



# Deep learning enabled real time speckle recognition and hyperspectral imaging using a multimode fiber array

ULAS KÜRÜM, PETER R. WIECHA, REBECCA FRENCH, AND OTTO L. MUSKENS\*

*Physics and Astronomy, Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, UK*

*\*O.Muskens@soton.ac.uk*

*<https://inanophotonics.southampton.ac.uk/>*

**Abstract:** We demonstrate the use of deep learning for fast spectral deconstruction of speckle patterns. The artificial neural network can be effectively trained using numerically constructed multispectral datasets taken from a measured spectral transmission matrix. Optimized neural networks trained on these datasets achieve reliable reconstruction of both discrete and continuous spectra from a monochromatic camera image. Deep learning is compared to analytical inversion methods as well as to a compressive sensing algorithm and shows favourable characteristics both in the oversampling and in the sparse undersampling (compressive) regimes. The deep learning approach offers significant advantages in robustness to drift or noise and in reconstruction speed. In a proof-of-principle demonstrator we achieve real time recovery of hyperspectral information using a multi-core, multi-mode fiber array as a random scattering medium.

Published by The Optical Society under the terms of the [Creative Commons Attribution 4.0 License](#). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

## 1. Introduction

Motivated by the need for imaging in complex environments and through opaque media, new techniques for characterizing and controlling multiple scattering are currently seeing a tremendous development [1]. This new toolbox opens up directions for controlling light in random media and exploiting it for applications such as imaging and optical information processing. At the same time, improved understanding allows us to retrieve more information from seemingly random scattering fields to see around corners and through opaque media [2, 3]. Control over light scattering has led to exciting new applications such as programmable multiport optical elements for classical and quantum states [4–7], quantum secure keys [8] and compressive sampling imaging systems [9].

The multispectral characteristics of speckle fields have been used successfully in a range of studies to realize speckle spectrometers [10–15]. By exploiting wavelength-dependent speckle patterns from a multimode fiber, a spectral resolution of picometers in the near-infrared and nanometers in the visible region has been demonstrated [16, 17]. Next to single-channel spectrometers, multiplexing of spectrally resolved speckle fields into hyperspectral imaging systems is of great interest. Compared to traditional approaches such as Integral Field Spectrometers [18, 19], complex media can offer new opportunities for combining broadband transmission with high spectral resolution [20–22]. In the spatial domain, multimode fibers as well as multi-core fiber bundles are a topic of study for a variety of imaging applications such as remote sensing and endoscopy [23–26].

Recently, a multicore multimode fiber bundle has been used as a frequency characterization element in a high-throughput imaging spectrometer for snapshot spatial and spectral measurements with sub nanometer spectral resolution [27]. A compressive sensing (CS) algorithm was successfully employed to retrieve spectral information. Convex regularization techniques such as

CS provide a suitable solution for a number of problems in computational imaging. However, in addition to their high computational cost, the applicability depends on a sparsity assumption and indeed performance is reduced for dense data. Faster data processing and requirement of robustness with respect to noise and drift could benefit from an entirely different approach based on artificial neural networks [28–34]. Computational methods using neural networks that are trainable for specific problems have recently been shown to be highly efficient and fast [35–37]. Recently, this approach has been used in various applications utilizing speckle patterns such as image reconstruction, object classification and recognition [38–44].

Here, we demonstrate the successful application of Deep Learning (DL) neural networks to the retrieval of spectral information from speckle images. Using a multi-mode, multicore fiber array as a multiplexed speckle spectrometer, we achieve real-time spectral imaging over several thousands of individual fiber cores. Besides being orders of magnitude faster than other, CS-based techniques, we investigate the robustness of DL to noise as well as to image shifts that could originate from thermal expansion or vibrations in the imaging system. We show the adaptability of DL in such conditions with good performance achieved by appropriate training. Results for DL are compared with CS and with analytical regularized inversion approaches. We find that DL performs well both in the compressive and oversampling regimes, combining a good balance in characteristics with fast reconstruction speeds and massive parallelized performance over many fiber cores.

## 2. Method

We use a multi-core, multimode fiber (MCMMF) as the complex scattering medium (Edmund Optics, Fiber optic image conduit). Mode mixing in each individual fiber results in a characteristic speckle pattern with a wavelength dependence determined by the fiber length and the angle of incidence of the incident light. For a direct comparison with previous results we used in our first calibration experiments the setup as described in [27]. For the projection of animations we developed a new setup as shown in Fig. 1(a) based on an acousto-optic tunable filter (AOTF) and spatial light modulator (SLM). In short, a supercontinuum light source (Fianium SC400) was spectrally filtered using an AOTF with a resolution of 5 nm. The filtered light was projected onto a single-mode fiber to ensure stability of the illumination beam with wavelength and to eliminate any other forms of spectral drift in the setup. The fiber output was reflected off the liquid crystal spatial light modulator (SLM, Holoeye Pluto) and was projected onto the MCMMF at an incident angle of  $4^\circ$  with an image demagnification of 5:1. The MCMMF consisted of 3012 fibers with individual core diameters of 50  $\mu\text{m}$ . Fibers of different lengths can be used for different applications depending on the bandwidth required [27]. After transmission through the fiber array, the output facet of the MCMMF was imaged onto the focal plane array of a 12-bit, 5 MPixel monochrome CMOS camera with a pixel size of  $2.2 \mu\text{m} \times 2.2 \mu\text{m}$  (AVT Guppy) using a 1:1 imaging system. Collected images were transferred to PC via IEEE 1394a and saved in uncompressed TIF format in  $2592 \text{ (H)} \times 1944 \text{ (V)}$  resolution in 12 bits.

Figure 1(b) illustrates the typical information obtained at the exit surface of the MCMMF in the form of wavelength dependent speckle patterns obtained from the individual fiber cores. Each pattern corresponds to a superposition of higher order fiber modes that is dependent on the wavelength, the length of the fiber and the angle of incidence. All fiber cores are slightly different and local variations in the material properties, strain, impurities and other random structural elements give rise to an individual set of speckle patterns for each fiber core in the array. The speckle patterns for every wavelength are stored into a multispectral transmission matrix for every core, which in principle allows retrieval of spectral information from arbitrary superposition states using a number of different techniques. Spectra consisting of more than one wavelength component, as well as continuous spectra, result in a superposition of many speckle patterns. Analytical inversion techniques like Moore-Penrose pseudo-inversion can be employed

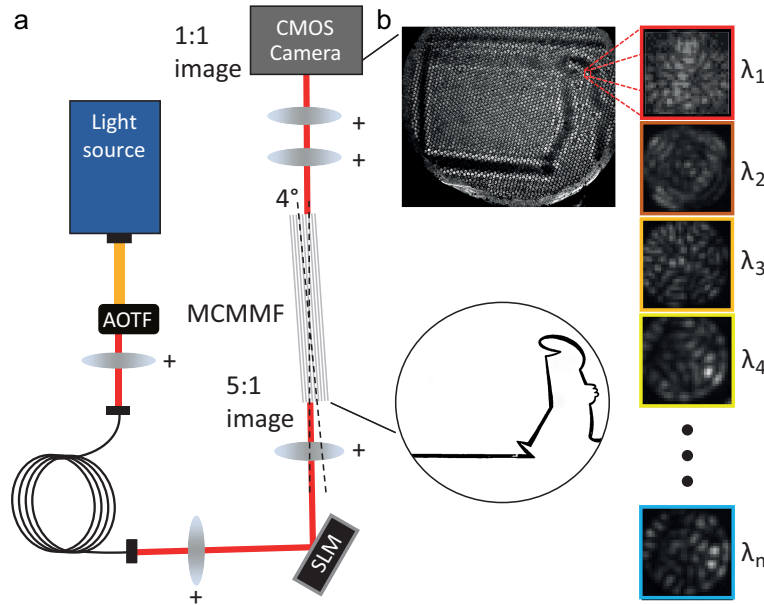


Fig. 1. (a) Scheme of the experimental setup including broadband supercontinuum laser source, acousto-optical tunable filter (AOTF), Spatial Light Modulator (SLM) used for image generation and the multi-core, multimode fiber (MCMF). (b) Original projected image and detected camera image of the exit interface of the fiber bundle at a single selected wavelength with typical speckle patterns of a selected fiber core for different input wavelengths ( $\lambda_1 - \lambda_n$ ). La Linea, with permission, copyright CAVA/QUIPOS.

to reconstruct the spectra from these superpositions, but their performance is strongly dependent on noise and appropriate regularization is needed. In this work we compare our DL approach with analytical inversion using Tikhonov regularization (TR) [9]. Moreover, analytical inversion is limited to the oversampled regime, and its breakdown is observed at the Shannon-Nyquist sampling limit [22]. Compressive sensing (CS) extends reconstruction into the undersampling regime under conditions of sparsity, in our work CS was implemented for comparison to DL using the python package “cvxpy” [45].

Spectral reconstruction via DL was implemented using a convolutional neural network (CNN), composed of a series of convolutional layers followed by two fully connected layers of 512 and 256 nodes with dropout regularization using 70% keep probability. The final, dense output layer of 43 neurons represents the spectrum, where each neuron corresponds to a discrete wavelength channel. The size of the CNN was manually optimized for each tested sampling condition. For a region of interest (ROI) size of 5x5 pixels, the best performing network consists of two convolutional layers with each a 2x2 kernel (CNN (i) in Fig. 2, yellow). On 20x20 pixels, a three-layer CNN with kernel size 3x3 throughout the network was found to perform best (CNN (ii) in Fig. 2, blue). Each convolution is followed by batch normalization and a leaky ReLU activation layer, all layers use valid padding. We found that any type of pooling consistently reduced the reconstruction quality, so no pooling was performed. The networks were implemented in python using keras as frontend for tensorflow [46].

To test the performance of the different approaches, multiple patterns were digitally added up together to simulate a real signal made of a given number,  $N_\lambda$ , of nonzero wavelength components with randomly varying intensities. The images of the speckle patterns were cropped to various ROI sizes to achieve different regimes of oversampling and undersampling, as given by the ratio

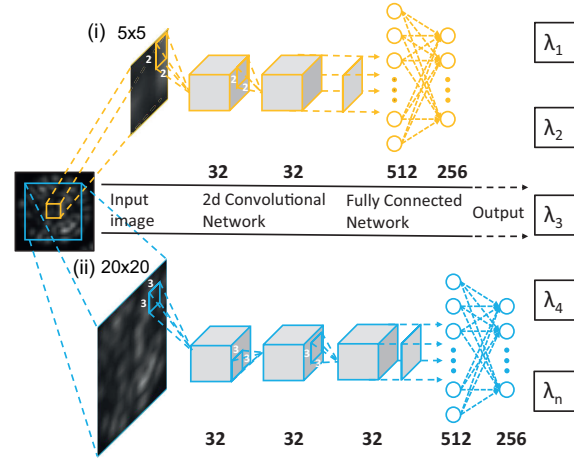


Fig. 2. Neural network structures used in this study for pixel areas of (i) 5x5 pixels ( $Y/X = 0.58$ ) and (ii) 20x20 pixels ( $Y/X = 9.30$ ).

of the total number of calibrated wavelengths,  $Y$ , to the total number of pixels of the ROI,  $X$ . For each multimode fiber, a total dataset of 31,000 images was generated, of which 29,000 were used to train the neural network, 1000 served for validation during training and the remaining 1000 were used for the final evaluation. A plot demonstrating training convergence is given in Fig. 7 in the Appendix.

### 3. Results and discussion

#### 3.1. Deep learning reconstruction quality

A direct numerical illustration of the reconstruction capabilities of the DL approach is shown in Fig. 3. In Fig. 3(a) and 3(b) the performance is shown for different sampling regimes ranging from oversampling ( $Y/X = 9.30$ ) down to deep undersampling ( $Y/X = 0.21$ ). The cartoons in Fig. 3(a) illustrate the quality of the reconstruction, where a smiley emoticon was used as the ground truth in a single wavelength channel. For a single non-zero wavelength  $N_\lambda=1$ , this is the only information contained in the spectrum. For the case  $N_\lambda = 10$ , nine other wavelengths in the spectrum are filled with a crossed-out symbol in the form of a capital “X”. We see that in both cases, the reconstruction of the target is very good in the oversampling regime, while the quality becomes poorer below the sampling limit,  $Y/X < 1$ . For  $N_\lambda = 10$  we can see the appearance of the cross in the image, indicating the presence of significant cross-talk between the spectral channels.

A quantitative analysis of this dependence of the reconstruction quality on the sampling rate is presented in Fig. 3(b), where the cross correlation of the reconstructed spectrum with the input spectrum (ground truth) is plotted versus the sampling ratio  $Y/X$ . We can clearly see the main trends identified in the illustration, namely a good quality of reconstruction in the oversampling regime and a degradation in performance for  $Y/X < 1$ . In the undersampling regime, we see that the reconstruction improves for lower number of nonzero wavelengths, with reasonably good performance (defined as correlation  $> 0.5$ ) for sampling rates as low as  $Y/X=0.21$  for just a single non-zero wavelength. Clearly, DL is able to infer meaningful results under conditions in which the information density is sparse and where analytical inversion techniques show a complete breakdown [22]. In other words, DL is able to cross over far into the compressive sensing regime and therefore exhibits similar characteristics to a CS-based approach. Having shown the strength of DL in the compressive sensing regime, it is of interest to investigate its capabilities in the



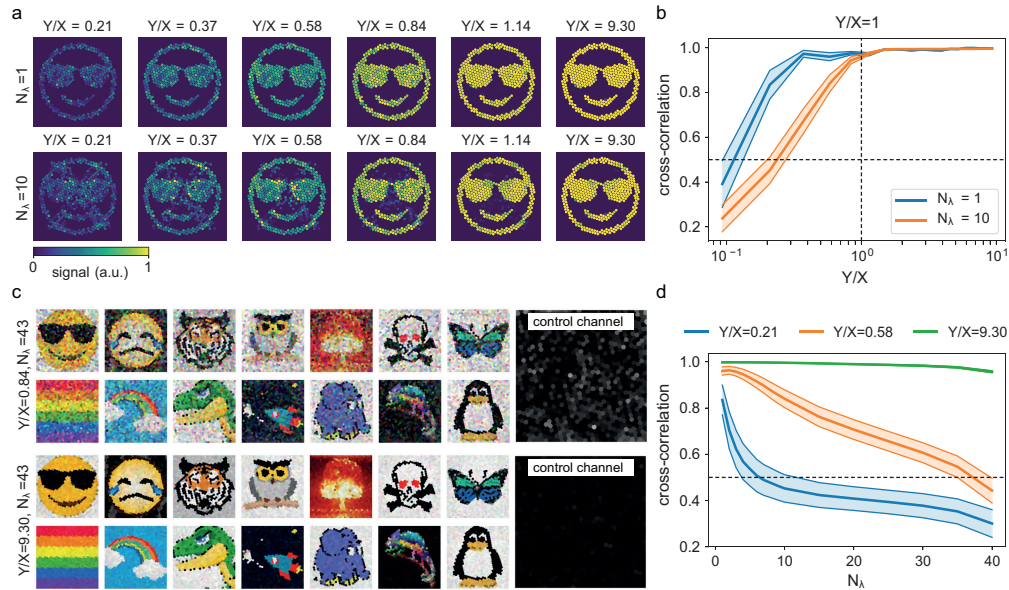


Fig. 3. (a) Numerical illustration and (b) calculation of reconstruction quality using DL for different sampling rates  $Y/X$ , for  $N_\lambda = 1$  and  $N_\lambda = 10$  non-zero wavelengths. One wavelength carries the encoded image (smiley), all other non-zero channels encode the image of a capital “X”, which becomes slightly visible at low sampling rates due to cross-talk (see Appendix). (c) Numerical illustration of image reconstruction using DL for dense spectra ( $N_\lambda = 42$ ) showing 14 RGB images that are encoded in 42 wavelength channels, the 43rd, blank channel serves for cross-talk control. Reconstructions are shown for undersampling  $Y/X = 0.84$  and oversampling  $Y/X = 9.30$  regimes. (d) Cross-correlation with ground truth as a function of number of non-zero wavelengths in the spectrum for different sampling rates. Results in (b,d) are averaged over the whole fiber stack and for 100 spectra per fiber core. Light areas indicate the standard deviation of the data. Dashed line at  $Y/X = 1$  corresponds to Nyquist-Shannon sampling limit. Dashed line at a cross-correlation of 0.5 indicates the threshold below which the reconstruction is considered to have failed. Underlying full spectral data for (a) and (c) are presented in Fig. 9 and Fig. 10 of the Appendix. All shown cliparts are from [www.openclipart.org](http://www.openclipart.org) and public domain.

regime of dense information, under conditions where the sparsity assumption underpinning CS ceases its validity. We start again with a numerical illustration in Fig. 3(c) to visualize the amount of information that can be encoded through speckle images. Using  $Y = 43$  available wavelength channels, we encoded separately the red, green and blue (RGB) channels of 14 independent RGB images using the experimentally obtained transmission matrix. The remaining unused wavelength channel allows to assess the residual cross-talk. Raw RGB reconstruction data are given in the Supporting Information.

Figure 3(c) shows the reconstructed RGB images obtained using our DL approach in either the regime of undersampling ( $Y/X = 0.84$ ) or oversampling ( $Y/X = 9.30$ ). In case of oversampling ( $Y/X > 1$ ), the neural network has much more input information to work with, which results in excellent image reconstruction quality and low residual cross-talk. For the undersampling regime, the images are still discernible but with significant reconstruction noise and cross-talk. These trends are again quantified in the accompanying analysis of Fig. 3(d), showing the cross-correlation with the ground truth versus number of non-zero wavelengths. We see that the network output is almost perfectly correlated in the oversampled case (perfect reconstruction),

which holds even for dense spectra, where signal is present in all wavelength channels ( $N_\lambda=43$ ). As seen in Fig. 3(c), for increasing number of wavelengths the effect of undersampling results in a rapid decrease of reconstruction fidelity. In the appendix we compare the same RGB data with reconstruction by TR and CS in the under- and oversampling regime and find that DL can compete with CS at a visual comparison and largely outperforms the analytical approach in the undersampled measurements.

### 3.2. Comparison of different reconstruction techniques

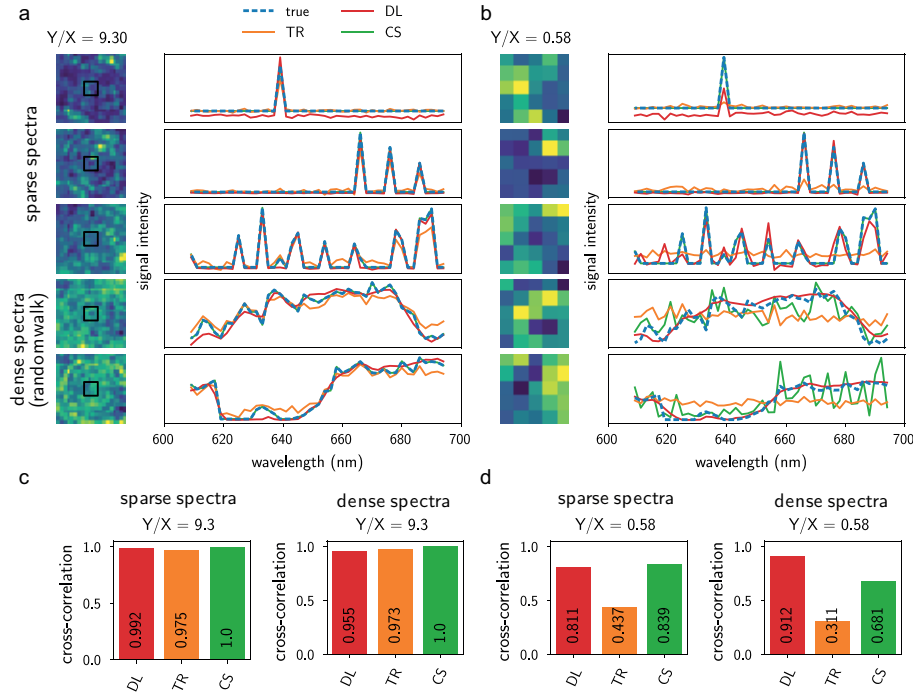


Fig. 4. (a) Examples of speckle images and reconstructed spectral information for sparse (three top rows) and dense spectra (two bottom rows, generated by a random-walk like algorithm). (b) oversampling regime with  $Y/X=9.3$ , right column: undersampling regime with  $Y/X=0.58$ . The black box inside the speckle pattern shows the ROI used in the undersampling case. (c-d) Histograms comparing average cross correlations from 1000 randomly generated sparse ( $< 50\%$  sparsity) and dense (all wavelengths non-zero) spectra, obtained with deep learning (DL), Tikhonov regularization inversion (TR) and compressive sensing (CS) in the oversampling (c) and undersampling (d) regime.

In Fig. 4 deep learning (DL) is compared directly with both the TR and CS reconstruction methods. In this benchmark, 1000 randomly generated spectra were generated numerically from the experimental transmission matrix. Figure 4(a) shows the oversampling case ( $Y/X=9.30$ ) corresponding to an ROI of  $20 \times 20$  pixels, while Fig. 4(b) shows the undersampling case ( $Y/X=0.58$ ) corresponding to only  $5 \times 5$  pixels. Several examples of typical spectra are shown (blue dash, ground truth), together with their corresponding reconstructions using DL (red), TR (orange) and CS (light blue). The lower two examples correspond to continuous spectra with a high density of spectral information.

Figure 4(c) and 4(d) gives the full quantitative analysis of the average of the cross-correlation between each of the 1000 randomly generated spectra (ground truth) and its respective reconstruction. In the oversampling regime, all methods perform well with correlation values  $> 0.95$ . In

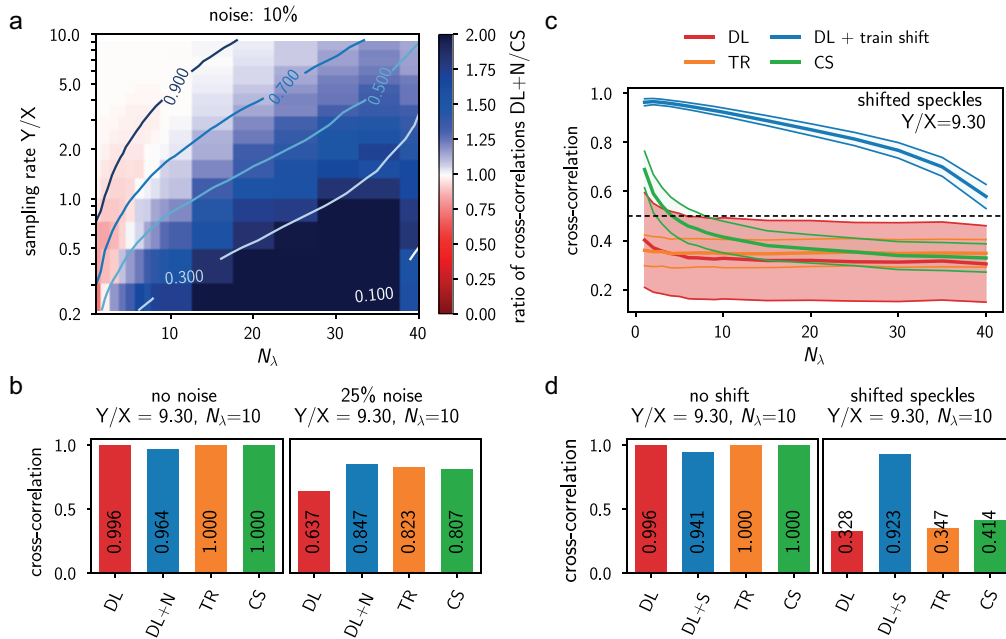


Fig. 5. (a) Map showing ratio of reconstruction quality (cross-correlations) for deep learning trained on noisy data (DL+N) over compressive sensing (CS) for 10% added noise. Blue indicates DL+N better than CS, red indicates CS better than DL+N. Contour lines indicate the cross-correlation of DL+N speckle reconstruction. (b) Calculated cross-correlations without noise and in presence of 25% noise. DL+N outperforms CS over a large part of parameter space where the cross-correlation  $< 0.9$ . (c) Robustness of the reconstruction against shifts of the speckle patterns by one pixel in a random direction. (d) Calculated cross-correlations without shift and with shift of 1 pixel. DL can be trained on data including shift (DL+S), which renders the method very robust in such scenario, largely outperforming TR and CS.

the undersampling regime TR fails completely (average cross-correlation  $< 0.5$ ) as it can be seen to produce a mostly flat spectrum for all cases irrespective of the spectral shape. DL yields a very good performance and even clearly outperforms CS for dense spectra in the undersampling case.

The slightly weaker performance of DL compared to CS in the oversampling case can be explained by the statistical training procedure in DL, while CS on the other hand is an analytical approach, yielding generally a close-to-optimum solution – however at a significantly higher computational cost compared to the neural network reconstruction, as discussed further below. In the undersampling case we observe that DL tends to result in inferred spectra that are smoother than the original input spectrum, whereas CS results in more spikes in the spectra even when the input is smooth.

The previous tests considered speckle reconstruction using a multispectral transmission matrix in the complete absence of noise. In a real-life scenario, one can expect some level of noise to be present in the image, either shot noise, electronic camera noise or other non-specific backgrounds. An imaging system may also experience some drift caused by environmental effects, such as vibrations and thermal variations. To assess the robustness of the different approaches against typical perturbations, we compare in Fig. 5 the respective performance of DL, CS and TR. While CS and TR are based on analytic methods which are intrinsically inflexible to variations, DL has the advantage of allowing some level of adaptability through the choice of training data.

### 3.3. Robustness against noise and spatial shifts

To investigate the adaptability of the DL approach, we trained the neural network using noisy and spatially shifted training data with the aim of making it more robust against these effects. To account for the effect of noise, we added normally distributed random intensity noise to every pixel of the speckle pattern. Figure 5(a) shows a parameter map showing the relative performance of DL trained on noisy data (DL+N) against CS, which is quantified as the ratio of their respective cross-correlations with the input spectra. Noise-adapted DL consistently outperforms CS over a large parameter-space. Blue regions indicate a superior DL performance. DL always performs better for relatively dense spectra with  $N_\lambda > 15$ . Even in most of the “white” parts of the colourplot, DL is outperforming TR and CS by at least some percent (see bar plots in Fig. 5(b)). CS is better performing only on sparse data at very low sampling rates. Furthermore TR contains a free parameter which has to be empirically adjusted to match a given noise level in order to achieve the specified performance [9]. The neural network trained on noisy data (DL+N) outperforms the normal DL when dealing with noisy spectra. This increased adaptability of the DL+N neural network comes at the cost of a relatively small hit on performance when dealing with noiseless data as seen in Fig. 5(b).

In principle, the performance of CS and TR on noisy data can be improved by adding an additional denoising step prior the spectral reconstruction, which of course adds additional computational cost. Since we are interested in developing a real-time-ready approach, computationally expensive image treatment on each speckle should be avoided. Thus, our analysis evaluates, how the different approaches can handle noisy data “out of the box”.

A larger drift of the CCD or of the fiber would completely change the transmission matrix of the system and hence would require a full re-characterization of the setup. In the following we test whether deep learning can adapt to such situations, where analytical methods and compressive sensing break down definitely. To simulate spatial drift, the cropped ROI was randomly shifted by one pixel in arbitrary directions. We compare in Fig. 5(c) and 5(d) the performance of networks that were trained with and without spatial shift. If the neural network is trained without such perturbed data, the performance of DL decreases, just like for the conventional methods TR and CS. However, if the training datasets contain accordingly perturbed data, DL shows a distinct performance boost at the reconstruction from non-perfect speckle images. In the case of shift (Fig. 5(c) and 5(d)), accordingly trained DL dramatically outperforms the other techniques on perturbed data, while again the loss of performance on unshifted data is relatively small.

We note that the training set might not only contain data accounting for shifts of the CCD, but also for shifts of the fiber or variations in the incident angle of light. All these effects would affect the speckle patterns in a deterministic way and could be corrected by an accordingly trained artificial neural network, whilst this is impossible with an analytical method.

### 3.4. Real time reconstruction of hyperspectral images

Finally, apart from its universal applicability and superior stability, the most important achievement of using DL over CS is its reconstruction speed. We therefore demonstrate that our deep learning based hyperspectral reconstruction method is capable of real-time image processing. To this end we projected a grayscale video via an SLM onto the MCMMF. During video playback we randomly changed the wavelength of the illumination using the AOTF. We analyzed in real time the 2700 speckle patterns for each core, captured by the CMOS using artificial neural networks (one network for each fiber), which were trained in advance and pre-loaded into RAM. Figure 6(a) shows frames from the original video data in the top row, and the three first wavelengths of the spectral reconstruction in the bottom rows (for full videos see Visualizations). Upon switching of the wavelength, the reconstructed image changes channel with limited cross-talk between channels. At 1.74s, the AOTF was switched during the frame acquisition, briefly resulting in two spectral components. We note that the reconstruction quality is a bit worse compared to the synthetic

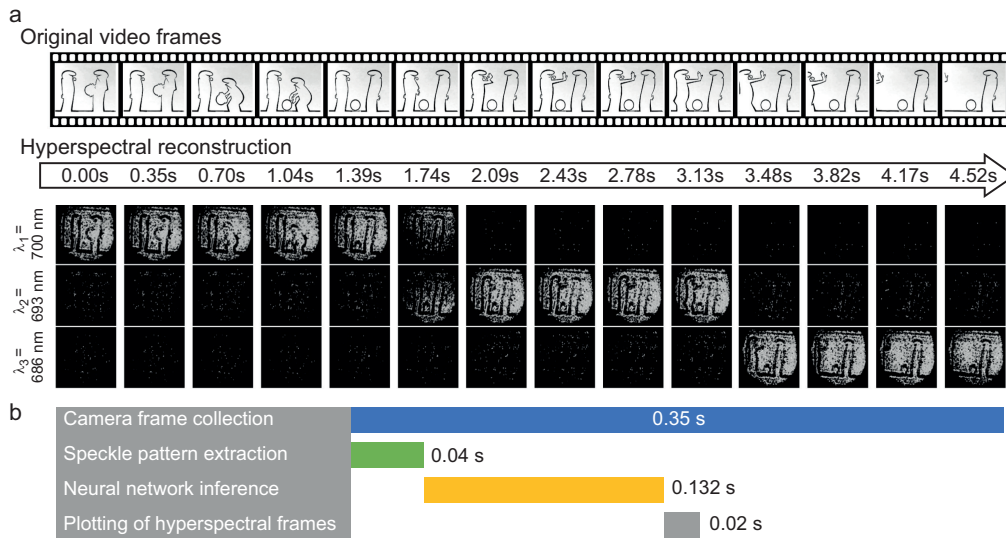


Fig. 6. (a) Real-time speckle-based hyperspectral video reconstruction via DL. A video is projected on the fiber bundle using an SLM in amplitude-modulation configuration. During playback the wavelength of the projecting light is changed. Top: Original frames of the input video (see also Visualizations). Bottom: Spectral reconstruction of three wavelength channels of the full multi-core fiber (approx. 2700 fiber cores). (b) Bar graph showing the timings of the different execution steps. Full visualizations are included in the Supporting Materials of this study. La Linea, with permission, copyright CAVA/QUIPOS.

data shown before, which is due to non-perfect intensity stability of the setup and some spurious residual wavelength correlations in the used, shorter (2.54cm) fiber. Further visualizations with the “La Linea” video ([Visualization 1](#), [Visualization 2](#), [Visualization 3](#), [Visualization 4](#)) and with an “Eclipse” video ([Visualization 5](#)) are also available. Finally, using microsecond-fast frequency sweeps with the AOTF, we were able to transmit spectrally broadened wavelength windows through the fiber bundle. Neural network based, multi-wavelength hyperspectral image reconstruction of experimental real-time data is demonstrated in [Visualization 2](#), [Visualization 3](#), and [Visualization 4](#) in the supplementary material.

The training of the networks for 2700 fiber cores takes about 8 hours on an Nvidia Quadro P6000 GPU. This is a one-time procedure since the speckle patterns can be maintained stable over long times. The hyperspectral image reconstruction itself is the time-critical process in many applications. The trained neural networks are capable to reconstruct all 2700 fiber cores in only 132 ms on an Intel i7-3770 CPU. In comparison, CS requires around 2 minutes per frame for processing of the full fiber bundle. TR should potentially be even faster than deep learning, but interestingly, we found that in our python implementation, TR is around 1.5 times slower than the neural network inference (190 ms per video frame), which we attribute to the relatively slow python code, whilst the neural networks are evaluated through the highly optimized tensorflow backend. Hence it should be possible to further accelerate TR. Visually, all three methods perform similarly on the experimental video reconstruction.

Figure 6(b) shows the durations of the steps of DL hyperspectral image reconstruction on the Intel i7-3770 CPU, working with 32GB of RAM for network pre-loading. Including the preprocessing and rendering of the final hyperspectral images, the algorithm in its current state can reconstruct about 5 full images per second, which is faster than the total acquisition and transfer time of our 12bit CMOS camera (about 0.35s). In the supplementary Visualizations, we also show >5fps reconstruction rate, using faster 8bit CMOS readout, which becomes essentially



limited by the neural network inference speed.

There is significant scope for further increasing the reconstruction speed of DL, in a first step already by performance optimization of the code. While not all 2700 networks fitted in the memory of our GPU, a benchmark on 1000 preloaded networks showed that the GPU offers around  $\times 2.5$  speed improvement, multi-GPU platforms will be accordingly faster. Training networks that each reconstruct speckles for multiple fibers in parallel can potentially result in a further significant performance boost and reduce the memory requirement of the approach, as is shown in Fig. 8 of the appendix. Even higher speed can potentially be obtained by developing hardware implementations of the networks, for example based on Field Programmable Gate Arrays (FPGAs) [47]. With respect to our current implementation using Python, a more direct hardware/software communication could also straightforwardly increase the frame collection rates, which however is outside the scope of this proof of principle study.

#### 4. Conclusion

In conclusion, we have shown that with deep learning, a multicore multimode fiber bundle can be used as a real-time hyperspectral camera, robust to noise and spatial shifts. Using wavelength- and fiber-core dependent speckle patterns, we have used deep learning to cope with large amounts of data at a video rate of several frames per second on conventional computer hardware. The imaging spectrometer and deep learning technique are versatile by design, and the calibrated wavelength range can be tailored to specific applications. The approach can be easily scaled to any number of wavelength channels, to desired spectral resolution via the length of the multicore fiber bundle and with respect to imaging spatial resolution. Deep learning in combination with speckle spectrometry enables a new class of low-cost, compact hyperspectral imaging systems with real-time data processing capabilities.

#### Supplementary materials

Visualizations 1-4: “La Linea” short animation demonstrating the spectral deconstruction using the MCMMF and deep learning algorithm for several AOTF wavelength sweep sequences. Visualization 5: “Eclipse” short animation demonstrating the spectral deconstruction using the MCMMF and deep learning algorithm.

#### Appendix

##### *Training convergence*

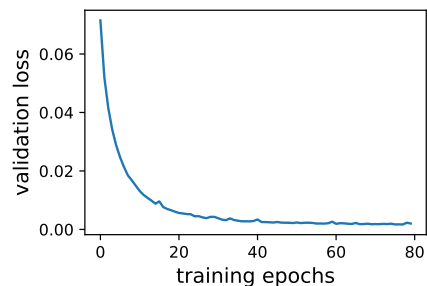


Fig. 7. Least squares loss on validation data (1000 spectra) as function of training epoch. The training set contains 29000 spectra.

To visualize the convergence of the training, in Fig. 7 the validation loss during the training of a

selected fiber channel is plotted as function of training epochs, showing convergence after 40 epochs.

### Multiple fiber-cores per neural network

One main assumption, underlying the multi-core fiber hyperspectral imaging approach, is that the speckles of the individual fiber-cores are completely uncorrelated. We therefore expect no improvement in reconstruction quality by observation of larger CCD areas (thus several fiber cores) with a single neural network. Indeed we can confirm this in according tests: Provided the training set is sufficiently large, the reconstruction quality is not affected by the number of simultaneously observed fibers. On the other hand, the reconstruction quality does also not improve compared to the case of a single observed fiber per network.

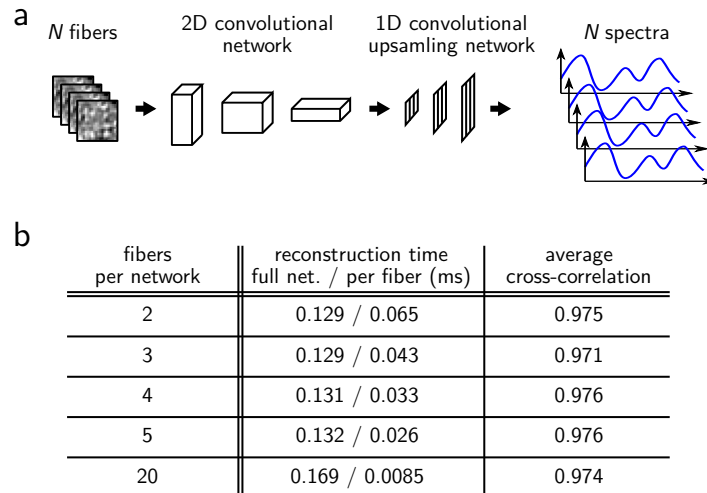


Fig. 8. Training multi-channel 2D convolutional – 1D convolutional upsampling networks on multiple fibers per network. (a) architecture of a multi-fiber network: speckle images from  $N$  different fibers are fed into the network and translated into  $N$  spectra at the output. (b) timing and reconstruction quality for a fixed network architecture for increasing number  $N$  of reconstructed fibers (corresponding to the number of input and output channels). Trained on  $10,000 \times N$  speckles. On 32GB RAM, training data for up to 20 fibers could be kept in memory. The trained networks show no significant difference in reconstruction quality, while the timing per fiber reduces almost linearly. The Memory requirement of each pretrained network is approximately constant, hence required RAM linearly decreases with the number of reconstructed fibers per network.

The first layers of convolutional networks usually develop some basic feature detection filters during training and thus should be valid for all fibers in our application. In conclusion, while the reconstruction quality is similar, deconstructing multiple fiber-speckles with a single network should allow to reduce evaluation time and memory requirements, which we demonstrate by according results, shown in Fig. 8. On the other hand, increasing the number of fibers per network in principle increases the size of the required training set quadratically, because the networks need to be trained on permutations of speckle-image / spectrum pairs for all fibers. Therefore, the gain in computation speed and memory reduction comes at the cost of a significantly slower training phase.

A possible alternative could be transfer learning, where a convolutional network is trained on a set of speckle/spectrum pairs for different fibers and then is used as a pre-trained feature detector for the individual fibers. This could in principle significantly accelerate training for each

fiber and also hugely reduce memory requirements at inference time. While it is not the scope of our proof-of-concept study to develop the technically most possible advanced implementation of neural network speckle reconstruction, it could be subject of future investigations.

#### Visual performance of deep learning hyperspectral reconstruction

Full reconstruction of all 43 wavelength channels for the data shown in Fig. 3a is shown in Fig. 9. In the undersampled measurements (Fig. 9(a-b)), cross-talk is visible by ghost-images of the letter “X” in blank channels.

Figure 10 shows the reconstructions of all wavelength channels for the data given in Fig. 3d.  $\lambda_{43}$  serves as cross-talk control channel.

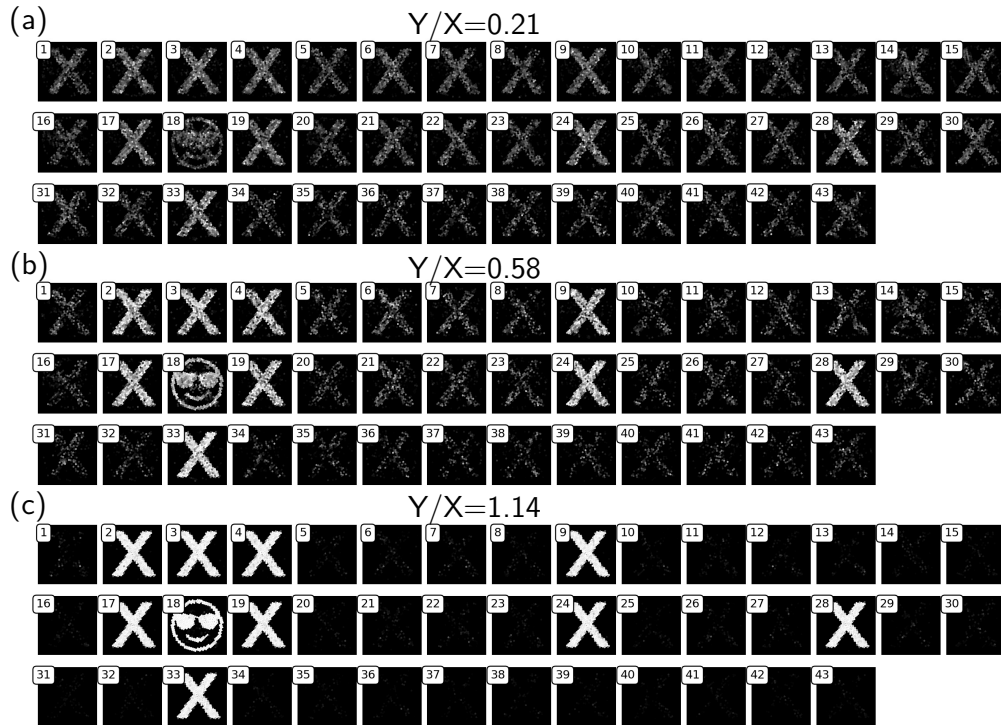


Fig. 9. Reconstruction of all spectral channels with 10 channels containing images at a sampling rate of (a)  $Y/X=0.21$ , (b)  $Y/X=0.58$ , (c) and  $Y/X=1.14$ . Channels carrying information are 2; 3; 4; 9; 17; 18; 19; 24; 28; 33.

#### Visual comparison of DL, CS and TR reconstruction

Figure 11 shows a visual comparison of deep learning with compressive sensing and the Tikhonov regularized transmission matrix approaches in the undersampling (Fig. 11(a)) and oversampling (Fig. 11(b)) regimes. Deep learning shows marginally worse reconstruction compared to CS, which could be due to slight overfitting and could probably be improved by optimization of the network and training hyper-parameters.

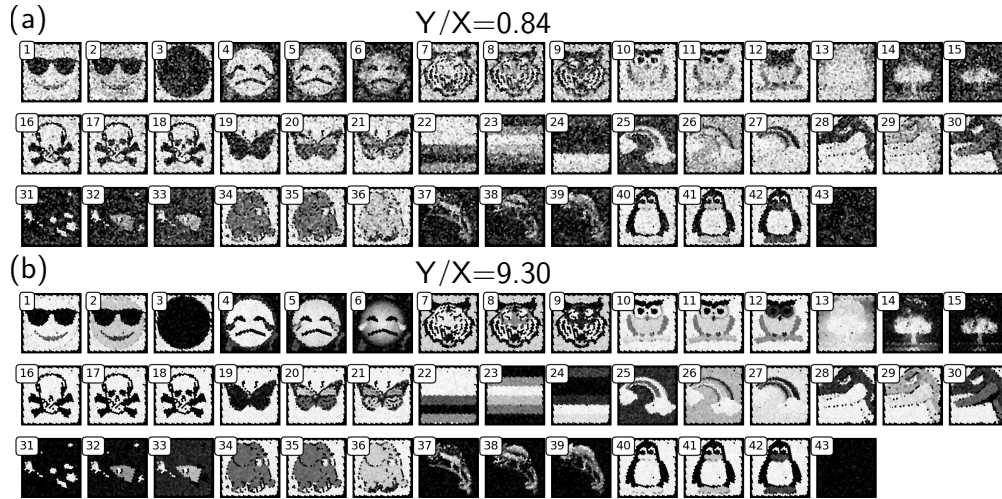


Fig. 10. Encoding and reconstruction of 14 RGB images in the speckle patterns of the multi-core fiber. Raw reconstructions of the spectral channels, containing the red, green and blue parts of the color images, at a sampling rates of (a)  $Y/X=0.84$  and (b)  $Y/X=9.3$ .

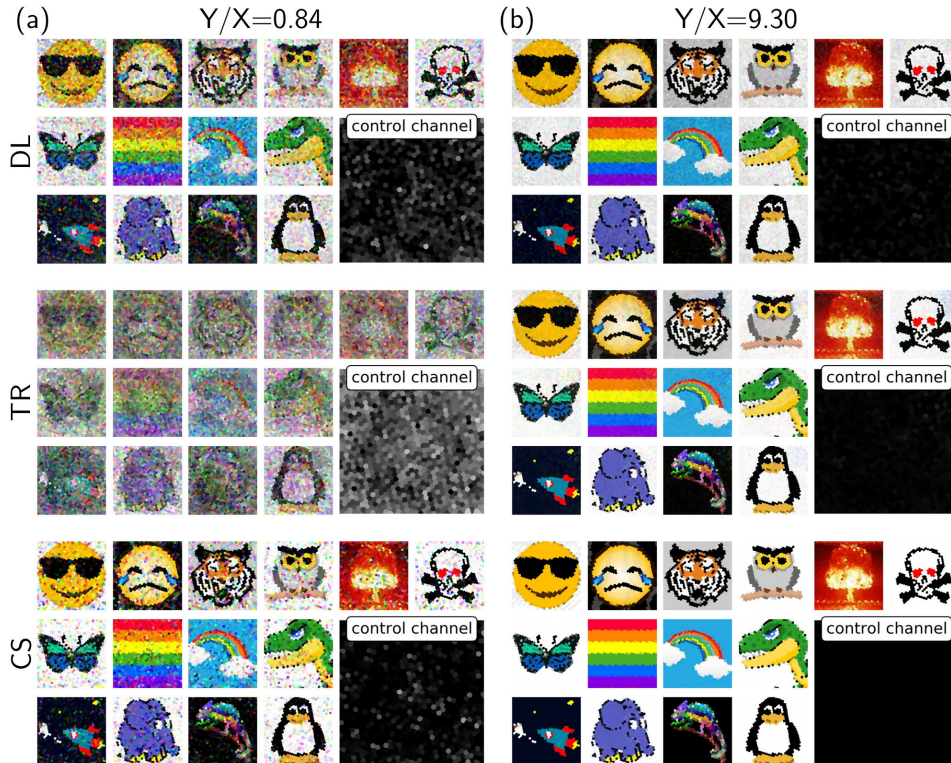


Fig. 11. Visual comparison of reconstruction quality of the different methods (deep learning [DL], Tikhonov regularization [TR] and compressive sensing [CS]) on the RGB dataset (see also Fig. 3). (a) sampling rate of  $Y/X=0.84$  and (b)  $Y/X=9.3$ .



## Funding

German Research Foundation (DFG) Research Fellowship (WI 5261/1-1); EPSRC (EP/J016918/1).

## Acknowledgments

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Quadro P6000 GPU used for this research. La Linea © CAVA/QUIPOS. La Linea usage rights granted to the University of Southampton for the purpose of this research has been approved by Quipos, Osvaldo Cavandoli's worldwide licensor. All data supporting this study are openly available from the University of Southampton repository (DOI: 10.5258/SOTON/D0942).

## References

1. S. Rotter and S. Gigan, "Light fields in complex media: Mesoscopic scattering meets wave control," *Rev. Mod. Phys.* **89**, 015005 (2017).
2. O. Katz, E. Small, and Y. Silberberg, "Looking around corners and through thin turbid layers in real time with scattered incoherent light," *Nat. Photonics* **6**, 549–553 (2012).
3. J. Bertolotti, E. G. van Putten, C. Blum, A. Lagendijk, W. L. Vos, and A. P. Mosk, "Non-invasive imaging through opaque scattering layers," *Nature* **491**, 232–234 (2012).
4. H. Defienne, M. Barbieri, B. Chalopin, B. Chatel, I. A. Walmsley, B. J. Smith, and S. Gigan, "Nonclassical light manipulation in a multiple-scattering medium," *Opt. Lett.* **39**, 6090–6093 (2014).
5. S. R. Huisman, T. J. Huisman, T. A. W. Wolterink, A. P. Mosk, and P. W. H. Pinkse, "Programmable multiport optical circuits in opaque scattering materials," *Opt. Express* **23**, 3102–3116 (2015).
6. T. Strudley, R. Bruck, B. Mills, and O. L. Muskens, "An ultrafast reconfigurable nanophotonic switch using wavefront shaping of light in a nonlinear nanomaterial," *Light. Sci. & Appl.* **3**, e207 (2014).
7. J. Park, J.-Y. Cho, C. Park, K. Lee, H. Lee, Y.-H. Cho, and Y. Park, "Scattering optical elements: Stand-alone optical elements exploiting multiple light scattering," *ACS Nano* **10**, 6871–6876 (2016).
8. S. A. Goorden, M. Horstmann, A. P. Mosk, B. Škorić, and P. W. H. Pinkse, "Quantum-secure authentication of a physical unclonable key," *Optica* **1**, 421–424 (2014).
9. A. Liutkus, D. Martina, S. Popoff, G. Chardon, O. Katz, G. Lerosey, S. Gigan, L. Daudet, and I. Carron, "Imaging with nature: Compressive imaging using a multiply scattering medium," *Sci. Reports* **4**, 5552 (2014).
10. H. Cao, "Perspective on speckle spectrometers," *J. Opt.* **19**, 060402 (2017).
11. B. Redding and H. Cao, "Using a multimode fiber as a high-resolution, low-loss spectrometer," *Opt. Lett.* **37**, 3384–3386 (2012).
12. M. Mazilu, T. Vettenburg, A. D. Falco, and K. Dholakia, "Random super-prism wavelength meter," *Opt. Lett.* **39**, 96–99 (2014).
13. M. Chakrabarti, M. L. Jakobsen, and S. G. Hanson, "Speckle-based spectrometer," *Opt. Lett.* **40**, 3264–3267 (2015).
14. S. F. Liew, B. Redding, M. A. Choma, H. D. Tagare, and H. Cao, "Broadband multimode fiber spectrometer," *Opt. Lett.* **41**, 2029–2032 (2016).
15. G. C. Valley, G. A. Sefler, and T. J. Shaw, "Multimode waveguide speckle patterns for compressive sensing," *Opt. Lett.* **41**, 2529–2532 (2016).
16. B. Redding, M. Alam, M. Seifert, and H. Cao, "High-resolution and broadband all-fiber spectrometers," *Optica* **1**, 175–180 (2014).
17. N. H. Wan, F. Meng, T. Schröder, R.-J. Shiue, E. H. Chen, and D. Englund, "High-resolution optical spectroscopy using multimode interference in a compact tapered fibre," *Nat. Commun.* **6**, 7762 (2015).
18. N. A. Hagen and M. W. Kudenov, "Review of snapshot spectral imaging technologies," *Opt. Eng.* **52**, 090901 (2013).
19. J. G. Dwight and T. S. Tkaczyk, "Lenslet array tunable snapshot imaging spectrometer (LATIS) for hyperspectral fluorescence microscopy," *Biomed. Opt. Express* **8**, 1950–1964 (2017).
20. S. K. Sahoo, D. Tang, and C. Dang, "Single-shot multispectral imaging with a monochromatic camera," *Optica* **4**, 1209–1213 (2017).
21. P. Wang and R. Menon, "Computational multispectral video imaging [invited]," *JOSA A* **35**, 189–199 (2018).
22. R. French, S. Gigan, and O. L. Muskens, "Speckle-based hyperspectral imaging combining multiple scattering and compressive sensing in nanowire mats," *Opt. Lett.* **42**, 1820–1823 (2017).
23. Y. Choi, C. Yoon, M. Kim, T. D. Yang, C. Fang-Yen, R. R. Dasari, K. J. Lee, and W. Choi, "Scanner-free and wide-field endoscopic imaging by using a single multimode optical fiber," *Phys. Rev. Lett.* **109**, 203901 (2012).
24. M. Plöschner, T. Tyc, and T. Čížmár, "Seeing through chaos in multimode fibres," *Nat. Photonics* **9**, 529–535 (2015).
25. E. E. Morales-Delgado, D. Psaltis, and C. Moser, "Two-photon imaging through a multimode fiber," *Opt. Express* **23**, 32158–32170 (2015).
26. A. Porat, E. R. Andresen, H. Rigneault, D. Oron, S. Gigan, and O. Katz, "Widefield lensless imaging through a fiber bundle via speckle correlations," *Opt. Express* **24**, 16835–16855 (2016).
27. R. French, S. Gigan, and O. L. Muskens, "Snapshot fiber spectral imaging using speckle correlations and compressive sensing," *Opt. Express* **26**, 32302–32316 (2018).



28. T. Villmann, M. Kästner, A. Backhaus, and U. Seiffert, "Processing hyperspectral data in machine learning," in *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning Proceedings*, vol. 21 (2013), pp. 1–10.
29. W. Zhao and S. Du, "Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Transactions on Geosci. Remote. Sens.* **54**, 4544–4554 (2016).
30. Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Transactions on Neural Networks Learn. Syst.* **27**, 1279–1289 (2016).
31. Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Transactions on Geosci. Remote. Sens.* **56**, 847–858 (2018).
32. E. Aptoula, M. C. Ozdemir, and B. Yanikoglu, "Deep learning with attribute profiles for hyperspectral image classification," *IEEE Geosci. Remote. Sens. Lett.* **13**, 1970–1974 (2016).
33. W. Li, G. Wu, and Q. Du, "Transferred deep learning for hyperspectral target detection," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, (IEEE, 2017), pp. 5177–5180.
34. K. Yao, R. Unni, and Y. Zheng, "Intelligent nanophotonics: merging photonics and artificial intelligence at the nanoscale," <https://arxiv.org/abs/1810.11709> (2018).
35. M. A. Nielsen, *Neural networks and deep learning* (Determination Press, 2015).
36. I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning* (MIT Press, 2016).
37. P. R. Wiecha, A. Lecestre, N. Mallet, and G. Larrieu, "Pushing the limits of optical information storage using deep learning," *Nat. Nanotechnol.* **14**, 237 (2019).
38. L. Yunzhe, X. Yujia, and T. Lei, "Deep speckle correlation: A deep learning approach toward scalable imaging through scattering media," *Optica* **5**, 1181–11819 (2018).
39. G. Satat, M. Tancik, O. Gupta, B. Heshmat, and R. Raskar, "Object classification through scattering media with deep learning on time resolved measurement," *Opt. Express* **25**, 17466–17479 (2017).
40. E. Valent and Y. Silberberg, "Scatterer recognition via analysis of speckle patterns," *Optica* **5**, 204–207 (2018).
41. R. Horisaki, R. Takagi, and J. Tanida, "Learning-based imaging through scattering media," *Opt. Express* **24**, 13738–13743 (2016).
42. B. Rahmani, D. Loterie, G. Konstantinou, D. Psaltis, and C. Moser, "Multimode optical fiber transmission with a deep learning network," *Light. Sci. & Appl.* **7**, 69 (2018).
43. N. Borhani, E. Kakkava, C. Moser, and D. Psaltis, "Learning to see through multimode fibers," *Optica* **5**, 960–966 (2018).
44. O. Moran, P. Caramazza, D. Faccio, and R. Murray-Smith, "Deep, complex, invertible networks for inversion of transmission effects in multimode optical fibres," in *Proceedings of the 32Nd International Conference on Neural Information Processing Systems*, (Curran Associates Inc., USA, 2018), NIPS'18, pp. 3284–3295.
45. A. Agrawal, R. Verschueren, S. Diamond, and S. Boyd, "A rewriting system for convex optimization problems," *J. Control. Decis.* **5**, 42–60 (2018).
46. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine Learning on heterogeneous systems," <https://www.tensorflow.org/> (2015).
47. K. Ovtcharov, O. Ruwase, J.-Y. Kim, J. Fowers, K. Strauss, and E. Chung, "Accelerating deep convolutional neural networks using specialized hardware," *Microsoft Res.* (2015).