Convolutional long short-term memory neural network equalizer for nonlinear Fourier transform-based optical transmission systems

OLEKSANDR KOTLYAR,[1,2,4] **MORTEZA KAMALIAN-KOPAE,**[1,5]
MARYNA PANKRATOVA,[1] **ANASTASIIA VASYLCHENKOVA,**[1,3]
JAROSLAW E. PRILEPSKY,[1] **AND SERGEI K. TURITSYN**[1]

[1]*Aston Institute of Photonic Technologies, Aston University, Birmingham, B4 7ET, UK*
[2]*School of Science and Technology, Örebro University, Örebro, 70182, Sweden*
[3]*Optical Networks Group, University College London, London, WC1E 6BT, UK*
[4]*oleksandr.kotlyar@oru.se*
[5]*m.kamalian-kopae1@aston.ac.uk*

Abstract: We evaluate improvement in the performance of the optical transmission systems operating with the continuous nonlinear Fourier spectrum by the artificial neural network equalisers installed at the receiver end. We propose here a novel equaliser designs based on bidirectional long short-term memory (BLSTM) gated recurrent neural network and compare their performance with the equaliser based on several fully connected layers. The proposed approach accounts for the correlations between different nonlinear spectral components. The application of BLSTM equaliser leads to a 16x improvement in terms of bit-error rate (BER) compared to the non-equalised case. The proposed equaliser makes it possible to reach the data rate of 170 Gbit/s for one polarisation conventional nonlinear Fourier transform (NFT) based system at 1000 km distance. We show that our new BLSTM equalisers significantly outperform the previously proposed scheme based on a feed-forward fully connected neural network. Moreover, we demonstrate that by adding a 1D convolutional layer for the data pre-processing before BLSTM recurrent layers, we can further enhance the performance of the BLSTM equaliser, reaching 23x BER improvement for the 170 Gbit/s system over 1000 km, staying below the 7% forward error correction hard decision threshold (HD-FEC).

1. Introduction

Constantly increasing demand for capacity of optical communication systems incites active studies to reach the data rates higher than those provided by the current-generation systems [1–3]. There are various factors limiting the performance of modern optical communication systems. In this work we mostly focus on the mitigation of optical signal distortions by fibre nonlinearity, however, our approach can be expanded for compensation of other impairments. The nonlinearity of the optical fibre is often considered to be one of the most challenging factor degrading the performance of optical long-haul lines due to the nonlinear cross-talk between different frequency components [1]. The nonlinearity compensation and mitigation methods have been (and continue to be) the active research area over the last decade [3,4]. Among the multiple alternative techniques, the nonlinear Fourier transform (NFT) signal processing methods and the transmission techniques dealing with nonlinear Fourier (NF) modes, have recently attracted a great deal of attention [5–7]. The main benefit of using the NFT-based processing in optical communication is the effective linearisation of the signal evolution within the idealised single-mode fibre model, where the propagation of light along the latter is assumed

to be well-approximated by the lossless nonlinear Schrödinger equation (NLS). The linear decoupled evolution of nonlinear modes inside the NF domain allows us to utilise the efficient solutions and modulation methods developed for linear communication, in order to improve system's performance [8]. Unfortunately, the properties of optical fibre in realistic systems deviate from the description rendered by the idealised model that matches exactly the NFT processing operations. In particular, the NLS with the account of non-uniform gain-loss profile and noise is no longer exactly solvable by the NFT approach. This leads to the channel and processing mismatch, resulting in eventual system's performance degradation [9–11]. We note that the properties of noise that gets projected onto the NF domain and becomes different compared to its time-domain progenitor, are rather nontrivial [9,12]. Therefore, it is interesting to examine in the NFT-based systems new designs of additional equaliser that can mitigate the data corruptions due to deviations from the ideal integrable models. This work investigates the efficiency of such equalisers utilising artificial neural networks (NN) of two types: the feed-forward and bidirectional long short-term memory (BLSTM) gated recurrent NN approaches which are used to deal with the noise and NFT processing impairments affecting the quality of data transition inside the NF domain.

Machine learning (ML) methods are nowadays recognised as a promising tool for the mitigation of a variety of signal distortions in fibre-optic communications [13–15], including NFT-based systems [16–21]. In our previous works, we applied several ML techniques in NFT-based optical transmission systems for received constellation points labelling using classification [22–24] and regression [24,25], with the goal of system's performance improvement. In particular, the ML methods for classification have been used to distinguish the received symbols, effectively creating new nonlinear hard-decision boundaries [22,23]. We used several supervised methods such as feed-forward neural networks (FFNN), k-nearest neighbours, support vector machines, and the unsupervised k-means clustering method. However, the classification approach is failing in the case when we have the intersection of clouds in the received constellation [24]. For this case, we have already shown that, in order to attain the higher performance gain, we can employ the FFNN-based equaliser *directly to the received nonlinear spectrum* (NS) after dispersion compensation: such a technique renders much better equalisation results [24]. It was demonstrated that the NF spectral equaliser can be used for the de-noising of received NS, giving the performance improvement of almost one order of magnitude in terms of BER. One of the benefits of that equaliser type was its relative training simplicity. However, for the high-rate systems with very poor performance metrics (e.g. when the noise significantly corrupted the initial data), the FFNN-based equaliser [24] was not able to decrease the BER below the desired 7% HD-FEC threshold 3.8×10^{-3} [26]. At the same time, we note that the obvious disadvantage of FFNN for the sequential data is that it does not share the features obtained along the data processing. Nonetheless, when using the recurrent neural networks (RNN), the outputs from neighbouring taps are used as the input to the network processing a given tap, thus adding "a state" or "a memory" to the network. The latter feature allows the processing devise to reveal a complicated correlated behaviour of the input symbol sequences [27]. This capability of the RNN is, in particular, useful in dealing with the stochastic memory ensuing from the impact of the amplified spontaneous emission (ASE) noise projected onto the NF domain [9,12,28] and (possibly) coupling with other channel impairments [11]. We also note that, due to their capability of handling memory effects, various RNN based techniques have recently become a subject of high interest for the nonlinearity compensation in different optical fibre systems [29,30].

Here, we extend the approach presented in [24], by employing the NN directly to the continuous NS (carrying our information) at the receiver of NFT-based system, see Fig. 1 for the schematic of the transmission system. Similar to Ref. [24], we use several NS taps for the equlisation; in some respect, it is related to the approach used in Ref. [19], where, however, the components of

continuous NS part were used to help retrieving the discrete NS components. In this work we adopt the BLSTM gated RNN for the received NS equalisation. Note that the BLSTM-based equalisers have recently been successfully applied for impairments compensation of "conventional" optical transmission systems [31,32], and here we extend this approach to the NFT systems. The BLSTM RNN accounts for the correlations in the nonlinear spectrum. At the same time, the utilisation of convolutional neural network (CNN) can help extracting the data mapped onto the NS of the signal. The reason for the latter is that our data carrying symbols are mapped on the Fourier components of the NS [see Eq. (4) below] which needs to be retrieved by using the appropriate matched filtering. Therefore, in order to improve the system's performance, we also propose a more advanced equaliser based on the hybrid NN that is made up of a stack of 1D convolutional and BLSTM layers. The performance of FFNN and BLSTM-based equalisers are compared in terms of achieved BER for different signal powers. We also anticipate that this performance improving method can be applicable for the equalisation of a wide range of systems dealing with continuous NS modulation.
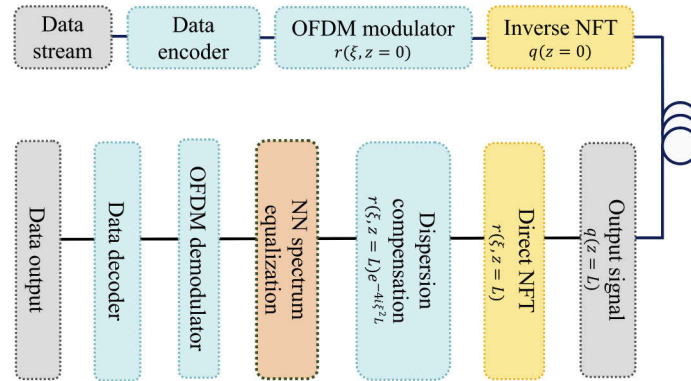
Fig. 1. The schematic of the NFT optical communication system with NN-based equaliser represented by the peach coloured block. For more detailed description see, e.g., Ref. [7, Fig. 2].

2. NFT-based optical communication systems

The propagation of the slow-varying one polarisation envelope $q(z, t)$ of the electromagnetic field along the optical fibre in the path-averaged approximation (see e.g. [10]) with the account of distributed optical noise is described by the perturbed NLS [1,33]:

$$iq_z - \frac{\beta_2}{2}q_{tt} + \gamma q|q|^2 = \eta(z,t), \tag{1}$$

here z is the distance along the fibre, t represents the retarded time in the frame co-moving with the envelope, and η describes the ASE noise. We consider the so-called focusing type of the NLS referring to anomalous dispersion with $\beta_2 < 0$, related to the standard single-mode fibre with $\beta_2 \approx -22$ ps²/km at the attenuation minimum [33]. The coefficient γ characterises Kerr nonlinearity, with the typical value $\gamma = 1.27$ W/km. The noise term $\eta(z, t)$ represents the impact of the ASE, modelled as an additive white Gaussian noise process with zero mean and $\langle\eta(z,t)\eta^*(z',t')\rangle = \Gamma\delta(z - z')\delta(t - t')$ ($\langle\ldots\rangle$ stands for averaging), with the typical power spectral density $\Gamma = 6.6 \cdot 10^{-21}$ W/GHz/km. The NFT operations are traditionally applied for the normalised signals, so for the intermediate computations we use the common normalisation

scheme

$$\frac{t}{T_0} \to t, \quad \frac{z}{Z_0} \to z, \quad \frac{q}{Q_0} \to q, \quad \text{where} \quad Z_0 = \frac{|\beta_2|}{2T_0^2}, \quad Q_0 = \sqrt{\frac{\gamma T_0^2}{|\beta_2|}},$$

and set the $T_0 = 0.9$ ns, thus arriving at the normalised NLS. Note that the values of parameters (γ, in particular) are effectively altered if we consider the approximate path-averaged NLS for lumped amplification schemes [10] or non-uniform Raman amplification [34].

The main idea of the NFT is that the signal can be unequivocally represented by the nonlinear spectrum components that evolve in the trivial and decoupled manner. To obtain the NS corresponding to a signal one needs to perform the forward NFT operation via solving the so-called Zakharov-Shabat system of ordinary linear differential equations [35] defined for the pair of auxiliary functions $\phi_{1,2}(t)$ [35]:

$$\frac{d\phi_1}{dt} = q(t)\,\phi_2 - i\xi\phi_1\,, \qquad \frac{d\phi_2}{dt} = -q^*(t)\,\phi_1 + i\xi\phi_2\,. \tag{2}$$

Here ξ is a (generally complex) spectral parameter, the nonlinear analogue of conventional Fourier frequency, and $q(t)$ is the signal to decompose (we omitted the explicit dependence on distance z); the asterisk means the complex conjugation. In the NFT-based transmission methods, the information-bearing signal is surrounded by the zero-padding guard intervals usually taken equal to the linear dispersion-induced memory [8]. So we assume that our signals are of a finite support and we operate in a *burst mode* [6–8]. In this paper we employ only the continuous NF spectrum similar to the systems considered in Refs. [8,9,12,34,36,37], and no discrete spectrum (soliton modes) is present, so that the parameter ξ is assumed real. For the forward NFT operation, we define a special solution of Eq. (2), $\Phi(t,\xi) = [\phi_1, \phi_2]^T$, with the "initial" condition at the trailing end of the pulse (assumed here to be at $t = 0$): $\Phi|_{t=0} = [1,0]^T$. Then, the solution at the leading end of the burst at $t = T$ must necessarily take the form: $\Phi_{t=T} = [a(\xi)e^{-i\xi T}, b(\xi)e^{i\xi T}]^T$, where the functions $a(\xi)$ and $b(\xi)$ are called *scattering coefficients* and constitute the core of NFT-based pulse decomposition. The continuous part of the nonlinear spectrum is defined by the ratio of these scattering coefficients:

$$r(\xi) = \ b(\xi)/a(\xi),$$

widely known as the reflection coefficient; we shall refer to it simply as NS. The expression for $r(\xi, L)$ at the receiver placed at distance L is quite simple:

$$r(\xi, z = L) = e^{4i\xi^2 L} r(\xi, z = 0). \tag{3}$$

Within our NFT-based transmission system, we modulate input data using the parameters of the NS $r(\xi)$. The inverse NFT stage is used to produce the time-domain waveform at Tx that is then launched into the fibre; we do not present the mathematical of inverse NFT details here, see more in Ref. [6]. In our work we use fast NFT and INFT algorithms developed by Wahls and collaborators [38]. We note that, in general, it is possible to use both eigenvalues and continuous spectrum as data carriers [7,39], but we do not address such systems here.

In this work, we use only the continuous part of the nonlinear spectrum as a data carrier, modulating $r(\xi)$ directly, as it was done in, e.g., Refs. [8,9,11,12,36,37], including the experimental demonstrations of the approach's feasibility [7,40–42]. We choose the Fourier image of the exponential-based orthogonal frequency division multiplexing (OFDM) modulation format [43]

with $N_{\text{sc}} = 512$ sub-carriers in the nonlinear continuous spectrum:

$$r(\xi, z = 0) = \sum_{k=-N_{\text{sc}}/2}^{N_{\text{sc}}/2} c_k \, \text{sinc}(2\xi - k), \tag{4}$$

where the information-bearing coefficients c_k are picked from the QAM alphabet. The transmission is implemented in a burst mode, where the full temporal support of a single symbol (burst) is 18 ns. Then, the achievable data rate for our system is 170 Gbit/s for 64-QAM, accounting the dispersion-broadening memory included into the overall burst length. This additional guard interval for preventing inter-symbol interference is chosen based on the dispersion-induced signal broadening according to the relation

$$T_g \approx 2\pi B L |\beta_2| \tag{5}$$

where B is the total signal bandwidth. A 1000 km experiment where the fibre is numerically simulated using split-step Fourier transform has been carried out. The ASE noise is added along the link in a distributed way. The simulation bandwidth is 1.4 THz.

Here we notice that in our current research we do not account for the effects emerging from realistic distributed or lumped amplification, because in our proof-of-concept study we aim at addressing the question: how the advanced NN-based equalisation would help with the main degrading factors affecting the NFT transmission performance, in particular, the interplay of noise and dispersion [44,45], and the corruptions emerging due to the finite accuracy of NFT processing [12,28,46]. Evidently, all these effects are present if we address the simplified model based on Eq. (1) with zero gain-loss profile, and use some standard NFT processing methods. At the same time, we argue that the validity of the equalisation methods described below will stay in force if we address more realistic systems with the non-zero gain-loss budget, or if we account for some other (say, component-induced) impairments. But for a more realistic scenario the retraining of the NNs using the corrupted nonlinear spectra (or even some moderate modifications of the NNs structure) can be important for our getting good improvement numbers.

3. Feed-forward neural network-based equaliser

Prior to applying the FFNN equaliser (Fig. 2), we estimate the performance of considered NFT-based communication system without the NN equalization. To detect received symbols, the complex plane of received c_k from Eq. (4) is divided into a square grid, and the symbols located in the wrong cell are identified as errors, incrementing BER's numerator. This detection scheme is referred to as "without equalisation" further in the text to reflect the absence of machine learning-based processing. In addition, we apply a rotation to the received constellation, meaning the uniform rotation of the whole constellation combined with our "without equalisation" scheme. The scheme is called "with constellation rotation" in the text. It can be seen from Fig. 3(a), the purple curve, that through "unequlised" detection scheme we obtain the BER values that are much higher than the HD-FEC threshold for the signal powers around the optimal value. The rotation of the constellation as a whole results in only minor improvement of the system performance: the cyan line in Fig. 3(a).

Now we turn to evaluating the transmission improvements provided by the NN-based equalisers. We start from analysis of a fully connected FFNN equaliser, the structure of which is similar to the previously considered one in Ref. [24], but now we also address the scenario when we have a higher number of nodes and layers. The schematic of the FFNN equaliser is given in Fig. 2. In particular, we employ the FFNN with 4 hidden layers with 96 nodes in each hidden layer or more advanced structure having 5 hidden layers with 256 nodes in each hidden layer. Then we compare the performance of the FFNN-based equalisers having a different number of hidden layers and nodes in them.
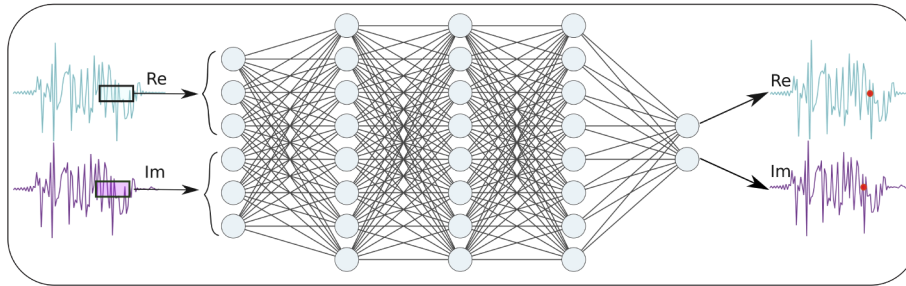
Fig. 2. The schematic of the feed-forward NN -based equaliser. In our work we present the results for the fully connected FFNN equalisers with 4 and 5 hidden layers, with 96 and 256 nodes in each layer, correspondingly.

(a)   (b)
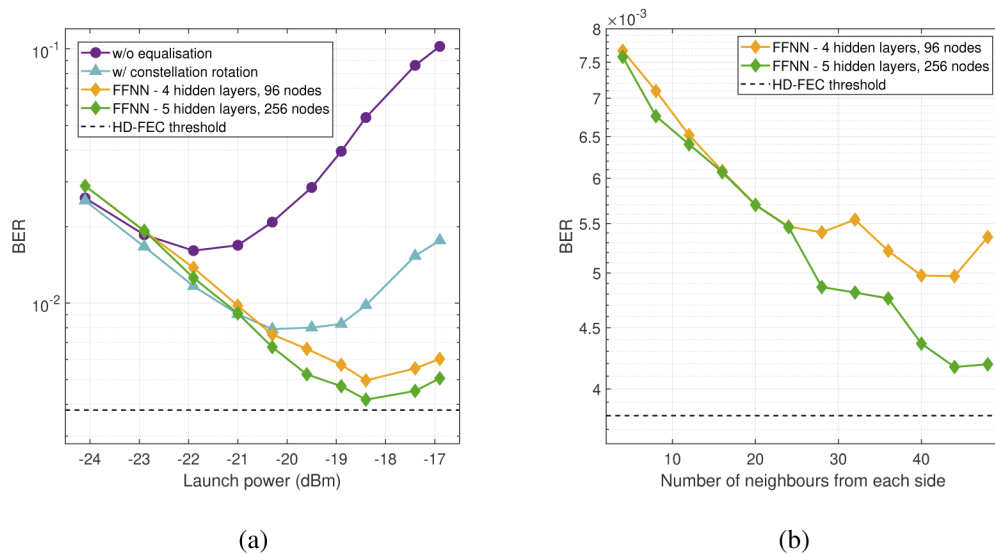
Fig. 3. (a) BER vs power w/o equalisation (the purple curve), with constellation rotation as a whole (cyan curve), and applying the FFNNs with different structure for the equalisation of the received spectrum, the yellow (96 nodes), and green (256 nodes) curves. The number of taps from each side of the NS sample of interest used for the FFNN-based equalisation is 44 for both curves. (b) BER as a function of the number of neighbouring taps for the NN equalisation for two network configurations at the optimal power (-18.4 dBm). The data rate is 170 Gbit/s, propagation distance 1000 km. The horizontal black dashed line represents the 7% HD-FEC threshold.

When dealing with the NN equalisers, we use the samples picked from the received continuous NS as an input for the NN. During the NN model training stage, the received NS is mapped onto the transmitted NS. Next, the trained FFNN model is used for the equalisation of the received NS for the test case. The distributed optical noise and the numerical processing noise emerging from NFT processing [12] both contribute to effective cross-talk and lead to the emerging correlations between the nonlinear spectrum components [9,12]. An important part of our approach is that this correlation between the nonlinear spectrum components is taken into account, similarly to [24] where we performed a multi-tap equalisation, see the left pane of Fig. 2. It is worth mentioning that this memory is not the channel memory which is taken care of by inserting temporal guard intervals between consecutive NFT "super-symbols" (each containing hundreds of QAM symbols), but the effective memory resulting from the NFT processing-induced impairments (i.e. the finite accuracy of the numerical NFT operation) and the mismatch between the true channel and NFT processing due to the presence of noise. When we equalise a sample from the spectrum, we simultaneously process up to 49 nearest elements (NS taps) at each side of the NS sample of interest. The number of nodes in the input layer is $2(N_{\mathrm{neigh}} + 1)$, where N_{neigh} is the number of taps that we take into account.

In this paper, for the NN implementation we employ the end-to-end open-source machine learning platform Tensorflow [47] with Keras [48] application programming interface. The FFNN was trained using Nesterov-accelerated adaptive moment estimation (NADAM) [49] optimisation algorithm. We use the leaky rectified linear unit (leaky ReLU) activation function because, as our tests showed, it provides a better performance than the ordinary ReLU, exponential linear unit (ELU) [50] and hyperbolic tangent activation functions. The real and imaginary parts of each spectrum point at the input (output) are considered as separate nodes, see the left (right) ends of Fig. 2. To create the input feature vector to the FFNN, we take the received NS sample of interest along with its N_{neigh} neighbours from both sides. Thus, the input feature vector is represented as the overlapping sub-sequences of the elements of received NS. Such sub-sequences created from real and imaginary parts of the NS are concatenated to be fed into the FFNN. The feature vectors from different noise and data realisations of the training sets are interleaved and shuffled before the training. During the training the input vectors are considered to be independent. Such a way of preparing the sequential feature vectors can be called a "sliding window with mixing". The training set was generated by 8×10^3 realisations of in-line noise (it is the number of independent runs), and another 1.2×10^3 noise realisations were used for our making up a validation set. Such numbers of independent runs used to generate the training (validation) sets provide us with 4.7×10^6 (0.7×10^6) pairs of target and input vectors for the training (validation). For evaluating the performance of equalisers, in the paper we use the distinct test set that consists of 31×10^3 in-line noise realisations (giving us 18×10^6 input-output pairs). The training is performed separately for each signal power.

The resulting BER vs power dependencies for the FFNN constructed from either 4 hidden layers each with 96 nodes or 5 hidden layers each with 256 nodes, are shown in Fig. 3(a) by yellow and green curves, respectively. It can be seen that the FFNN with 5 hidden layers outperforms the simpler NN. At the optimal power, the largest FFNN renders the BER value 4.2×10^{-3}, which is close to the HD-FEC threshold, but still above it. The dependence of system performance as a function of the number of processing taps that we take for the equalisation, is shown in Fig. 3(b). It can be seen from the figure that the performance is increasing with the growth of the number of neighbours up to a certain value that corresponds to the effective NS correlation length.

We observed that the further increase in the number of layers and neurons in hidden layers up to ~ 500 did not provide any noticeable performance improvement. We also checked the possible effect of introducing a big difference in the number of neurons in the hidden layers. We considered the FFNN structure that consists of six hidden layers with a gradually decreasing number of neurons: three layers of 512, followed by layers with 256, 128, and 64 neurons. This

FFNN model also did not provide any considerable BER improvement, giving the BER value at the optimal power point 4.4×10^{-3}.

To estimate the computational complexity of the NN equaliser we can compute the number of multiplications required for processing of one element (tap) of a nonlinear spectra. For the equalisers presented in the paper equalisation of one tap corresponds to a processing of a sequence constructed of equalising tap and its neighbours from both sides. Concretely, for FFNN with N hidden layers number of multiplications can be computed as $\sum_{i=0}^{N} n_i n_{i+1}$, where n_i is the number of nodes in the layer i, n_0 and n_{N+1} correspond to the input layer and output layers respectively. It is seen that complexity of FFNN-based eqalizer is proportional to the number of weights. Processing of one element of nonlinear spectrum by FFNN-based equaliser consisting of 5 hidden layers and 256 nodes in each layers and taking into account 44 taps from both sides, that showed the best performance [green lines in Fig. 3(b)], requires about 3.8×10^5 multiplications.

We conclude that the considered FFNN-based equaliser can provide decent improvement in the performance, though, this scheme is not optimal for the considered communication system. Now we set the stage for the main point of the paper: the application of the RNN-based equalisation to the received NS.

## 4. Equaliser is based on long short-term memory gated recurrent neural network

The FFNN considered in the previous section can be regarded as the approximation of a nonlinear function defining the mapping from the received spectra to the transmitted one. The nonlinear spectral sample of interest and its neighbours are used as inputs of the FFNN, meaning that the particular interrelations between the NS components are not taking into account. The result of the FFNN equalisation depends on the values of the input (the received NS samples) and is determined only by the FFNN weights values. Moreover, we can only feed the real and imaginary parts of the NS to the FFNN as a stack, i.e. during current tap processing the FFNN is not taking into account the result of processing of the previous (or next) taps in the input sample.

Therefore, we apply now the bidirectional RNNs that render us the opportunity to account the result of processing previous (next) tap while equalising the current tap when learning the mapping function from the received to transmitted NS. Moreover, the RNNs are able to operate with multivariate data. This allows us to feed to the RNN-equaliser with the real and imaginary parts of NS jointly.

In particular, the long short-term memory (LSTM) [51] is a special type of RNN that is capable of learning long-term sequential dependencies and overcoming vanishing gradient or exploding gradient problems that are inherent to ordinary RNNs [52]. In this section, we consider two new equaliser designs based on the LSTM RNNs. We employ the LSTM realisation from Keras that utilises a forget gate [53]. In order to take advantage of the fast cuDNN implementation for RNN training we left default *tanh* and *sigmoid* as non-gate and recurrent (gate) activation functions, respectively. The schematic of the new equalisers is given in Fig. 4. As in the previous section, we fed the received NS to the equaliser, taking into account several neighbours from both sides of the NS element of interest. But, unlike the case of the FFNN equaliser, we fed the imaginary and real parts of the spectra simultaneously as a multivariate sequence with two variables. The main part of the equaliser is the gated RNN stacked of two bidirectional LSTM (BLSTM) layers with 96 nodes in each layer. Here for the BLSTM training we employ, as same as in previous section, the NADAM optimisation algorithm to minimize the loss. During the spectra equalisation, the RNN can extract the sequential features from one tap to the next (and to the previous in the forward direction of BLSTM). Therefore, we expected that such an equaliser can outperform the FFNN operating with a stack of data without understanding the interrelation or ordering among the data points. Then, with the aim to improve the BLSTM equaliser performance, we add a 1D convolutional layer (the optional 1D convolutional layer in Fig. 4) before the BLSTM layers.

The 1D CNN works as a filter to extract features from the NS. Actually, the hybrid NN obtained by combination of CNN and LSTM are well-known to work efficiently for, e.g., the problems of named entity recognition in language processing [54], visual recognition and description [55], and rare sound event detection [56]. Such that we naturally expected to have a better performance as a result of our adding the 1D CNN. We utilise only one 1D convolutional layer consisting of 64 filters, without using any nonlinear activation functions and max-pooling layers. The dimensionality of input multivariate sequence to BLSTM changes from 2 (corresponding to the real and imaginary parts of the NS sample) to 64 (the number of convolutional filters). The output layer is a fully connected feed-forward layer with two nodes for the real and imaginary parts of an equalised NS sample. As the output of the network we have an equalised NS sample, the same as we had in the previous section.
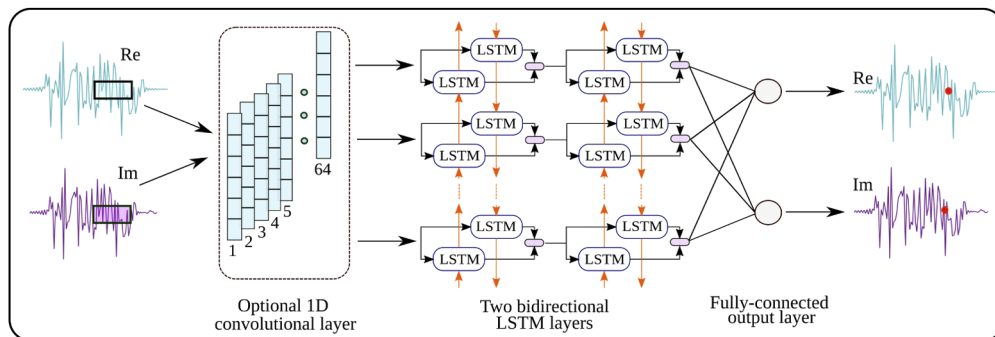

Fig. 4. The schematic of the spectra equaliser implemented as a 1D CNN layer combined with the BLSTM. As the inputs to the NN we take the received NS samples using a sliding window. The output of the NN is an equalised NS element.

The performance evaluation of the proposed BLSTM equalisers, i.e. the effective BER after the equalisation vs the input power, is shown in Fig. 4. For the case of the "pure" BLSTM RNN without the CNN, the respective BER vs power dependence for the equalised system is presented in Fig. 5(a) by the red curve. It can be seen that the equaliser reduces the BER value below the desirable 7% HD-FEC level for optimal power values. However, this is achieved only for a narrow range of optimal powers.

In order to decrease BER further, we added a 1D convolutional layer combining it with the BLSTM layers, the respective BER vs power is shown by a blue curve in Fig. 5(a). As expected, the convolution layer allows improving the BER value further expanding acceptable power window in the vicinity of the optimal point. The above-discussed curves in Fig. 5(a) were obtained by utilizing of 16 neighbours from both sides, the number which allowed us to train the resulting NNs fast enough.

In Fig. 5(b) we demonstrate that by increasing the number of neighbours for the BLSTM and 1D CNN - BLSTM equalisers we can reduce the obtained BER values. However, unfortunately, feeding the BLSTM with the sequences longer than 33 samples (16 neighbouring taps) resulted in unstable model training. For the bigger number of taps the training behaviour became irregular, so we leave the question of finding the minimum in the dependencies shown in Fig. 5(b) for some future studies. We note here that the loss landscape for NNs is often non-convex with multiple local minima and maxima [57], and the model tends to get stuck into some local minimum during the training. We reached the training minimum for 16 neighbours relatively quickly, while the same procedure for 20 neighbours (the last points on both curves in Fig. 5) required much more time. Thus, we used 16 neighbouring taps for the BLSTM and CNN - BLSTM equalisation, the results of which is presented in Fig. 5(a), to avoid lengthy computations.

(a)

(b)
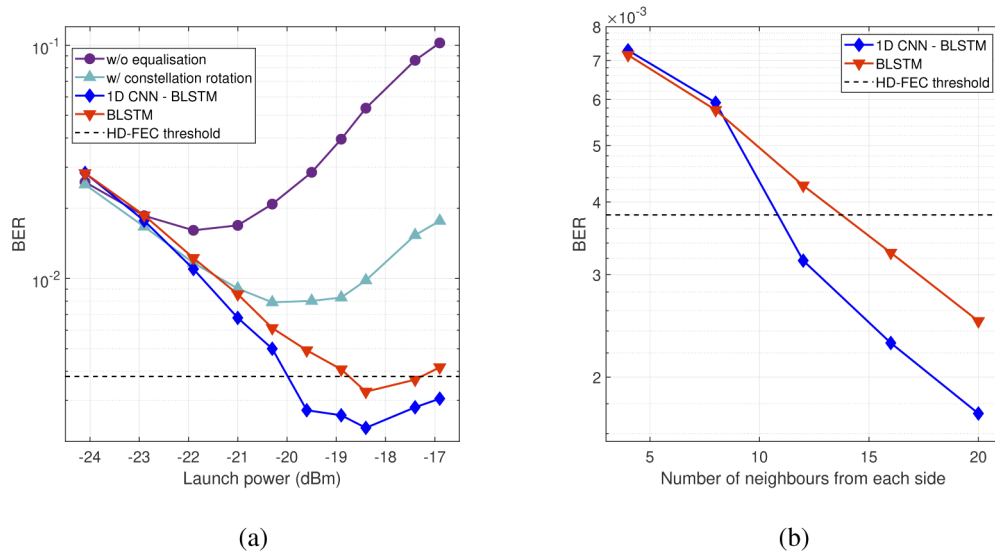
Fig. 5. (a) BER vs power, without equalisation (purple curve), with constellation rotation as a whole (cyan curve), with BLSTM RNN equalisation (red curve) and 1D CNN - BLSTM RNN equalisation (blue curve). Number of considered neighbours from each side of the NS element during equalisation is 16. (b) BER as a function of the number of neighbours (taps) used in the NN equalisation for the optimal power (-18.4 dBm). The data rate is 170 Gbit/s, propagation distance 1000 km. Horizontal black dashed line indicates the 7% HD FEC threshold.

In Fig. 6 we present the resulting constellations of the received symbols before and after the described NN-based equalisation. The transmitted symbols are marked with red, and the received symbols are blue. The received symbols are shown in Fig. 6(a) while the equalised symbols using the FFNN are depicted in Fig. 6(b). The equalised symbols as the result of applying BLSTM and CNN - BLSTM are shown in Fig. 6(c) and Fig. 6(d), respectively, which demonstrate a clear advantage of the CNN - BLSTM equaliser in restoring the transmitted symbols.

Finally, let us briefly turn to the processing complexity per sample by the BLSTM and 1D CNN - BLSTM architectures. The number of multiplications required for the processing of the sequence of the length l and dimensionality n_i by standard bidirectional LSTM layer [58] is $2l[4n_c^2 + 4n_in_c + 3n_c]$, where n_c is the number of LSTM memory cells and the number of multiplications is proportional to the number of weights [51]. The complexity of one 1D convolutional layer per sequence of length l and dimensionality d (that is equal to 2 in our case) in the case of no padding can be estimated as $kdf(l - (k - 1))$, where k is the size of the filter and f is the number of filters. The total complexity of the equalizer is a sum of multiplications related to all layers including the output fully-connected layer. The number of multiplications required for processing of the sample length accounting for 16 neighbouring taps from both sides [presented by blue and red curves at Fig. 5(a)] by above described BLSTM and 1D CNN - BLSTM equalisers is about 9.8×10^6 and 10.4×10^6 respectively. Therefore, while BLSTM and 1D CNN - BLSTM equalisers provide better performance in terms of BER, they have higher computational complexity comparing to FFNN-based equaliser that should be taken into account while choosing appropriate model.
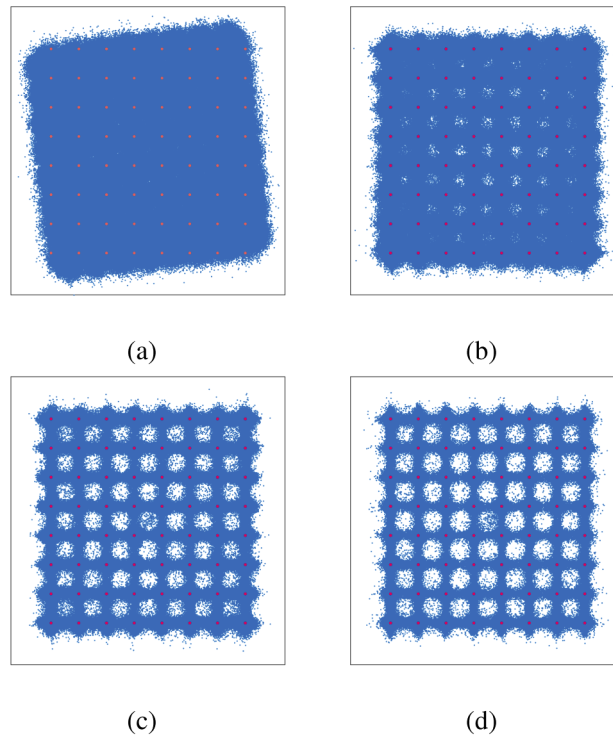
Fig. 6. Received constellation symbols a) without equalisation, b) after the FFNN-based equaliser consisting of four hidden layers with 96 nodes in each layer, (c) after the BLSTM equaliser with two bidirectional LSTM hidden layers with 96 nodes in each, and (d) after the 1D CNN - BLSTM hybrid NN equaliser. All constellations are at the launch power of -18.4 dBm for 170 Gbit/s NFT-based transmission.

5. Conclusion

To sum up, in this paper we addressed the question of how we can improve the performance of the NFT-based optical transmission systems and make them more robust in the practical environment. The signal transmission using NFT-based systems is degraded due to a mismatch between the real channel (first of all, ASE noise) and the NFT processing operations that assume the ideal integrable model. With this, near the optimum point, the performance of NFT systems is also heavily affected by the finite accuracy of the NFT processing operations themselves, resulting in the eventual decrease of the quality metrics as functions of power. In this work we designed two novel NN-based equalisers (and examined in detail the performance of a previously proposed NN architecture) that improve the quality of the NFT-based transmission systems based on the modulation of the continuous part of NS. The equalisers were applied to the received NS. Several taps of the NS have been processed jointly to account for the memory effects. We compared the performance improvement rendered by the FFNN, BLSTM, and 1D CNN - BLSTM equalisers. Our findings demonstrate that the proposed new type of equalisers based on the RNNs can noticeably outperform the FFNN equaliser. We achieved a 16x BER performance improvement with the equaliser based on the RNNs with two bidirectional LSTM layers and a 23x BER improvement with the equaliser based on hybrid 1D CNN - BLSTM for a 170 Gbit/s NFT communication system, up to 1000 km propagation distance. However, we also observed some challenges in training the complicated NN structures involving the BLSTM part, such that we were not able to reach the minimum in the BER as a function of neighbouring taps. Therefore,

we believe that our approach has the potential to deliver even more improvement when combined with additional regularisations and optimisations.

Disclosures. The authors declare no conflicts of interest.

References

1. R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," J. Lightwave Technol. **28**(4), 662–701 (2010).
2. E. Agrell, M. Karlsson, A. Chraplyvy, D. J. Richardson, P. M. Krummrich, P. Winzer, K. Roberts, J. K. Fischer, S. J. Savory, B. J. Eggleton, and M. Secondini, "Roadmap of optical communications," J. Opt. **18**(6), 063002 (2016).
3. P. J. Winzer, D. T. Neilson, and A. R. Chraplyvy, "Fiber-optic transmission and networking: the previous 20 and the next 20 years," Opt. Express **26**(18), 24190–24239 (2018).
4. J. C. Cartledge, F. P. Guiomar, F. R. Kschischang, G. Liga, and M. P. Yankov, "Digital signal processing for fiber nonlinearities," Opt. Express **25**(3), 1916–1936 (2017).
5. M. Yousefi and F. Kschischang, "Information transmission using the nonlinear Fourier transform, Parts I–III," IEEE Trans. Inf. Theory **60**(7), 4312–4328 (2014).
6. S. K. Turitsyn, J. E. Prilepsky, S. T. Le, S. Wahls, L. L. Frumin, M. Kamalian, and S. A. Derevyanko, "Nonlinear Fourier transform for optical data processing and transmission: advances and perspectives," Optica **4**(3), 307 (2017).
7. S. T. Le, V. Aref, and H. Buelow, "Nonlinear signal multiplexing for communication beyond the kerr nonlinearity limit," Nat. Photonics **11**(9), 570–576 (2017).
8. S. T. Le, J. E. Prilepsky, and S. K. Turitsyn, "Nonlinear inverse synthesis for high spectral efficiency transmission in optical fibers," Opt. Express **22**(22), 26720 (2014).
9. S. A. Derevyanko, J. E. Prilepsky, and S. K. Turitsyn, "Capacity estimates for optical transmission based on the nonlinear Fourier transform," Nat. Commun. **7**(1), 12710 (2016).
10. M. Kamalian, J. E. Prilepsky, S. T. Le, and S. K. Turitsyn, "On the design of NFT-based communication systems with lumped amplification," J. Lightwave Technol. **35**(24), 5464–5472 (2017).
11. X. Yangzhang, D. Lavery, P. Bayvel, and M. I. Yousefi, "Impact of perturbations on nonlinear frequency-division multiplexing," J. Lightwave Technol. **36**(2), 485–494 (2018).
12. M. Pankratova, A. Vasylchenkova, S. A. Derevyanko, N. B. Chichkov, and J. E. Prilepsky, "Signal-noise interaction in optical-fiber communication systems employing nonlinear frequency-division multiplexing," Phys. Rev. Appl. **13**(5), 054021 (2020).
13. J. Thrane, J. Wass, M. Piels, J. C. M. Diniz, R. Jones, and D. Zibar, "Machine learning techniques for optical performance monitoring from directly detected PDM-QAM signals," J. Lightwave Technol. **35**(4), 868–875 (2017).
14. F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," IEEE Commun. Surv. Tutorials **21**(2), 1383–1408 (2019).
15. D. Zibar, F. Da Ros, G. Brajato, and U. C. de Moura, "Toward intelligence in photonic systems," Opt. Photonics News **31**(3), 34–41 (2020).
16. R. T. Jones, S. Gaiarin, M. P. Yankov, and D. Zibar, "Time-domain neural network receiver for nonlinear frequency division multiplexed systems," IEEE Photonics Technol. Lett. **30**(12), 1079–1082 (2018).
17. S. Gaiarin, F. Da Ros, N. De Renzis, E. P. da Silva, and D. Zibar, "Dual-polarization nfdm transmission using distributed raman amplification and nft-domain equalization," IEEE Photonics Technol. Lett. **30**(22), 1983–1986 (2018).
18. Y. Wu, L. Xi, X. Zhang, Z. Zheng, J. Wei, S. Du, W. Zhang, and X. Zhang, "Robust neural network receiver for multiple-eigenvalue modulated nonlinear frequency division multiplexing system," Opt. Express **28**(12), 18304–18316 (2020).
19. J. Koch, K. Chan, S. Kühl, C. G. Schaeffer, and S. Pachnicke, "Neural networks based equalization of experimental transmission using the nonlinear fourier transformation," in *46th European Conference on Optical Communication (ECOC 2020)*, (2020).
20. K. Mishina, S. Sato, S. Yamamoto, Y. Yoshida, D. Hisano, and A. Maruta, "Demodulation of eigenvalue modulated signal based on eigenvalue-domain neural network," in *Optical Fiber Communication Conference (OFC) 2020*, (Optical Society of America, 2020), p. W3D.1.
21. E. V. Sedov, I. S. Chekhovskoy, J. E. Prilepsky, and M. P. Fedoruk, "Application of neural networks to determine the discrete spectrum of the direct zakharov–shabat problem," Quantum Electron. **50**(12), 1105–1109 (2020).
22. O. Kotlyar, M. Kamalian, M. Pankratova, J. E. Prilepsky, and S. K. Turitsyn, "Machine learning for performance improvement of periodic NFT-based communication system," in *European Conference on Optical Communication (ECOC)*, (2019).
23. O. Kotlyar, M. Pankratova, M. Kamalian, A. Vasylchenkova, J. E. Prilepsky, and S. K. Turitsyn, "Unsupervised and supervised machine learning for performance improvement of NFT optical transmission," in *2018 IEEE British and Irish Conference on Optics and Photonics (BICOP)*, (2018).

24. O. Kotlyar, M. Pankratova, M. Kamalian-Kopae, A. Vasylchenkova, J. E. Prilepsky, and S. K. Turitsyn, "Combining nonlinear fourier transform and neural network-based processing in optical communications," Opt. Lett. **45**(13), 3462–3465 (2020).

25. M. Kamalian-Kopae, A. Vasylchenkova, O. Kotlyar, M. Pankratova, J. Prilepsky, and S. Turitsyn, "Artificial neural network-based equaliser in the nonlinear Fourier domain for fibre-optic communication applications," in *2019 Conference on Lasers and Electro-Optics Europe and European Quantum Electronics Conference (CLEO/Europe-EQEC)*, (IEEE, 2019).

26. E. Agrell and M. Secondini, "Information-theoretic tools for optical communications engineers," in *2018 IEEE Photonics Conference (IPC)*, (2018), pp. 1–5.

27. A. Maas, Q. V. Le, T. M. O'Neil, O. Vinyals, P. Nguyen, and A. Y. Ng, "Recurrent neural networks for noise reduction in robust asr," in INTERSPEECH, (2012).

28. S. Derevyanko, D. Shepelsky, M. Pankratova, A. Vasylchenkova, N. Chichkov, and J. Prilepsky, "Analytical model of nonlinear noise in the b-modulated optical transmission systems," in *CLEO: Science and Innovations*, (Optical Society of America, 2020), pp. SF2L–5.

29. M. Schaedler, F. Pittala, G. Böcherer, C. Bluemm, M. Kuschnerov, and S. Pachnicke, "Recurrent neural network soft-demapping for nonlinear isi in 800gbit/s dwdm coherent optical transmissions," in *46th European Conference on Optical Communication (ECOC 2020)*, (2020).

30. S. Deligiannidis, C. Mesaritakis, and A. Bogris, "Performance and complexity evaluation of recurrent neural network models for fibre nonlinear equalization in digital coherent systems," in *46th European Conference on Optical Communication (ECOC 2020)*, (2020).

31. B. Karanov, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-end optimized transmission over dispersive intensity-modulated channels using bidirectional recurrent neural networks," Opt. Express **27**(14), 19650–19663 (2019).

32. S. Deligiannidis, A. Bogris, C. Mesaritakis, and Y. Kopsinis, "Compensation of fiber nonlinearities in digital coherent systems leveraging long short-term memory neural networks," J. Lightwave Technol. **38**(21), 5991–5999 (2020).

33. G. P. Agrawal, *Fiber-optic communication systems*, vol. 222 (John Wiley & Sons, 2012).

34. S. T. Le, J. E. Prilepsky, P. Rosa, J. D. Ania-Castañón, and S. K. Turitsyn, "Nonlinear inverse synthesis for optical links with distributed raman amplification," J. Lightwave Technol. **34**(8), 1778–1786 (2016).

35. V. E. Zakharov and A. B. Shabat, "Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media," Soviet Physics JETP **34**, 62 (1972).

36. J. E. Prilepsky, S. A. Derevyanko, K. J. Blow, I. Gabitov, and S. K. Turitsyn, "Nonlinear inverse synthesis and eigenvalue division multiplexing in optical fiber channels," Phys. Rev. Lett. **113**(1), 013901 (2014).

37. M. Yousefi and X. Yangzhang, "Linear and nonlinear frequency-division multiplexing," IEEE Trans. Inf. Theory **66**(1), 478–495 (2020).

38. S. Wahls, S. Chimmalgi, and P. J. Prins, "FNFT: A software library for computing nonlinear fourier transforms," J. Open Source Softw. **3**(23), 597–598 (2018).

39. A. Vasylchenkova, J. E. Prylepsky, N. B. Chichkov, and S. K. Turitsyn, "Multieigenvalue communication paired with b-modulation," in *45th European Conference on Optical Communication (ECOC 2019)*, (2019), pp. 1–4.

40. S. T. Le, I. D. Philips, J. E. Prilepsky, P. Harper, A. D. Ellis, and S. K. Turitsyn, "Demonstration of nonlinear inverse synthesis transmission over transoceanic distances," J. Lightwave Technol. **34**(10), 2459–2466 (2016).

41. W. A. Gemechu, T. Gui, J. Goossens, M. Song, S. Wabnitz, H. Hafermann, A. P. T. Lau, M. I. Yousefi, and Y. Jaouën, "Dual polarization nonlinear frequency division multiplexing transmission," IEEE Photonics Technol. Lett. **30**(18), 1589–1592 (2018).

42. F. Da Ros, S. Civelli, S. Gaiarin, E. P. da Silva, N. De Renzis, M. Secondini, and D. Zibar, "Dual-polarization nfdm transmission with continuous and discrete spectral modulation," J. Lightwave Technol. **37**(10), 2335–2343 (2019).

43. W. Shieh, H. Bao, and Y. Tang, "Coherent optical ofdm: theory and design," Opt. Express **16**(2), 841–859 (2008).

44. S. Civelli, E. Forestieri, and M. Secondini, "Why noise and dispersion may seriously hamper nonlinear frequency-division multiplexing," IEEE Photonics Technol. Lett. **29**(16), 1332–1335 (2017).

45. S. Civelli, E. Forestieri, and M. Secondini, "Mitigating the impact of noise on nonlinear frequency division multiplexing," Appl. Sci. **10**(24), 9099 (2020).

46. S. Derevyanko, M. Balogun, O. Aluf, D. Shepelsky, and J. E. Prilepsky, "Channel model and the achievable information rates of the optical nonlinear frequency division-multiplexed systems employing continuous b-modulation," Opt. Express **29**(5), 6384–6406 (2021).

47. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "Tensorflow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, (2016), pp. 265–283.

48. F. Chollet, *et al.*, "Keras," https://keras.io (2015).

49. T. Dozat, "Incorporating Nesterov momentum into Adam," in *International Conference on Learning Representations (ICLR)*, (2016).

50. D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," arXiv preprint arXiv:1511.07289 2 (2016).

51. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput. **9**(8), 1735–1780 (1997).

52. R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *International conference on machine learning*, (2013), pp. 1310–1318.

53. F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," Neural Comput. **12**(10), 2451–2471 (2000).

54. J. P. Chiu and E. Nichols, "Named entity recognition with bidirectional lstm-cnns," Trans. Assoc. Comput. Linguist. **4**, 357–370 (2016).

55. J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2015), pp. 2625–2634.

56. H. Lim and J. Park, "Rare sound event detection using 1d convolutional recurrent neural networks," in *Detection and Classification of Acoustic Scenes and Events 2017*, (2017), pp. 80–84.

57. H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein, "Visualizing the loss landscape of neural nets," in *Advances in Neural Information Processing Systems*, vol. 31 (2018), pp. 6389–6399.

58. H. Sak, A. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," (2014).