to the amplitude of the predicted image and an Hadamard multiplication layer that serves to represent a possible not uniformity of the beam and correct scaling issues.

plex matrix multiplication. The non-linear activation function is named Parametric Rectified Linear Unit (PReLU) [153] - basically a continuous function formed by the conjunction of two half-lines in 0 (where it is not differentiable) -. Instead, the Hadamard layer is applied to the predicted image, as showed in Fig. B.1, in order to learn eventual spatial intensity differences and correcting scaling issues.

B.1 Weight regularisations

Two types of regularisation are considered: unitary regularisation and the l_2 norm. The former, unitary regularisation, is motivated by the fact that the modes of the fibre are expected to be orthogonal. A similar approach was introduced by Brock et al. in [154] , where the authors attempted to find an orthogonal representation for a layer of interest. In this case, the regularisation is extended to the complex domain implementing:

$$\mathscr{L}_{unitary}(W) = \|WW^{\dagger} - I\|_1 \text{ for } W \in \mathbb{C}^{m\times n}, \tag{B.1}$$

with $W^{\dagger}$ representing the conjugate transpose of W. In this way, the matrix W is pushed toward semi-unitary.

The latter has been already introduced in Chapter 1. However, in this case the network is complex-valued. Thus, the regularisation should be applied to both amplitude and phase:

$$\mathscr{L}(W, \alpha_r, \alpha_\phi) = \sum_i \alpha_r |W_i|^2 + \alpha_\phi \underline{/W_i} \tag{B.2}$$

where the operator $\underline{/\cdot}$ indicates the phase extraction, W_i the i-th element of the matrix W, α_r and α_ϕ the parameters for, respectively, amplitude and phase. However, since no phase in this experiment is used, α_ϕ is set to zero. As a consequence, the remaining term is equivalent to a l_2 norm applied separately to the real ($\Re\{\cdot\}$) and imaginary ($\Im\{\cdot\}$) channels, with which W is represented in the actual code:

$$\mathscr{L}(W, \alpha_r) = \sum_i \alpha_r (\Re\{W_i\}^2 + \Im\{W_i\}^2), \tag{B.3}$$

which makes it particularly easy to implement.

B.2 Results

The results relative to the intensity-to-intensity model are here reported. In this case, the training dataset is composed by patterns taken from the databases already shown in Fig. 3.4: handwritten digits, cloths from the Fashion-MNIST database, Hadamard patterns and random pat-

terns. In particular, the amount of pattern used for each dataset are, respectively: 12000, 12000, 784 and 20000. The size of this patterns is 28×28 pixels (which defines the number of elements of the Hadamard base). Instead, the size of the speckles is 112×112. Firstly, the performances of the model for three different cases are compared: without any regularisation, with l_2 regularisation ($\lambda = 0.03$) and with unitary regularisation. The fibre used for this experiment is the 1m-long. The neural network is tested on 3000 MNIST and 3000 Fashion-MNIST patterns. As it can be noticed from Fig. B.2, the reconstructions are in good accordance with the ground truth patterns. However, for this particular experiment there is not a strong evidence of a better performance due to a particular regularisation employed.

(a) No regularisation (b) l_2 weight regularisation (c) Unitary regularisation

Figure B.2: Complex-valued neural network reconstruction of unseen patterns for three different type of regularisation: (a) no regularisation, (b) l_2 norm and (c) unitary regularisation. This experiment was conducted on the 1m-long fibre. The size of the input speckles is 112×112. The resolution of the images in the figure is 28×28.

This is confirmed also considering the MSE metric applied to the test images (comparing retrieved and ground truth images). In Table B.1, a comparison between three different regularisation options, in terms of MSE, is reported. Furthermore, the complex-valued network is compared to a common real-valued fully-connected layer (therefore just one channel, i.e. half of the parameter to be tuned) with l_2 regularisation and without. As it can be observed, once again the different regularisations lead to similar results. On the other hand, complex-valued networks

Table B.1: Model comparison for real-valued and complex-valued network with different regularisation. Speckle resolution is 112×112, 1m-long fibre. MSE is used as metrix on the whole test dataset.

Model		
Name	Regularisation	MSE
Real-valued	l_2 weight regularisation	1034.62
Real-valued	None	1025.96
Complex	Unitary regularisation	989.28
Complex	l_2 weight regularisation	962.33
Complex	None	960.31

present an overall better performance.

In Fig. B.3, the complex-valued network, after undergoing the same training process, using the l_2 regularisation, is tested on a sample taken from the Muybridge collection, therefore on an image not belonging to any of the classes composing the training datasets. In this way, a comparison is proposed between the results for a 1m-long fibre and for a 10-m long fibre. As it can be noticed, the results are qualitatively similar even if the size and the number of the speckles vary between the two different cases. This is a signature of a deeper mode-coupling phenomena. In fact, the longer the fibre the greater the possibility for bending, imperfections or temperature gradient to locally modify the index of refraction. Thus, the role of the speckle resolution will be investigated in the next section.

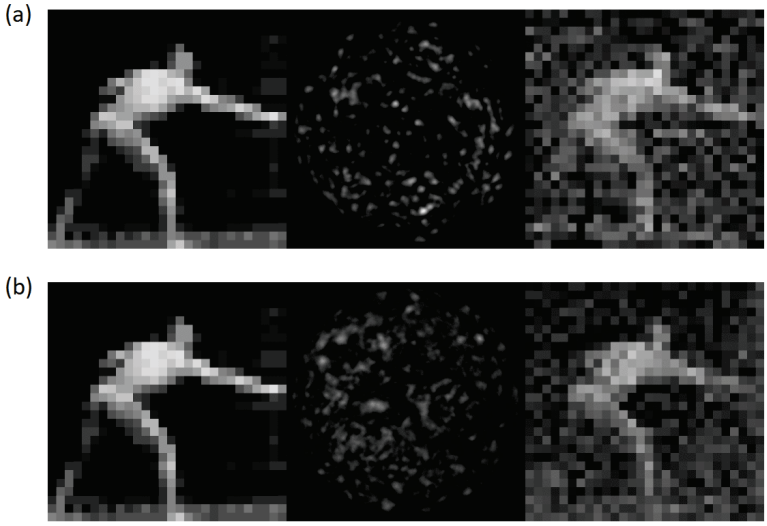

Figure B.3: Muybridge's sample reconstruction for fibres with length equal to 1 m (a) and 10 m (b). On the left column, it is reported the ground truth image (28×28 pixels); on the central column, the speckle pattern (112×112); on the the right column, the predicted image (28×28 pixels). The different speckle complexity between (a) and (b) arises from the propagation through a longer fibre, therefore an enhancement of modes-coupling due to imperfections or external agents.

B.2.1 Impact of speckle resolution

In the same experimental condition, the impact of the resolution of the speckles, i.e. the input of the network, on the final reconstruction is evaluated. A 28×28 ground truth image to be restored is fixed. Then, the speckle size is varied from 14×14 to 224×224. As it can be observed, the network is still able to provide an approximate reconstruction even starting from a 14 speckle pixel-resolution. Clearly, by increasing this resolution the predicted image improves. On the other hand, the step between 112×112 and 224×224 does not show any major improvement. Evidently, when the speckles are well resolved there is no need to consider a larger resolution.

This information is particularly relevant to identify an optimal minimum resolution that allows to save precious RAM memory. As a matter of fact, the memory requirement scales with $O(R_i^2 R_o^2)$, where R_i and R_o indicate, respectively, the pixel resolution of the input and output image.

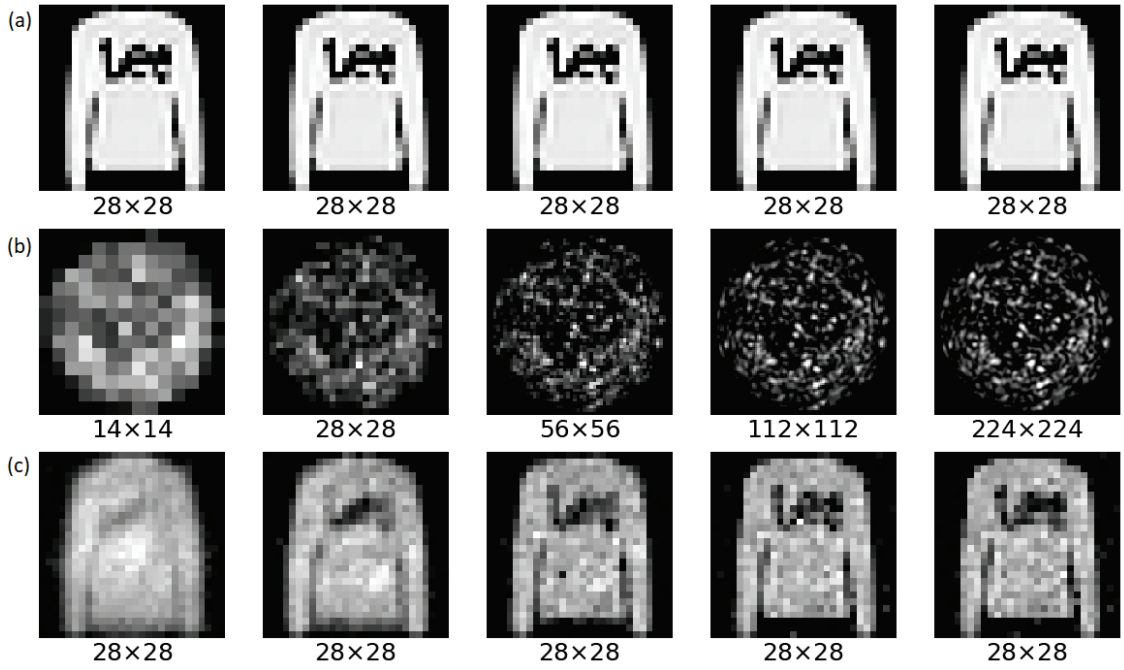

Figure B.4: Comparison between reconstructed images for different speckle resolution. (a) Ground truth image; (b) speckle pattern; (c) predicted image. The speckle side-resolution is varied over the values 14, 28, 56, 112, and 224. As it can be observed, the more the speckle are resolved, the better the reconstruction. However, no major improvement is shown when the speckles are already well resolved.

Appendix C

Complex-valued neural network Python code

In this appendix, the code used to implement the complex-valued neural network is reported. This code can be downloaded following the The DOI link [133]. The Python version employed is the "3.6.5". Moreover, the Keras library [23] and the Tensorflow framework [25] are necessary to run this code. Finally, some of the results shown in Chapter 3 can be reproduced by means of the data provided at the same DOI link (for the 1 m-long fibre).

C.1 Defining the model

The present approach relies on a single complex-valued fully-connected layer. This is implemented as a customer layer for Keras, under the name of "ComplexDense". The complex weights format is complex64, stored on two separated matrices one for the real part and one for the imaginary part, each one with format "float32". The weight initialisation follows a uniform random distribution between ± 0.002. The hyperparamenters includes the input and output resolution, the λ parameter for the l_2 norm, the number of epochs and the batch size. The l_2 norm is implemented as described in Appendix B.1.

Listing C.1: Hyperparameters and model specification

```
speckle_dim = 120
out_dim = 92
lamb = 0.03
ep = 850
bs = 16

inp = Input(shape=(speckle_dim**2,2))
comp = ComplexDense(out_dim**2, use_bias=False,
```

```
                    kernel_initializer=RandomUniform(-.002, .002),
                    kernel_regularizer=regularizers.l2(lamb))(inp)
amp = Amplitude()(comp)

model = Model(inputs=inp, outputs=amp)
```

Here the "ComplexDense" class is defined. Essentially, it performs a simple but complex-valued matrix multiplication.

Listing C.2: Custom ComplexDense Layer

```
class ComplexDense(Layer):

def __init__(self, output_dim,
             activation=None,
             use_bias=True,
                 kernel_initializer='glorot_uniform',
             bias_initializer='zeros',
             kernel_regularizer=None,
             **kwargs):
        super(ComplexDense, self).__init__(**kwargs)
    self.output_dim = output_dim
    self.activation = activations.get(activation)
    self.use_bias = use_bias
    self.kernel_initializer = initializers.get(kernel_initializer)
    self.bias_initializer = initializers.get(bias_initializer)
    self.kernel_regularizer = regularizers.get(kernel_regularizer)


def build(self, input_shape):
    self.kernel = self.add_weight(name='kernel',
                                  shape=(input_shape[1],
                                     self.output_dim, 2),
                                  initializer=self.kernel_initializer,
                                  regularizer=self.kernel_regularizer,
                                  trainable=True)

    if self.use_bias:
        self.bias = self.add_weight(name='bias',
                                    shape=(self.output_dim, 2),
                                    initializer=self.bias_initializer,
                                    trainable=True)
```

```
        else:
            self.bias = None
        super(ComplexDense, self).build(input_shape)

    def call(self, X):
        # True Complex Multiplication (by channel combination)
        complex_X = channels_to_complex(X)
        complex_W = channels_to_complex(self.kernel)

        complex_res = complex_X @ complex_W

        if self.use_bias:
            complex_b = channels_to_complex(self.bias)
                 = K.bias_add(complex_res, complex_b)

        output = complex_to_channels(complex_res)

        if self.activation is not None:
            output = self.activation(output)

        return output

    def compute_output_shape(self, input_shape):
        return (input_shape[0], self.output_dim, 2)

    def get_config(self):
        config = {'output_dim': self.output_dim,
                  'use_bias': self.use_bias,
                  'kernel_initializer':
                     initializers.serialize(self.kernel_initializer),
                  'bias_initializer':
                     initializers.serialize(self.bias_initializer),
                  'kernel_regularizer':
                     regularizers.serialize(self.kernel_regularizer)
                   }
        base_config = super(ComplexDense, self).get_config()
        return dict(list(base_config.items()) + list(config.items()))
```

C.2 Parameter optimisation

Stochastic gradient descent and the MSE metric are employed for the cost function. Thus, the model is now ready to be fitted. `x_train`, `y_train`, `x_val`, `y_val`, `x_test` and `x_test` refer to the training, validation and test datasets. In particular, `x_` indicates the speckle images and `y_` the ground truth images. The training datasets is composed by $45,000$ samples, the validation one by $5,000$.

Listing C.3: Compile and optimise parameters

```
model.compile(optimizer=SGD(lr=1e-5), loss='mse', metrics=['mse'])
model_chk = ModelCheckpoint(weights_filepath, monitor='mse',
   verbose=0,
                            save_best_only=False,
                            save_weights_only=False, mode='auto',
                               period=1)
reduce_lr = ReduceLROnPlateau(monitor='loss', factor=0.1, patience=2,
             min_lr=lr/1e3, verbose=1,)
early_stop = EarlyStopping(monitor='loss', min_delta=0.0001,
   patience=8)

model.fit(x_train, y_train, validation_data = (x_val, y_val),
          epochs = ep, batch_size = bs,
          callbacks = [model_chk, reduce_lr, early_stop], shuffle =
             True)
```

The number of epochs (iterations) is 850. The training process takes around 2 days on a GPU Nvidia GeForce GTX 1080 Ti. Once the network is trained a reconstructed image (or more) can be obtained using:

```
pred_test = model.predict(x_test)**2
```

In this case, the amplitude is predicted, therefore the square operator needs to be applied to restore a ground truth image (encoded as an intensity image).

Bibliography

[1] Piergiorgio Caramazza, Alessandro Boccolini, Daniel Buschek, Matthias Hullin, Catherine F Higham, Robert Henderson, Roderick Murray-Smith, and Daniele Faccio. Neural network identification of people hidden from view with a single-pixel, single-photon detector. *Scientific Reports*, 8(1):11945, 2018.

[2] Piergiorgio Caramazza, Oisín Moran, Roderick Murray-Smith, and Daniele Faccio. Transmission of natural scene images through a multimode fibre. *Nature Communications*, 10(1):2029, 2019.

[3] Oisín Moran, Piergiorgio Caramazza, Daniele Faccio, and Roderick Murray-Smith. Deep, complex, invertible networks for inversion of transmission effects in multimode optical fibres. In *Advances in Neural Information Processing Systems*, pages 3280–3291, 2018.

[4] Piergiorgio Caramazza, Roderick Murray-Smith, and Daniele Faccio. Neural network classification for intensity imaging through multimode optical fibres. In *Computational Optical Sensing and Imaging*, pages CW3B–4. Optical Society of America, 2018.

[5] Andrew Ng: Why AI is the new electricity. https://news.stanford.edu/thedish/2017/03/14/andrew-ng-why-ai-is-the-new-electricity/.

[6] What's the difference between AI, Machine Learning, and Deep Learning? https://https://blogs.oracle.com/bigdata/difference-ai-machine-learning-deep-learning.

[7] Peter Dayan and Yael Niv. Reinforcement learning: the good, the bad and the ugly. *Current opinion in neurobiology*, 18(2):185–196, 2008.

[8] How much data do we create every day? The mind-blowing stats everyone should read. https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#6f5c0e3260ba.

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

[10] 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*, pages 248–255. Ieee, 2009.

[11] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[12] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[13] Marvin Minsky and Seymour A Papert. *Perceptrons: An introduction to computational geometry*. MIT press, 2017.

[14] Bernard Widrow. Adaptive "adaline" neuron using chemical "memistors". *Stanford Electron. Labs*, 1553-2, 1960.

[15] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1, 1988.

[16] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[17] Sebastian Raschka. *Python machine learning*. Packt Publishing Ltd, 2015.

[18] David H Hubel and Torsten N Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1):215–243, 1968.

[19] M.C.J. Peemen. *Improving the efficiency of deep convolutional networks*. PhD thesis, Department of Electrical Engineering, 10 2017. Proefschrift.

[20] Patrice Y Simard, David Steinkraus, John C Platt, et al. Best practices for convolutional neural networks applied to visual document analysis. In *Icdar*, volume 3, 2003.

[21] Tom Schaul, Justin Bayer, Daan Wierstra, Yi Sun, Martin Felder, Frank Sehnke, Thomas Rückstieß, and Jürgen Schmidhuber. PyBrain. *Journal of Machine Learning Research*, 11:743–746, 2010.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python . *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[23] François Chollet et al. Keras. https://keras.io, 2015.

[24] Rami Al-Rfou, Guillaume Alain, Amjad Almahairi, Christof Angermueller, Dzmitry Bahdanau, Nicolas Ballas, Frédéric Bastien, Justin Bayer, Anatoly Belikov, Alexander Belopolsky, et al. Theano: A python framework for fast computation of mathematical expressions. *arXiv preprint arXiv:1605.02688*, 2016.

[25] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[26] Bruce Denby. Neural networks in high energy physics: a ten year perspective. *Computer Physics Communications*, 119(2-3):219–231, 1999.

[27] Rudolf Frühwirth. Selection of optimal subsets of tracks with a feed-back neural network. *Computer Physics Communications*, 78(1-2):23–28, 1993.

[28] Paulo Abreu, W Adam, T Adye, E Agasi, GD Alekseev, A Algeri, P Allen, S Almehed, SJ Alvsvaag, U Amaldi, et al. Classification of the hadronic decays of the z0 into b and c quark pairs using a neural network. *Physics Letters B*, 295(3-4):383–395, 1992.

[29] S Abachi, B Abbott, M Abolins, BS Acharya, I Adam, DL Adams, M Adams, S Ahn, H Aihara, GA Alves, et al. Direct measurement of the top quark mass. *Physical Review Letters*, 79(7):1197, 1997.

[30] Dan Guest, Kyle Cranmer, and Daniel Whiteson. Deep learning and its application to LHC physics. *Annual Review of Nuclear and Particle Science*, 68:161–181, 2018.

[31] Babak Alipanahi, Andrew Delong, Matthew T Weirauch, and Brendan J Frey. Predicting the sequence specificities of dna-and rna-binding proteins by deep learning. *Nature Biotechnology*, 33(8):831, 2015.

[32] Hui Y Xiong, Babak Alipanahi, Leo J Lee, Hannes Bretschneider, Daniele Merico, Ryan KC Yuen, Yimin Hua, Serge Gueroussov, Hamed S Najafabadi, Timothy R Hughes, et al. The human splicing code reveals new insights into the genetic determinants of disease. *Science*, 347(6218):1254806, 2015.

[33] Feng Ning, Damien Delhomme, Yann LeCun, Fabio Piano, Léon Bottou, and Paolo Emilio Barbano. Toward automatic phenotyping of developing embryos from videos. *IEEE Transactions on Image Processing*, 14:1360–1371, 2005.

[34] Viren Jain, Joseph F Murray, Fabian Roth, Srinivas Turaga, Valentin Zhigulin, Kevin L Briggman, Moritz N Helmstaedter, Winfried Denk, and H Sebastian Seung. Supervised learning of image restoration with convolutional networks. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007.

[35] Dan Ciresan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in Neural Information Processing Systems*, pages 2843–2851, 2012.

[36] Kevin Schawinski, M Dennis Turp, and Ce Zhang. Exploring galaxy evolution with generative models. *arXiv preprint arXiv:1812.01114*, 2018.

[37] Ham Yoo-Geun, Kim Jeong-Hwan, and Luo Jing-Jia. Deep learning for multi-year enso forecasts. *Nature*, 573:pages568–572, 2019.

[38] Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9(1):4950, 2018.

[39] Kathleen Champion, Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116(45):22445–22451, 2019.

[40] Raban Iten, Tony Metger, Henrik Wilming, Lídia Del Rio, and Renato Renner. Discovering physical concepts with neural networks. *Physical Review Letters*, 124(1):010508, 2020.

[41] Joseph N Mait, Gary W Euliss, and Ravindra A Athale. Computational imaging. *Advances in Optics and Photonics*, 10(2):409–483, 2018.

[42] George Barbastathis, Aydogan Ozcan, and Guohai Situ. On the use of deep learning for computational imaging. *Optica*, 6(8):921–943, 2019.

[43] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2015.

[44] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 769–777, 2015.

[45] Alexandre Goy, Kwabena Arthur, Shuai Li, and George Barbastathis. Low photon count phase retrieval using deep learning. *Physical Review Letters*, 121(24):243902, 2018.

[46] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018.

[47] Yair Rivenson, Yibo Zhang, Harun Günaydın, Da Teng, and Aydogan Ozcan. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light: Science & Applications*, 7(2):17141, 2018.

[48] Ayan Sinha, Justin Lee, Shuai Li, and George Barbastathis. Lensless computational imaging through deep learning. *Optica*, 4(9):1117–1125, 2017.

[49] Ryoichi Horisaki, Ryosuke Takagi, and Jun Tanida. Learning-based imaging through scattering media. *Optics Express*, 24(13):13738–13743, 2016.

[50] Meng Lyu, Hao Wang, Guowei Li, and Guohai Situ. Exploit imaging through opaque wall via deep learning. *arXiv preprint arXiv:1708.07881*, 2017.

[51] Guy Satat, Matthew Tancik, Otkrist Gupta, Barmak Heshmat, and Ramesh Raskar. Object classification through scattering media with deep learning on time resolved measurement. *Optics Express*, 25(15):17466–17479, 2017.

[52] Yunqiang Yang and Aly E Fathy. See-through-wall imaging using ultra wideband short-pulse radar system. In *2005 IEEE Antennas and Propagation Society International Symposium*, volume 3, pages 334–337. IEEE, 2005.

[53] Stanley E Borek. An overview of through the wall surveillance for homeland security. In *34th Applied Imagery and Pattern Recognition Workshop (AIPR'05)*, pages 6–pp. IEEE, 2005.

[54] Fauzia Ahmad, Yimin Zhang, and Moeness G Amin. Three-dimensional wideband beamforming for imaging through a single wall. *IEEE Geoscience and Remote Sensing Letters*, 5(2):176–179, 2008.

[55] Daniele Faccio and Andreas Velten. A trillion frames per second: the techniques and applications of light-in-flight photography. *Reports on Progress in Physics*, 81(10):105901, 2018.

[56] JA Giordmaine, PM Rentzepis, SL Shapiro, and KW Wecht. Two-photon excitation of fluorescence by picosecond light pulses. *Applied Physics Letters*, 11(7):216–218, 1967.

[57] MA Duguay and AT Mattick. Ultrahigh speed photography of picosecond light pulses and echoes. *Applied Optics*, 10(9):2162–2170, 1971.

[58] Alessandra Andreoni, Maria Bondani, Marco AC Potenza, and Yury N Denisyuk. Holographic properties of the second-harmonic cross correlation of object and reference optical wave fields. *JOSA B*, 17(6):966–972, 2000.

[59] Yan Han, Ozdal Boyraz, and Bahram Jalali. Tera-sample per second real-time waveform digitizer. *Applied Physics Letters*, 87(24):241116, 2005.

[60] Keisuke Goda, Kevin K Tsia, and Bahram Jalali. Amplified dispersive fourier-transform imaging for ultrafast displacement sensing and barcode reading. *Applied Physics Letters*, 93(13):131109, 2008.

[61] K Goda, KK Tsia, and B Jalali. Serial time-encoded amplified imaging for real-time observation of fast dynamic phenomena. *Nature*, 458(7242):1145, 2009.

[62] Keiichi Nakagawa, Atsushi Iwasaki, Yu Oishi, Ryoichi Horisaki, Akira Tsukamoto, Aoi Nakamura, Kenichi Hirosawa, Hongen Liao, Takashi Ushida, Keisuke Goda, et al. Sequentially timed all-optical mapping photography (stamp). *Nature Photonics*, 8(9):695, 2014.

[63] Andor. The technology behind iccds. http://www.andor.com/learning-academy/intensified-ccd-cameras-the-technologybehind-iccds.

[64] G Häusler, JM Herrmann, R Kummer, and MW Lindner. Observation of light propagation in volume scatterers with 10 11-fold slow motion. *Optics Letters*, 21(14):1087–1089, 1996.

[65] Nils Abramson. Light-in-flight recording by holography. *Optics Letters*, 3(4):121–123, 1978.

[66] Rudolf Schwarte, Zhanping Xu, Horst-Guenther Heinol, Joachim Olk, Ruediger Klein, Bernd Buxbaum, Helmut Fischer, and Juergen Schulte. New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device (pmd). In *Sensors, Sensor Systems, and Sensor Data Processing*, volume 3100, pages 245–253. International Society for Optics and Photonics, 1997.

[67] Achuta Kadambi, Refael Whyte, Ayush Bhandari, Lee Streeter, Christopher Barsi, Adrian Dorrington, and Ramesh Raskar. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Transactions on Graphics (ToG)*, 32(6):167, 2013.

[68] J Itatani, F Quéré, Gennady L Yudin, M Yu Ivanov, Ferenc Krausz, and Paul B Corkum. Attosecond streak camera. *Physical Review Letters*, 88(17):173903, 2002.

[69] Liang Gao, Jinyang Liang, Chiye Li, and Lihong V Wang. Single-shot compressed ultrafast photography at one hundred billion frames per second. *Nature*, 516(7529):74, 2014.

[70] Silvano Donati and Tiziana Tambosso. Single-photon detectors: from traditional pmt to solid-state SPAD-based technology. *IEEE Journal of Selected Topics in Quantum Electronics*, 20(6):204–211, 2014.

[71] John A Rouse. Three dimensional computer modeling of electron optical systems. In *Advances in Imaging and Electron Physics*, volume 208, pages 129–241. Elsevier, 2018.

[72] Chandra M Natarajan, Michael G Tanner, and Robert H Hadfield. Superconducting nanowire single-photon detectors: physics and applications. *Superconductor Science and Technology*, 25(6):063001, 2012.

[73] Gerald Buller and Andrew Wallace. Ranging and three-dimensional imaging using time-correlated single-photon counting and point-by-point acquisition. *IEEE Journal of Selected Topics in Quantum Electronics*, 13(4):1006–1015, 2007.

[74] MA Duguay and J-W Hansen. An ultrafast light gate. *Applied physics letters*, 15(6):192–194, 1969.

[75] Nils H Abramson. *Light in flight or the holodiagram: the columbi egg of optics*. SPIE Press, 1996.

[76] Nils H Abramson. "light-in-flight" - Abramson. https://www.youtube.com/watch?v=MepaY3mzwS4.

[77] Daniele Faccio. Can we "freeze" light? Prof. Daniele Faccio, TEDxHeriotWattUniversity. https://www.youtube.com/watch?v=icu2pA47FH4.

[78] Hibberd L. "first light", museum exhibition at the musee des arts et metiers, paris - hibberd l. https://lilyhibberd.com/First_Light.html.

[79] MIT. Visualizing video at the speed of light - one trillion frames per second. https://www.youtube.com/watch?v=EtsXgODHMWk.

[80] Andreas Velten, Everett Lawson, Andrew Bardagjy, Moungi Bawendi, and Ramesh Raskar. Slow art with a trillion frames per second camera. In *ACM SIGGRAPH 2011 Talks*, page 44. ACM, 2011.

[81] Genevieve Gariepy, Nikola Krstajić, Robert Henderson, Chunyong Li, Robert R Thomson, Gerald S Buller, Barmak Heshmat, Ramesh Raskar, Jonathan Leach, and Daniele Faccio. Single-photon sensitive light-in-fight imaging. *Nature Communications*, 6:6021, 2015.

[82] John M Dudley, Goëry Genty, and Stéphane Coen. Supercontinuum generation in photonic crystal fiber. *Reviews of Modern Physics*, 78(4):1135, 2006.

[83] Ryan Warburton, Constantin Aniculaesei, Matteo Clerici, Yoann Altmann, Genevieve Gariepy, Richard McCracken, Derryck Reid, Steve McLaughlin, Marco Petrovich, John Hayes, et al. Observation of laser pulse propagation in optical fibers with a SPAD camera. *Scientific Reports*, 7:43302, 2017.

[84] Robert W Boyd. Slow and fast light: fundamentals and applications. Technical Report 18-19, 2009.

[85] Kali Wilson, Bethany Little, Genevieve Gariepy, Robert Henderson, John Howell, and Daniele Faccio. Slow light in flight imaging. *Physical Review A*, 95(2):023830, 2017.

[86] Matteo Clerici, Gabriel C Spalding, Ryan Warburton, Ashley Lyons, Constantin Aniculaesei, Joseph M Richards, Jonathan Leach, Robert Henderson, and Daniele Faccio. Observation of image pair creation and annihilation from superluminal scattering sources. *Science Advances*, 2(4):e1501691, 2016.

[87] Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Moungi Bawendi, Diego Gutierrez, and Ramesh Raskar. Femto-photography: capturing and visualizing the propagation of light. *ACM Transactions on Graphics (ToG)*, 32(4):44, 2013.

[88] Genevieve Gariepy, Francesco Tonolini, Robert Henderson, Jonathan Leach, and Daniele Faccio. Detection and tracking of moving objects hidden from view. *Nature Photonics*, 10(1):23, 2016.

[89] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 3:745, 2012.

[90] Nikhil Naik, Shuang Zhao, Andreas Velten, Ramesh Raskar, and Kavita Bala. Single view reflectance capture using multiplexed scattering and time-of-flight imaging. In *ACM Transactions on Graphics (ToG)*, volume 30, page 171. ACM, 2011.

[91] Felix Heide, Lei Xiao, Wolfgang Heidrich, and Matthias B Hullin. Diffuse mirrors: 3d reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors.

In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3222–3229, 2014.

[92] Mauro Buttafava, Jessica Zeman, Alberto Tosi, Kevin Eliceiri, and Andreas Velten. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Optics Express*, 23(16):20997–21011, 2015.

[93] Matthew O'Toole, David B Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555(7696):338, 2018.

[94] Rohit Pandharkar, Andreas Velten, Andrew Bardagjy, Everett Lawson, Moungi Bawendi, and Ramesh Raskar. Estimating motion and size of moving non-line-of-sight objects in cluttered environments. In *CVPR 2011*, pages 265–272. IEEE, 2011.

[95] Susan Chan, Ryan E Warburton, Genevieve Gariepy, Jonathan Leach, and Daniele Faccio. Non-line-of-sight tracking of people at long range. *Optics Express*, 25(9):10109–10117, 2017.

[96] Jonathan Klein, Christoph Peters, Jaime Martín, Martin Laurenzis, and Matthias B Hullin. Tracking objects outside the line of sight using 2D intensity images. *Scientific Reports*, 6:32491, 2016.

[97] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[98] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 807–814, 2010.

[99] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International Conference on Machine Learning*, pages 1139–1147, 2013.

[100] Adrian Ankiewicz and G-D Peng. Generalized gaussian approximation for single-mode fibers. *Journal of lightwave technology*, 10(1):22–27, 1992.

[101] Jae-Ho Han, Junghoon Lee, and Jin U Kang. Pixelation effect removal from fiber bundle probe based optical coherence tomography imaging. *Optics Express*, 18(7):7427–7439, 2010.

[102] 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.

[103] AA Friesem and U Levy. Parallel image transmission by a single optical fiber. *Optics Letters*, 2(5):133–135, 1978.

[104] Anthony M Tai and AA Friesem. Transmission of two-dimensional images through a single optical fiber by wavelength–time encoding. *Optics Letters*, 8(1):57–59, 1983.

[105] Erich Spitz and Alain Werts. Transmission des images à travers une fibre optique. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences Serie B*, 264(14):1015–+, 1967.

[106] Gilmore J Dunning and RC Lind. Demonstration of image transmission through fibers by optical phase conjugation. *Optics Letters*, 7(11):558–560, 1982.

[107] 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.

[108] Ivo M Vellekoop and AP Mosk. Focusing coherent light through opaque strongly scattering media. *Optics Letters*, 32(16):2309–2311, 2007.

[109] Roberto Di Leonardo and Silvio Bianchi. Hologram transmission through multi-mode optical fibers. *Optics Express*, 19(1):247–254, 2011.

[110] Sébastien Popoff, Geoffroy Lerosey, Mathias Fink, Albert Claude Boccara, and Sylvain Gigan. Image transmission through an opaque material. *Nature Communications*, 1:81, 2010.

[111] SM Popoff, G Lerosey, R Carminati, M Fink, AC Boccara, and S Gigan. Measuring the transmission matrix in optics: an approach to the study and control of light propagation in disordered media. *Physical Review Letters*, 104(10):100601, 2010.

[112] Kathy J Horadam. *Hadamard matrices and their applications*. Princeton university press, 2012.

[113] Youngwoon Choi, Taeseok Daniel Yang, Christopher Fang-Yen, Pilsung Kang, Kyoung Jin Lee, Ramachandra R Dasari, Michael S Feld, and Wonshik Choi. Overcoming the diffraction limit using multiple light scattering in a highly disordered medium. *Physical Review Letters*, 107(2):023902, 2011.

[114] Tomáš Čižmár, Michael Mazilu, and Kishan Dholakia. In situ wavefront correction and its application to micromanipulation. *Nature Photonics*, 4(6):388, 2010.

[115] 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.

[116] Tomáš Čižmár and Kishan Dholakia. Exploiting multimode waveguides for pure fibre-based imaging. *Nature Communications*, 3:1027, 2012.

[117] Sergey Turtaev, Ivo T Leite, Tristan Altwegg-Boussac, Janelle MP Pakan, Nathalie L Rochefort, and Tomáš Čižmár. High-fidelity multimode fibre-based endoscopy for deep brain in vivo imaging. *Light: Science & Applications*, 7(1):92, 2018.

[118] Ioannis N Papadopoulos, Salma Farahi, Christophe Moser, and Demetri Psaltis. High-resolution, lensless endoscope based on digital scanning through a multimode optical fiber. *Biomedical Optics Express*, 4(2):260–270, 2013.

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

[120] Dirk E Boonzajer Flaes, Jan Stopka, Sergey Turtaev, Johannes F De Boer, Tomáš Tyc, and Tomáš Čižmár. Robustness of light-transport processes to bending deformations in graded-index multimode waveguides. *Physical Review Letters*, 120(23):233901, 2018.

[121] 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.

[122] Donald B Conkey, Antonio M Caravaca-Aguirre, and Rafael Piestun. High-speed scattering medium characterization with application to focusing light through turbid media. *Optics Express*, 20(2):1733–1740, 2012.

[123] Antonio M Caravaca-Aguirre, Eyal Niv, Donald B Conkey, and Rafael Piestun. Real-time resilient focusing through a bending multimode fiber. *Optics Express*, 21(10):12881–12887, 2013.

[124] Martin Plöschner, B Straka, K Dholakia, and T Čižmár. GPU accelerated toolbox for real-time beam-shaping in multimode fibres. *Optics Express*, 22(3):2933–2947, 2014.

[125] Martin Plöschner and Tomáš Čižmár. Compact multimode fiber beam-shaping system based on gpu accelerated digital holography. *Optics Letters*, 40(2):197–200, 2015.

[126] Ryosuke Takagi, Ryoichi Horisaki, and Jun Tanida. Object recognition through a multimode fiber. *Optical Review*, 24(2):117–120, 2017.

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

[128] 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):69, 2018.

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

[130] Alex Turpin, Ivan Vishniakou, and Johannes d. Seelig. Light scattering control in transmission and reflection with neural networks. *Optics Express*, 26(23):30911–30929, 2018.

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

[132] Eadweard Muybridge. *The human figure in motion*. Courier Corporation, 2012.

[133] Transmission of natural scenes through a 1 m-long multimode fibre. https://doi.org/10.5525/gla.researchdata.751.

[134] Mitsuo Takeda and Takaaki Kishigami. Complex neural fields with a hopfield-like energy function and an analogy to optical fields generated in phase-conjugate resonators. *JOSA A*, 9(12):2182–2191, 1992.

[135] Akira Hirose. Complex-valued neural networks: The merits and their origins. In *2009 International Joint Conference on Neural Networks*, pages 1237–1244. IEEE, 2009.

[136] Chiheb Trabelsi, Olexa Bilaniuk, Ying Zhang, Dmitriy Serdyuk, Sandeep Subramanian, João Felipe Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, and Christopher J Pal. Deep complex networks. *arXiv preprint arXiv:1705.09792*, 2017.

[137] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[138] Angélique Drémeau, Antoine Liutkus, David Martina, Ori Katz, Christophe Schülke, Florent Krzakala, Sylvain Gigan, and Laurent Daudet. Reference-less measurement of the transmission matrix of a highly scattering material using a dmd and phase retrieval techniques. *Optics Express*, 23(9):11898–11911, 2015.

[139] Arnaud Dubois, Laurent Vabre, Albert-Claude Boccara, and Emmanuel Beaurepaire. High-resolution full-field optical coherence tomography with a linnik microscope. *Applied Optics*, 41(4):805–812, 2002.

[140] Mathias Fink, Gabriel Montaldo, and Mickael Tanter. Time reversal acoustics. In *IEEE Ultrasonics Symposium, 2004*, volume 2, pages 850–859. IEEE, 2004.

[141] Richard C Aster, Brian Borchers, and Clifford H Thurber. *Parameter estimation and inverse problems*. Elsevier, 2018.

[142] Claire Prada, Sébastien Manneville, Dimitri Spoliansky, and Mathias Fink. Decomposition of the time reversal operator: Detection and selective focusing on two scatterers. *The Journal of the Acoustical Society of America*, 99(4):2067–2076, 1996.

[143] Alexandre Aubry and Arnaud Derode. Singular value distribution of the propagation matrix in random scattering media. *Waves in Random and Complex Media*, 20(3):333–363, 2010.

[144] Alan Edelman and Yuyang Wang. Random matrix theory and its innovative applications. In *Advances in Applied Mathematics, Modeling, and Computational Science*, pages 91–116. Springer, 2013.

[145] Antonia M Tulino, Sergio Verdú, et al. Random matrix theory and wireless communications. *Foundations and Trends® in Communications and Information Theory*, 1(1):1–182, 2004.

[146] Stefan Rotter and Sylvain Gigan. Light fields in complex media: Mesoscopic scattering meets wave control. *Reviews of Modern Physics*, 89(1):015005, 2017.

[147] Vladimir A Marčenko and Leonid Andreevich Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, 1967.

[148] Wai-Hon Lee. Iii computer-generated holograms: Techniques and applications. In *Progress in Optics*, volume 16, pages 119–232. Elsevier, 1978.

[149] Changhyeong Yoon, Youngwoon Choi, Moonseok Kim, Jungho Moon, Donggyu Kim, and Wonshik Choi. Experimental measurement of the number of modes for a multimode optical fiber. *Optics Letters*, 37(21):4558–4560, 2012.

[150] Liang Deng, Joseph D Yan, Daniel S Elson, and Lei Su. Characterization of an imaging multimode optical fiber using a digital micro-mirror device based single-beam system. *Optics Express*, 26(14):18436–18447, 2018.

[151] Poom Chiarawongse, Huanan Li, Wen Xiong, Chia Wei Hsu, Hui Cao, and Tsampikos Kottos. Statistical description of transport in multimode fibers with mode-dependent loss. *New Journal of Physics*, 20(11):113028, 2018.

[152] Alex Turpin, Gabriella Musarra, Valentin Kapitany, Francesco Tonolini, Ashley Lyons, Ilya Starshynov, Federica Villa, Enrico Conca, Francesco Fioranelli, Roderick Murray-Smith, and Daniele Faccio. Spatial images from temporal data. *arXiv preprint arXiv:1912.01413*, 2019.

[153] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[154] Andrew Brock, Theodore Lim, James M Ritchie, and Nick Weston. Neural photo editing with introspective adversarial networks. *arXiv preprint arXiv:1609.07093*, 2016.

List of publications

Published during the PhD

Piergiorgio Caramazza, Kali Wilson, Genevieve Gariepy, Jonathan Leach, Stephen McLaughlin, Daniele Faccio and Yoann Altmann. Enhancing the recovery of a temporal sequence of images using joint deconvolution. *Scientific Reports*, 8(1):5257, 2018.

Piergiorgio Caramazza, Alessandro Boccolini, Daniel Buschek, Matthias Hullin, Catherine F Higham, Robert Henderson, Roderick Murray-Smith and Daniele Faccio. Neural network identification of people hidden from view with a single-pixel, single-photon detector. *Scientific Reports*, 8(1):11945, 2018.

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*, pages 3280–3291, 2018.

Piergiorgio Caramazza, Oisín Moran, Roderick Murray-Smith and Daniele Faccio. Transmission of natural scene images through a multimode fibre. *Nature Communications*, 10(1):2029, 2019.

Other publications

F Di Mei, P Caramazza, D Pierangeli, G Di Domenico, H Ilan, A J Agranat, P Di Porto and E Del Re. Intrinsic negative mass from nonlinearity. *Physical Review Letters*, 116(15):153902, 2016.