

Article

# Global Receptive Field Designed Complex-Valued Convolutional Neural Network Equalizer for Optical Fiber Communications

Lu Han <sup>1</sup>, Yongjun Wang <sup>1,\*</sup>, Haifeng Yang <sup>1</sup>, Yang Zhao <sup>2</sup> and Chao Li <sup>3</sup>

<sup>1</sup> School of Electronic Engineering, Beijing University of Posts and Telecommunications (BUPT), Beijing 100876, China; clear@bupt.edu.cn (L.H.); hifeng@bupt.edu.cn (H.Y.)

<sup>2</sup> China Rocket Co., Ltd., No. 188, West Road, South 4th Ring, Beijing 100070, China; yzhao@chinarocket.cn

<sup>3</sup> Inspur Electronic Information Industry Co., Ltd., No. 15, Lingxiao Road, Beijing 100194, China; lichao22@ieisystem.com

\* Correspondence: wangyj@bupt.edu.cn

**Abstract:** In this paper, an improved complex-valued convolutional neural network (CvCNN) structure to be placed at the received side is proposed for nonlinearity compensation in a coherent optical system. This complex-valued global convolutional kernel-assisted convolutional neural network equalizer (CvGNN) has been verified in terms of Q-factor performance and complexity compared to seven other related nonlinear equalizers based on both the 64 QAM experimental platform and the QPSK numerical platform. The global convolution operation of the proposed CvGNN is more suitable for the calculation process of perturbation coefficients, and the global receptive field can also be more effective at extracting effective information from perturbation feature maps. The introduction of CvCNN can directly focus on the complex-valued perturbation feature maps themselves without separately processing the real and imaginary parts, which is more in line with the waveform-dependent physical characteristics of optical signals. Based on the experimental platform, compared with the real-valued neural network with small convolutional kernel (RvCNNC), the proposed CvGNNC improves the Q-factor by  $\sim 2.95$  dB at the optimal transmission power, while reducing the time complexity by  $\sim 44.7\%$ .

**Keywords:** complex-valued convolutional neural network; global receptive field; coherent optical communication system



**Citation:** Han, L.; Wang, Y.; Yang, H.; Zhao, Y.; Li, C. Global Receptive Field Designed Complex-Valued

Convolutional Neural Network

Equalizer for Optical Fiber

Communications. *Photonics* **2024**, *11*,

431. [https://doi.org/10.3390/](https://doi.org/10.3390/photronics11050431)

[photronics11050431](https://doi.org/10.3390/photronics11050431)

Received: 1 April 2024

Revised: 24 April 2024

Accepted: 30 April 2024

Published: 5 May 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Driven by various emerging internet services, the quantity of global data has exploded, and there is an increasing demand to perform massive amounts of data transmission and processing. For efficient use of spectrum resources, information is generally modulated in a M-ary quadrature-amplitude (M-QAM) modulation format, which is extremely susceptible to serious nonlinear impairments [1]. Moreover, long-distance transmission causes serious nonlinearity accumulation [2]. Digital backpropagation (DBP) based on split-step Fourier transform (SSFT) is an effective method for nonlinearity compensation (NLC), but it requires considerable computational resources [3,4].

Neural network (NN) algorithms have achieved excellent performance in many areas of science and technology. Various structures have been proposed and implemented in optical fiber communication systems, which have experimentally demonstrated that NNs can effectively map end-to-end relationships because of their excellent ability to fit the linear and nonlinear transform between the input and output. Ref. [5] proposed a NN equalizer based on the perturbation theory and selected triplets as input features and perturbation terms as NN output. Ref. [6] compared several NN equalizers and proved that symbol sequences can be treated as time series, so nonlinear tasks can be treated

as classification or regression tasks for time series in recurrent neural network (RNN). In our previous work [7], we regarded the triplets as feature units (FU) and constructed a dual-channel feature map. We also proposed using real-valued convolutional network classifier (RvCNNC) to process these feature maps and complete the task of nonlinearity equalization. However, it is well known that digital signals are usually represented in plural form for optical signal processing in optical fiber communication systems. Most existing NNs are real-valued frameworks that ignore the correlation between the real and imaginary components of complex signals [8]. A comprehensive design of a complex-valued neural network (CvNN) incorporating phase information in the research process was proposed, where the network parameters and the backpropagation algorithm were extended into the complex domain [9,10]. Applying CvNN simplifies waveform signal processing, and CvNN is a compatible model for wave-related information processing, as it can simultaneously deal with phase and amplitude components [11]. In [12], a CvNN was used for four-level pulse amplitude modulation (PAM-4) coherent detection. This approach achieves a better bit error rate (BER) performance than a real-valued neural network (RvNN) equalizer.

Recently, CNNs have faced challenges caused by vision transformers (ViTs) in many tasks [13,14]. To break these bottlenecks, scholars have proposed increasing the size of the effective receptive field (ERF) to improve the system performance. A global convolutional neural network (GNN) using a large convolutional kernel is proposed in [15] to eliminate the influence of shape bias and to establish a closer connection between the input feature map and the output classification result. The global convolution kernel has been proposed in [16], and the scholars have proven that the GNN performance can be greatly improved by expanding the size of the convolutional kernel and ERF.

Our novel contributions are summarized as follows:

- Complex-valued input feature map: For input data, we reconstruct a complex-valued single channel feature map from received symbols on the basis of perturbation theory.
- Equalizer design: In our proposed CNN equalizer [7], all FUs and FU position-related information is interrelated and essential. Thus, the equalizer in this paper was designed based on two aims: 1. To design a convolution kernel with a global receptive field; 2. to apply the global kernel into the complex-valued convolutional neural network (CvCNN). For the output of classifiers, we set 64-class classification labels for the received symbols, and for the output of regressors, we set difference values between the received and transmitted symbols. Based on different output data types and loss functions, we can build a nonlinear equalizer consisting of a classifier and a regressor.
- Experimental result validation: We built a 120 Gb/s polarization division multiplexing (PDM) 64QAM experimental platform with 375 km transmission distance. We evaluate our algorithm based on two aspects, the Q-factor performance and the complexity performance. In coherent optical fiber communication systems, we estimate the time complexity based on the number of floating point multiplications (FLOPs) to equalize one symbol and ignore other operations with lower impact. Moreover, the space complexity is depicted as the number of parameters required to implement the NN model.

The rest of our paper is organized as follows. In Section 2, we discuss the basic principle of perturbation theory, the structure of FUs, and the structure of the CvGNN. In Section 3, the configuration of the coherent optical fiber communication system is presented. In Section 4, the experimental results are given, along with an analysis of the performance. In Section 5, the complexity comparison between different equalizers is discussed. Finally, we conclude the paper in Section 6.

## 2. Theoretical Analysis

In this section, we review the feature map construction process by applying the perturbation theory. And based on the inherent symmetry of perturbation terms, we produce a

spatial folding method to reduce the dimensionality of feature maps. Then, a CNN structure based on dimensionally reduced feature maps is proposed for nonlinear equalization.

### 2.1. Feature Map Construction

In PDM optical fiber communication systems, the continuous signals can be denoted as  $\vec{Q}(z, t) = [Q_x(z, t), Q_y(z, t)]^T$ , where  $z$  and  $t$  represent the transmission distance and time, respectively. In optical fiber transmission link, the  $\vec{Q}(z, t)$  follows the Manakov equation as [17]:

$$\frac{\partial \vec{Q}(z, t)}{\partial z} + \frac{\alpha}{2} \vec{Q}(z, t) + i \frac{\beta_2}{2} \frac{\partial^2 \vec{Q}(z, t)}{\partial t^2} = i \frac{8}{9} \gamma |\vec{Q}(z, t)|^2 \vec{Q}(z, t), \quad (1)$$

where  $\alpha$ ,  $\beta_2$  and  $\gamma$  refer to the linear loss, group velocity dispersion, and nonlinear Kerr coefficient in the fiber optic links, respectively. The optical signal at the transmitting side can be expressed as follows:

$$\vec{Q}(z = 0, t) = \sum_k \vec{A}[k] g(t - kT_s), \quad (2)$$

where  $\vec{A}[k] = [A_x[k], A_y[k]]^T$  denotes the amplitude of the  $k_{th}$  symbol,  $T_s$  refers to the duration of the symbol, and  $g$  denotes the waveform of a carrier pulse. Based on the first-order perturbation theory and the assumption of large dispersion in the optical fiber, the nonlinear impairments can be assumed to be perturbation terms. If the received symbol sequence is expressed as  $\vec{B}[k] = [B_x[k], B_y[k]]^T$ , the corresponding perturbation term  $\vec{\delta}[k]$  can be obtained as follows:

$$\left\{ \begin{array}{l} \vec{\delta}[k] = \vec{B}[k] - \vec{A}[k] = \sum_{m,n} \overrightarrow{FU}_m^n C_{m,n} \\ \overrightarrow{FU}_m^n = \left\{ \begin{array}{l} B_x[k+m](B_x[k+n]B_x^*[k+m+n] \\ \quad + B_y[k+n]B_y^*[k+m+n]) \\ B_y[k+m](B_y[k+n]B_y^*[k+m+n] \\ \quad + B_x[k+n]B_x^*[k+m+n]) \end{array} \right\} \end{array} \right. , \quad (3)$$

where  $m$ ,  $n$  and  $m + n$  are symbol indexes with respect to the  $k_{th}$  symbol. The perturbation term is the vector dot product of  $\overrightarrow{FU}_m^n$  and perturbation coefficient  $C_{m,n}$ , and  $C_{m,n}$  can be calculated by the system link parameters using the following Equations (4a)–(4c) [18]:

$$C_{m,n} = i \frac{8}{9} \gamma \frac{\tau^2}{\sqrt{3}\beta_2} E_1 \left( -i \frac{mnT^2}{\beta_2 L} \right), \text{ Subject to : } \quad m \neq 0, n \neq 0 \quad (4a)$$

$$C_{m,n} = i \frac{8}{9} \gamma \frac{\tau^2}{\sqrt{3}\beta_2} E_1 \left( -i \frac{(m-n)\tau^2 T^2}{3|\beta_2|^2 L^2} \right), \text{ Subject to : } \quad mn = 0, m+n \neq 0 \quad (4b)$$

$$C_{m,n} = i \frac{8}{9} \gamma \frac{\tau^2}{\sqrt{3}\beta_2} \int_0^L dz \frac{1}{\sqrt{\frac{\tau^4}{3\beta_2^2} + z^2}}, \text{ Subject to : } \quad m = 0, n = 0 \quad (4c)$$

where  $\tau$  is the pulse width and  $E_1(x)$  represents the exponential integral function. Figure 1a shows the normalized amplitude of the perturbation coefficient at different  $m$  and  $n$ , and  $S$  is the maximum value of  $m$  and  $n$ . As shown in Figure 1b,c, we organize many different FUs to complete a feature map with two channels, and the space position relationships between different FUs are preserved. Moreover, in order to better preserve the inherent connection between the real part and imaginary part, we combined the two channels to form a complex-valued single-channel feature map as shown in Figure 1d. In addition, we propose using a CvCNN to classify a complex-valued single-channel feature map and

complete the nonlinear equalization of corresponding symbols. Later in the article, we use  $\mathbb{C}$  to indicate a value that corresponds to the complex domain.

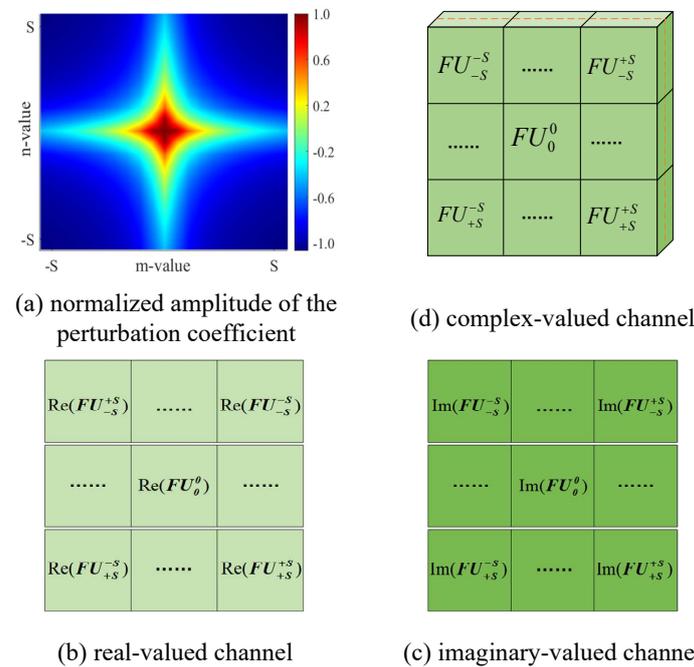


Figure 1. The construction method for input features.

Figure 2 shows the structure of the CvCNN. The convolution operation of CvCNN is the same as that of RvCNN. The difference is that all parameters are in the complex-valued domain, and the operations satisfy the complex-valued operations. The network input for complex neuron is  $X = X_R + iX_I$ , where  $X_R$  and  $X_I$  represent the real and imaginary components of  $X$ , the weight of the convolutional kernel is  $\omega = \omega_R + i\omega_I$ , and the bias is  $b = b_R + ib_I$ . So the corresponding complex neuron output is

$$Y = Y_R + iY_I = X * \omega + b = (X_R * \omega_R - X_I * \omega_I + b_R) + i(X_R * \omega_I + X_I * \omega_R + b_I). \tag{5}$$

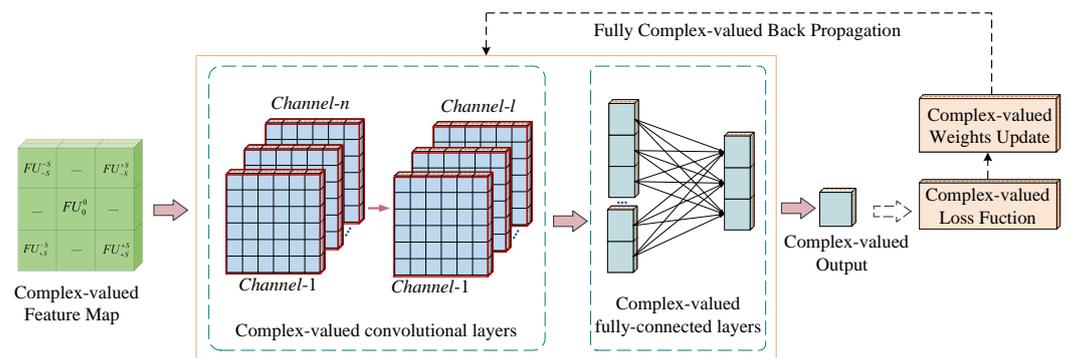


Figure 2. The nonlinear equalizer structure for CvCNN.

The nonlinear activation function can introduce nonlinearity to the affine transformations in neural networks. In this paper, we adopt the complex-valued Split-type A activation function [19], which means we activate the real and imaginary components, respectively. This can be denoted as shown in the following equation:

$$Z = f(Y_R) + if(Y_I) = Z_R + Z_I, \tag{6}$$

where  $Z$  is the final output. The most commonly used nonlinear activation functions in RvNN are Relu, Leaky Relu, Elu, and Tanh. Thus, in this paper, we adopt  $\mathbb{C}$  Relu,  $\mathbb{C}$  Leaky Relu,  $\mathbb{C}$  Elu, and  $\mathbb{C}$  Tanh as activation functions to verify the effect of the CvNN nonlinear equalizer.

In this paper, classification tasks are discussed, and the  $\mathbb{C}$ SoftMax function is used to map the probability of each category, which can be expressed as follows:

$$\mathbb{C}SoftMax(y_j) = \frac{e^{|y_j|}}{\sum_{c=1}^C e^{|y_c|}} \tag{7}$$

where  $y_j$  is the output value of the  $j$ -th ( $i = 1, 2, 3, \dots, C$ ) neural and  $C$  is the number of classifications in the output layer. When training a CvNN, the error is backpropagated from the output layer to the input layer using fully complex-valued gradient descent. Given the complex-valued output in polar form  $z_j = r_j e^{i\theta_j} \in \mathbb{C}$  in the CvNN, the complex-valued cross-entropy loss function can be expressed as follows [20]:

$$\mathcal{L}_{comp} = -\frac{1}{n} \sum_{j=1}^n \sum_{k=1}^C y_{jk} \log(|r_{jk}| e^{i\theta_{jk}}) \tag{8}$$

where  $y_{jk}$  denotes  $k$ -th element of the one-hot encoded label. The process of learning with complex domain backpropagation is similar to the learning process in the real domain, and involves finding the optimal weights  $\omega$  to minimize the loss. The loss calculated after the forward pass is backpropagated to each neuron in the network, and the weights are adjusted in the backward pass. And in CvNN,  $\omega$  can be updated by following Equation [21]:

$$\omega^{(l+1)} = \omega^{(l)} - \left( \eta_R^l \nabla_{\omega^*} E(\omega^{(l)}) + i \eta_I^l \nabla_{\omega} E(\omega^{(l)}) \right), \tag{9}$$

where  $\eta_l$  is the learning rate at the  $l$ -th iteration, and  $\nabla_{\omega^*} E$  defines the direction of the maximum rate of change with respect to  $\omega^*$ .

### 2.2. Global Convolutional Kernel in CNN

In the CNN, the convolutional layer is generally used to automatically extract the visual features of the feature map for filtering operations. The size of the convolution kernel defines the size range of the convolution, representing the size of the receptive field which refers to the range of input figure processed simultaneously in the network. Therefore, a global convolutional kernel can enhance ERF and can obtain more feature information simultaneously. Figure 3 shows the ERF ranges of different convolutional kernels, as well as different structures of CNNs using normal convolutional kernels and global convolutional kernels.

In the Figure 3a, by adopting a normal kernel design, we can obtain  $1 \times 1$  output maps (OMs) after multiple layers of convolution, which can be denoted as

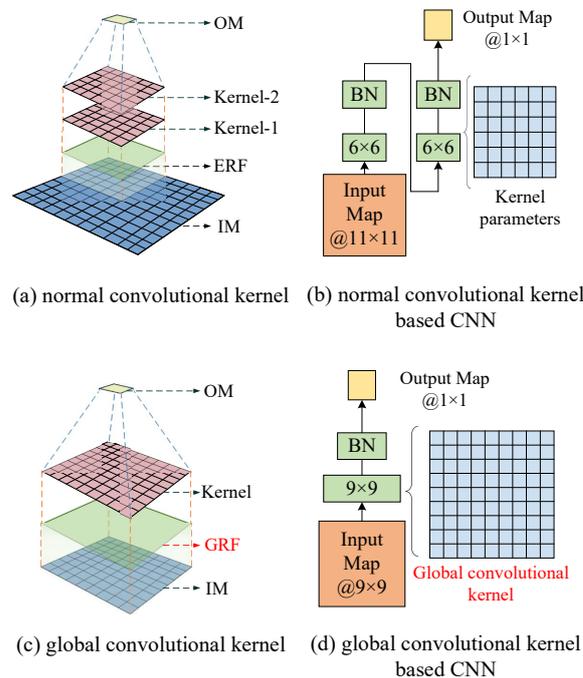
$$OM_{(1,1)} = \sum_{m,n} \left( \sum IM_{(m,n)} * \text{kernel-1} \right) * \text{kernel-2}, \tag{10}$$

where  $IM_{(m,n)}$  represents the input feature map,  $m$  and  $n$  represent the pixel index, kernel-1 is the convolution kernel for convolution layer-1, and kernel-2 is the convolution kernel for convolution layer-2. In the Figure 3c, it can be seen that compared with the normal kernel, only one-layer convolution is needed in GNN, which can be denoted as

$$OM_{(1,1)} = \sum_{m,n} IM_{(m,n)} * \text{kernel}_{(m,n)} \leftrightarrow \sum_{m,n} FU_m^n C_{m,n}. \tag{11}$$

The global kernel has a global receptive field (GRF), which can control the entire map information and better capture the correlation between pixels in the feature map and the

boundary information. Furthermore, GNN is more similar to the calculation process of perturbation term, and it is more meaningful to use GNN for nonlinear compensation.



**Figure 3.** Diagram of the different size of ERF and the different structure of CNN.

The structure of CNN is shown in Figure 3b, whose IM is  $11 \times 11$ . A double convolution layer with  $6 \times 6$  convolution kernel and batch normalization (BN) is adopted in CNN. In this paper, a global convolution kernel in Figure 3d whose size is equivalent to the size of IM is adopted because the GNN can more efficiently extract useful information without affecting the system’s performance. So, under the premise of balancing performance and complexity, the IM and global convolutional kernel sizes in GNN can be set as the optimized value  $9 \times 9$ .

### 3. Experimental Setup

Figure 4 depicts the experimental setup of the 120 Gb/s PDM 64 QAM coherent optical communication system. At the transmitter, two pseudo-random bit sequences (PRBS) are generated by MATLAB, and the sequences are combined to construct a strong random sequence that will not be learned by the NN or other advanced algorithms. Additionally, the data pattern used in the training and testing datasets has a maximum of 0.5% normalized cross-correlation to ensure the independence of the data [22]. Then, the data map to 64QAM and are loaded into an arbitrary waveform generator (AWG) with a sampling rate of 25 GSa/s. The I-channel and Q-channel signals are amplified by electric amplifiers (EAs) and then sent to the I/Q modulator with an external cavity laser (ECL). The PDM module consists of a polarization retention optical coupler (PM-OC), optical delay line (DL), polarization controller (PC), and polarizing beam combiner (PBC) to achieve polarization multiplexing. A variable optical attenuator (VOA) is used to adjust the power of the optical signal. A 375 km ( $5 \times 75$ ) standard single-mode optical fiber (SSMF) with a span of 5 is adopted, and at each end of the span, an Erbium-doped fiber amplifier (EDFA) is used to compensate for the linear loss. At the receiver, coherent detection technology is applied, and an ECL with a 100 kHz linewidth is used as the local oscillator (LO). Two PBSs are used to separate the polarization of the optical signal and LO. The X-polarization of the optical signal and LO is mixed by the  $90^\circ$  optical hybrid and detected by a balanced photonic detector (BPD). After that, two electric signals are obtained, including the X-polarization I component (X-I) and Q component (X-Q). Similarly, for the Y-polarization direction,

two electric signals are obtained, including the Y-polarization I component (Y-I) and Q component (Y-Q). A 4-channel digital phosphor oscilloscope (DPO) with a sampling rate of 100 GSa/s is used to digitize the signals.

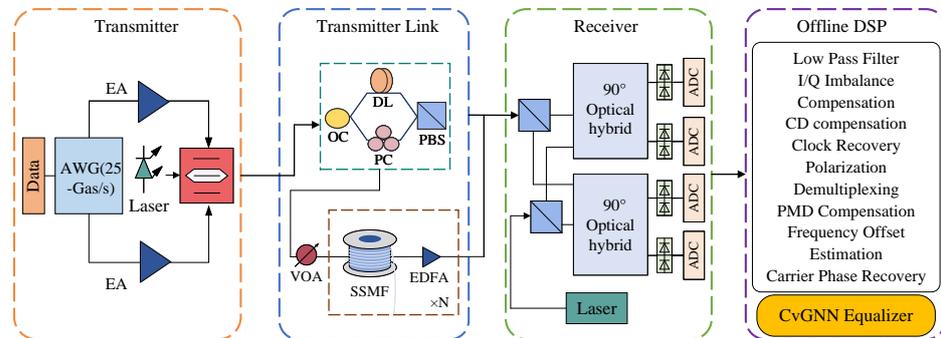


Figure 4. Experimental setup.

Offline digital signal processing (DSP) is applied to improve the signal quality. In order to better improve the overall quality of the signal, linear equalization is performed to repair the signal, and then a nonlinear equalization algorithm is adopted to enable it to learn and compensate for nonlinear damage more cleanly. Linear compensation mainly includes low-pass filter (LPF), I/Q imbalance compensation, chromatic dispersion (CD) compensation, clock recovery, polarization demultiplexing, polarization mode dispersion (PMD) compensation, frequency offset estimation (FOE), and carrier phase recovery (CPR). And the CvGNN equalizer is applied to achieve nonlinearity compensation.

The CvGNN equalizer is built, trained, and evaluated in PyTorch 3.8.1. The personal computer platform owns an AMD Ryzen7 CPU @ 2.90 GHz, and the Random Access Memory (RAM) is 16 GB. In our model, the Kaiming initialization method is applied to initialize initial weights [23], and the complex-valued Adam optimizer is employed to optimize the CvGNNC. When the output data type is labeled and the loss function is complex-valued cross-entropy, the equalizers act as classifiers. When the output data type shows the different values between the received and transmitted symbols, and the loss function is complex-valued mean square error, the equalizers act as regressors. The datasets for each LOP contain approximately  $2^{20}$  symbols, and we divided them into 70% for training and 30% for testing. The maximum training epochs are set to 1000, the initial learning rate is set to 0.003, and every 30 epochs, the learning rate drops to 90% of the original rate to prevent the learning from falling into the overfitting state.

#### 4. Results and Analysis

As mentioned above, the activation functions are essential for our NN equalizers. Taking the 1 dBm LOP as an example, we compared and verified the system Q-factor with four different activation functions using a CvGNN equalizer, as shown in Figure 5. The abscissa represents the epochs: an epoch occurs when all the training data are sent to the NN for training once. The ordinate is the Q-factor, which can better distinguish the system performance when BER is low and can be calculated by the BER using the following equation:

$$Q = 20 \log_{10} \left( \sqrt{2} \operatorname{erfcinv}(2 * \operatorname{BER}) \right). \quad (12)$$

From the Figure 5, we can determine that although C Tanh has the fastest convergence speed, it performs poorly, and C Leaky Relu performs well.

Therefore, we choose the C Leaky Relu as the nonlinear activation function in this paper. Based on the optimal parameters, we compare multiple NNs, including the system Q-factor performance under same time complexity, and the complexity observed when the Q-factor performance is achieved is similar. The specific structures of eight equalizers are displayed in Figure 6, and the Q-factor performance is shown in Figure 7.

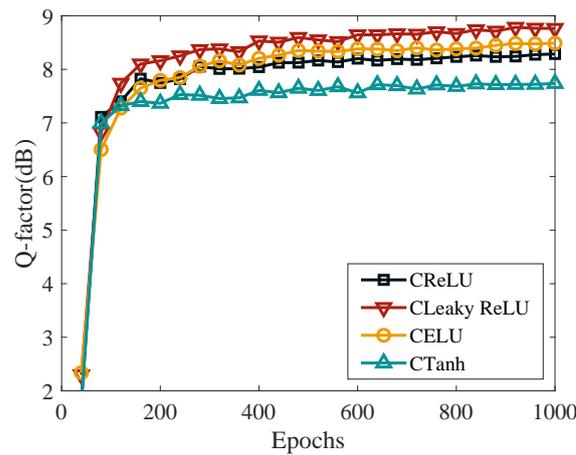


Figure 5. Q-factor trace of CvGNNC with different activation functions.

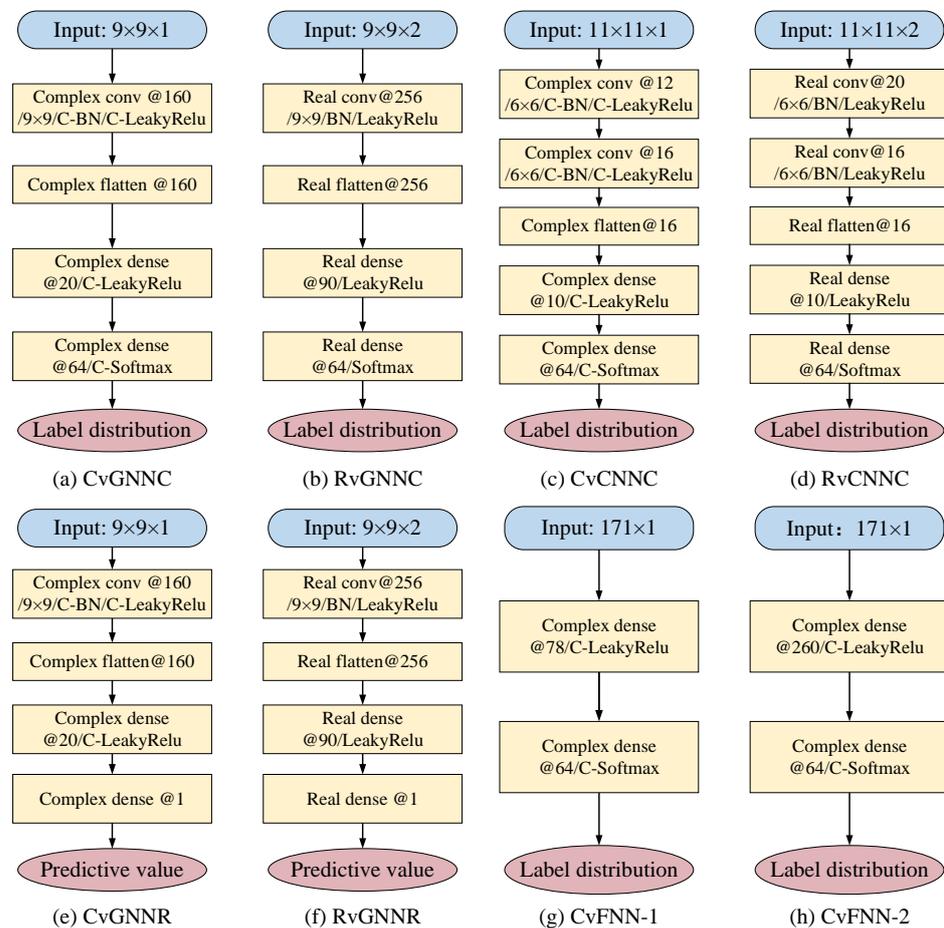
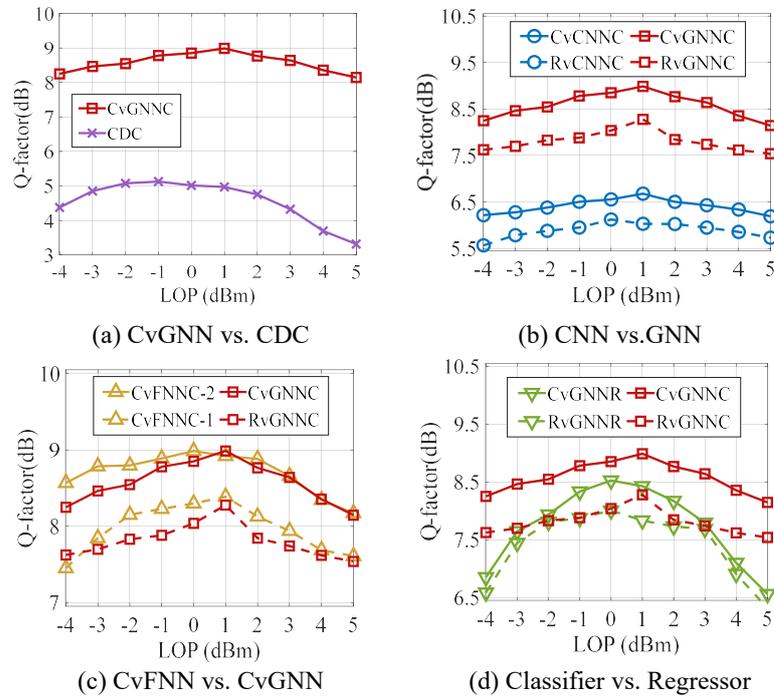


Figure 6. Structural design of different nonlinear equalizers.

In Figure 6a–f, the NN input is a feature map based on the perturbation theory. In Figure 6g,h, the NN input is the symbol sequence in which the nonlinearity equalization problem appears to be a time sequence problem. Figure 6a,e are the CvGNN classifier (CvGNNC) and regressor (CvGNNR) based on the global kernel design, the number of channels is 160, and the number of hidden layer neurons is 20. Figure 6b,f are the RvGNN classifier (RvGNNC) and regressor (RvGNNR) based on the global kernel design; the number of channels is 256; and the number of hidden layer neurons is 90. Figure 6c,d are the CvCNN classifier (CvCNNC) and RvCNN classifier (RvCNNC) based on the normal kernel. Figure 6g,h are the complex-valued fully connected NN classifiers (CvFNNC). The

CvFNNC-1 shown in Figure 6g has a total of 171 symbols at the input side with a hidden layer of 78 neurons, and the time complexity is equal compared with other equalizers; CvFNNC-2, shown in Figure 6h, has a hidden layer of 260 neurons, and its time complexity is 3.5 times that of other NNs.



**Figure 7.** Nonlinear equalization performance of different neural networks with the same time complexity.

Figure 7 presents the performance of eight nonlinear equalizers based on the different NN structures mentioned in Figure 6, which is expressed as Q-factor performance, and the LOP is in the range of  $-4$  dBm to  $5$  dBm. In Figure 7a, the performance of CvGNNC is compared with that achieved after chromatic dispersion compensation (CDC), proving that using a nonlinear equalizer at the receiver end can significantly improve the Q-factor performance of the system. As shown in Figure 7b, when LOP is  $1$  dBm, the best Q-factor of CvGNNC is  $8.98$  dB, which is  $1.15$  dB higher than RvGNNC, and CvCNNC also performs better than RvCNNC. It is proven that with the same time complexity, the CvNN system’s performance is better than that of RvNN because CvNN has greater advantages for the complex-valued operation of complex-valued perturbation characteristics. Furthermore, by using CvGNNC, we achieve  $2.31$  dB and  $2.95$  dB Q-factor improvement compared with CvCNNC and RvCNNC. It is proven that adopting a global kernel with GRF can extract FUs and their relationship information more efficiently. Therefore, a perturbation theory-aided CvGNNC can have a better compensation effect. Figure 7c proves that NN using the feature map constructed by perturbation theory more easily fits the relationship between input and output than the NN using front- and back-linked symbols as input features, both of which are CvNNs. When the time complexity of CvFNNC-2 is 3.5 times that of CvGNNC, its system performance is consistent with that of CvGNNC. When the complexity of the two NNs is the same, the Q-factor of CvGNNC is  $0.60$  dB higher than CvFNNC-1 at an LOP of  $1$  dBm. Moreover, the performance of CvFNNC-1 is  $0.55$  dB higher than that of RvGNNC, which further confirms the superiority of the CvNN in the optical fiber communication system. Figure 7d shows that for CNN, the performance of classifiers is better than that of regressors, which also proves that CvCNN is better than RvCNN for same pattern recognition tasks. The application of NN classifiers in the nonlinear equalization of optical fiber communication should be more extensive.

Additionally, as shown in Figure 8, we provide a 130 GBaud, 1200 km DP-QPSK simulation setup to corroborate the results presented in this paper. The optical fiber channel simulation is based on the split-step Fourier method (SSFM) and is implemented by MATLAB 2020a. The dispersion, nonlinear effect, and phase noise are added, and the optical signal-to-noise ratio (OSNR) is set at 30 dB. The optical transmission link is 1200 km SSMF with 20 spans, and each span incorporates an EDFA to fully compensate for linear impairments. The comprehensive simulation parameters refer to the actual optical fiber parameters, as shown in Table 1. The offline DSP is consistent with the experimental system. Simulation results are shown in the Figure 9, the 1 step-per-span (SPS) DBP and 50 SPS DBP are performed for comparison with the proposed nonlinearity equalizer CvGNNC.

Table 1. Simulation Parameters.

Parameter	$R_B$ (GBaud)	$L_{Span}$ (km)	$N_{Span}$	$L_{step}$ (m)	$\alpha$ (dB/km)	$\beta_2$ (ps <sup>2</sup> /km)	$\gamma$ /(W*km)
Value	130	60	20	50	0.2	21.667	1.3

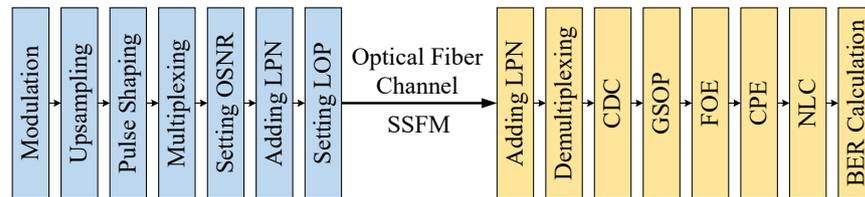


Figure 8. Simulation setup.

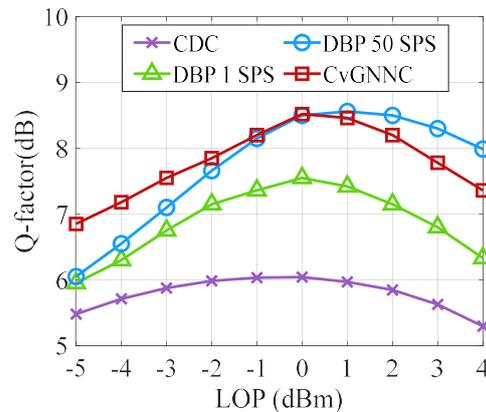


Figure 9. Nonlinear equalization performance.

It is evident that CvGNNC outperforms the 1 SPS DBP, achieving a 0.97 dB improvement in the Q-factor at the LOP of 0 dBm. Furthermore, when LOP ranges from −2 dBm to 1 dBm, the performance of CvGNNC is comparable to that of 50 SPS DBP. However, within the linear range, the CvGNNC outperforms 50 SPS DBP. Thus, unlike DBP algorithm, NNs are employed to fit end-to-end nonlinear models, enabling CvGNNC not only to balance nonlinear impairments but also to address residual linear impairments during the equalization process. Moreover, NNs offer the advantage of avoiding computationally intensive processes and do not rely on extensive or precise channel knowledge. Many algorithms have been proposed to simplify DBP-based equalizers, yet their complexity remains higher than that of NN-based equalizers [24–27]. In the same way, based on fair equalization performance, the CvGNNC exhibits significantly lower computational complexity compared to the multi-step DBP-based equalizer. Obviously, these results robustly validate the conclusions articulated in this paper.

### 5. Complexity Comparison Discussion

In this section, we analyze the proposed CvGNNC with other nonlinear equalizers in terms of time complexity and space complexity. The time complexity determines the training or prediction time of the model, and depicted as the number of FLOPs required to equalize one symbol. The space complexity is closely related to the model capacity, which can be determines the number of parameters network required.

The complexity of the CNN model is mainly concentrated in the convolutional layer and the fully connected layer. The complexity of the fully connected neural network is composed of multiple fully connected layers. Equation (13) defines the time complexity measured by the number of FLOPs, and the space complexity measured by the number of parameters, which are defined as  $N_{F\_conv}$  and  $N_{P\_conv}$  in the convolutional layer, and those in the fully connected layer are defined as  $N_{F\_fully}$  and  $N_{P\_fully}$ . And in the  $i_{th}$  convolutional layer, the  $C_{Ii}$  represents the number of input channels,  $C_{Oi}$  represents the number of output channels,  $M_i$  represents the size of the feature map, and  $K_i$  represents the size of the convolution kernel. Meanwhile, in the  $j_{th}$  fully connected layer,  $F_{Ij}$  and  $F_{Oj}$  represent the number of input and output neurons, while  $NC$  and  $NF$  represent the number of layers in the convolutional layer and fully connected layer, respectively. The specific calculation processes are as follows:

$$\left\{ \begin{array}{l} N_{F\_conv} = \sum_{i=1}^{NC} C_{Ii} * C_{Oi} * (M_i - K_i + 1)^2 * K_i^2, \\ N_{P\_conv} = \sum_{i=1}^{NC} C_{Oi} * [C_{Ii} * K_i^2 + (M_i - K_i + 1)^2], \\ N_{F\_fully} = \sum_{j=1}^{NF} F_{Ij} * F_{Oj}, \\ N_{P\_fully} = \sum_{j=1}^{NF} F_{Ij} * (F_{Oj} + 1). \end{array} \right. \quad (13)$$

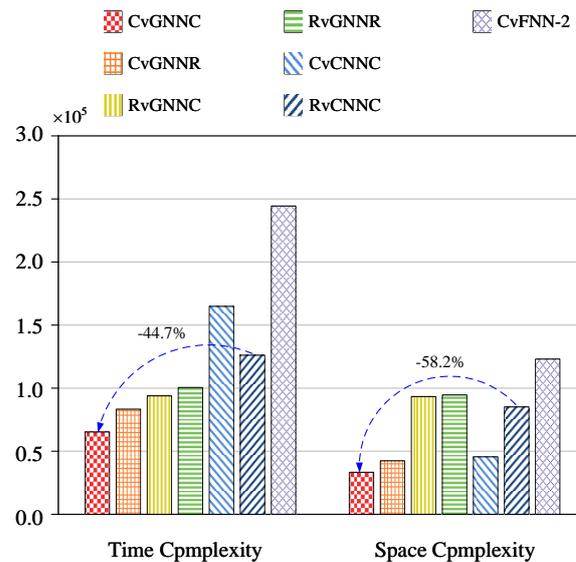
In CvNN, the data, weight and activation function are all located in the complex field, and their operations are in the complex domain. It is known that one complex-valued FLOP is equivalent to four real-valued FLOPs. Therefore, FLOPs in the CvNN should be multiplied by four times on the basis of the same RvNN structure. In the CvNN, the parameters are different for the real and imaginary components, so when calculating the network parameters in CvNN, it should be multiplied twice on the basis of the same RvNN structure. In this paper, because the number of FLOPs introduced by addition can be ignored, we consider only the number of operations provided by multiplication when calculating the complexity.

In Table 2, we list the calculation formulas of the time complexity and space complexity of multiple NNs related to the CvNN under the same performance, as well as their actual values. We can see in the table that when we calculate the time and space complexity of CvGNNC, CvGNNR, CvCNNC, and CvFNN, we multiply them by four and two times on the basis of the normal calculation. As shown in Table 1, in the CvGNNC, the number of FLOPs required is  $6.97 \times 10^4$ , and the number of parameters required is  $3.35 \times 10^4$ . Compared with RvGNNC, the time complexity and space complexity decrease by 16.3% and 16.2%, respectively. Moreover, the number of parameters required for CvCNNC is 46.5% lower than that of RvCNNC, while the number of FLOPs required is higher than that required for RvCNNC. This indicates that CvNN may introduces greater time complexity, but the application of GNN in CvNN can effectively suppress this phenomenon. In addition, we can see that compared with CvCNNC, 57.7% time complexity reduction and 22.0% space complexity reduction are obtained by CvGNNC; compared with RvCNNC, the time complexity and the space complexity of RvGNNC are reduced by 33.9% and 50.2%, respectively, which proves the superiority of the CvGNN in reducing complexity.

**Table 2.** Complexity calculation process of different equalizers.

	Time Complexity	Space Complexity	NC	NF	FLOPs ( $\times 10^4$ )	Parameters ( $\times 10^4$ )
CvGNNC	$4 \times (N_{F\_conv}^{cgc} + N_{F\_fully}^{cgc})$	$2 \times (N_{P\_conv}^{cgc} + N_{P\_fully}^{cgc})$	1	2	6.97	3.56
CvGNNR	$4 \times (N_{F\_conv}^{cgr} + N_{F\_fully}^{cgr})$	$2 \times (N_{P\_conv}^{cgr} + N_{P\_fully}^{cgr})$	1	2	8.33	4.24
RvGNNC	$N_{F\_conv}^{rgc} + N_{F\_fully}^{rgc}$	$N_{P\_conv}^{rgc} + N_{P\_fully}^{rgc}$	1	2	9.39	9.33
RvGNNR	$N_{F\_conv}^{rgr} + N_{F\_fully}^{rgr}$	$N_{P\_conv}^{rgr} + N_{P\_fully}^{rgr}$	1	2	10.05	9.46
CvCNNC	$4 \times (N_{F\_conv}^{ccc} + N_{F\_fully}^{ccc})$	$2 \times (N_{P\_conv}^{ccc} + N_{P\_fully}^{ccc})$	2	2	16.49	4.56
RvCNNC	$N_{F\_conv}^{rcc} + N_{F\_fully}^{rcc}$	$N_{P\_conv}^{rcc} + N_{P\_fully}^{rcc}$	2	2	12.61	8.52
CvFNN-2	$4 \times N_{F\_fully}^{efc}$	$2 \times N_{P\_fully}^{efc}$	/	2	24.44	12.31

Figure 10 intuitively compares the space and time complexity. From the figure, it can be concluded that, for the same performance, the time complexity and space complexity of CvFNNC-2 are much higher than that of our proposed CvGNNC. This also proves that the data obtained by constructing the feature map more easily fit the relationship between the input and output terminals. Additionally, we can easily see that when the performance is the same, the time and space complexity required by regressors is slightly higher than that of classifiers. Therefore, in the application of optical fiber communication systems, the classifiers are better than the regressors.



**Figure 10.** The computational complexity of different NNs, including time and space complexity.

### 6. Conclusions

In this paper, we propose an equalization technique using a CvGNN at the receiver of optical fiber communication systems. Based on the perturbation theory, we construct a complex-valued single-channel feature map as input to make it more suitable for complex-valued neural networks. We expand the convolution kernel to be equivalent to the feature map to expand the ERF and reduce the model depth and then trade off the performance and complexity to obtain the best parameters. Based on the optimal parameters, we select the CvGNNC, CvGNNR, RvGNNC, RvGNNR, CvCNNC, RvCNNC, and CvFNNC to compare the equalization performance and the equalization complexity. We find that at the same time complexity, a global convolutional kernel structure can further improve the performance compared with a normal convolutional kernel structure, and CvNNs are proven to be more suitable for optical fiber communication signal processing than RvNNs. In the same

performance, the time complexity and space complexity of the equalizer we proposed are 44.7% and 58.2% lower than those of the normal kernel real-valued network, respectively, which further proves the superiority of CvGNNC. Because of its low complexity and outstanding performance advantages, it can be better applied to the actual optical fiber communication systems.

**Author Contributions:** Conceptualization, L.H. and Y.W.; methodology, L.H. and Y.W.; software, L.H. and H.Y.; validation, L.H., Y.W., C.L. and H.Y.; formal analysis, L.H. and Y.W.; investigation, L.H. and Y.W.; resources, L.H.; writing—original draft preparation, L.H.; writing—review and editing, Y.W., Y.Z. and C.L.; visualization, L.H.; supervision, Y.W.; project administration, Y.W.; funding acquisition, Y.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the National Key Research and Development Program of China (2021YFB2900703) and the National Natural Science Foundation of China (62075014).

**Data Availability Statement:** Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

**Conflicts of Interest:** Author Yang Zhao was employed by the company China Rocket Co., Ltd., and Chao Li was employed by the company Inspur Electronic Information Industry Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Ip, E. Nonlinear compensation using backpropagation for polarization-multiplexed transmission. *IEEE J. Light. Technol.* **2010**, *28*, 939–951. [\[CrossRef\]](#)
2. Redyuk, A.; Averyanov, E.; Sidelnikov, O.; Fedoruk, M.; Turitsyn, S. Compensation of Nonlinear Impairments Using Inverse Perturbation Theory With Reduced Complexity. *IEEE J. Light. Technol.* **2020**, *38*, 1250–1257. [\[CrossRef\]](#)
3. Zheng, Z.; Lv, X.; Zhang, F.; Wang, D.; Sun, E.; Zhu, Y.; Zou, K.; Chen, Z. Fiber Nonlinearity Mitigation in 32-Gbaud 16QAM Nyquist-WDM Systems. *IEEE J. Light. Technol.* **2016**, *34*, 2182–2187. [\[CrossRef\]](#)
4. Napoli, A.; Maalej, Z.; Sleiffer, V.A.J.M.; Kuschnerov, M.; Rafique, D.; Timmers, E.; Spinnler, B. Reduced complexity digital back-propagation methods for optical communication systems. *IEEE J. Light. Technol.* **2014**, *32*, 1351–1362. [\[CrossRef\]](#)
5. Zhang, S.; Yaman, F.; Nakamura, K.; Inoue, T. Field and lab experimental demonstration of nonlinear impairment compensation using neural networks. *Nat. Commun.* **2019**, *10*, 3033. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Freire, P.J.; Napoli, A.; Spinnler, B.; Costa, N.; Turitsyn, S.K.; Prilepsky, J.E. Neural Networks-Based Equalizers for Coherent Optical Transmission: Caveats and Pitfalls. *J. Light. Technol.* **2022**, *28*, 7600223. [\[CrossRef\]](#)
7. Li, C.; Wang, Y.; Wang, J.; Yao, H.; Liu, X.; Gao, R.; Yang, L.; Xu, H.; Zhang, Q.; Ma, P.; et al. Convolutional Neural Network-Aided DP-64 QAM Coherent Optical Communication Systems. *IEEE J. Light. Technol.* **2022**, *40*, 3564–3572. [\[CrossRef\]](#)
8. Freire, P.J.; Abode, D.; Prilepsky, J.E.; Costa, N.; Spinnler, B.; Napoli, A. Transfer Learning for Neural Networks-Based Equalizers in Coherent Optical Systems. *IEEE J. Light. Technol.* **2021**, *39*, 6733–6745. [\[CrossRef\]](#)
9. Hirose, A.; Yoshida, S. Generalization Characteristics of Complex-Valued Feedforward Neural Networks in Relation to Signal Coherence. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 541–551. [\[CrossRef\]](#)
10. Lee, C.; Hasegawa, H.; Gao, S. Complex-Valued Neural Networks: A Comprehensive Survey. *IEEE/CAA J. Autom. Sin.* **2022**, *9*, 1406–1426. [\[CrossRef\]](#)
11. Bogdanov, S.A.; Sidelnikov, O.S. Use of complex fully connected neural networks to compensate for nonlinear effects in fibre-optic communication lines. *Quantum Electron.* **2021**, *51*, 459. [\[CrossRef\]](#)
12. Zhou, W.; Shi, J.; Zhao, L.; Wang, K.; Wang, C.; Wang, Y.; Kong, M.; Wang, F.; Liu, C.; Ding, J.; et al. Comparison of Real- and Complex-Valued NN Equalizers for Photonics-Aided 90-Gbps D-band PAM-4 Coherent Detection. *IEEE J. Light. Technol.* **2021**, *39*, 6858–6868. [\[CrossRef\]](#)
13. Yuan, L.; Hou, Q.; Jiang, Z.; Feng, J.; Yan, S. VOLO: Vision Outlooker for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 6575–6586. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Moutik, O.; Sekkat, H.; Tigani, S.; Chehri, A.; Saadane, R.; Tchakoucht, T.A.; Paul, A. Convolutional Neural Networks or Vision Transformers: Who Will Win the Race for Action Recognitions in Visual Data? *Sensors* **2023**, *23*, 734. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Peng, C.; Zhang, X.; Yu, G.; Luo, G.; Sun, J. Large Kernel Matters—Improve Semantic Segmentation by Global Convolutional Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
16. Ding, X.; Zhang, X.; Han, J.; Ding, G. Scaling Up Your Kernels to 31×31: Revisiting Large Kernel Design in CNNs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021.
17. Ding, J.; Liu, T.; Xu, T.; Hu, W.; Popov, S.; Leeson, M.S.; Zhao, J.; Xu, T. Intra-Channel Nonlinearity Mitigation in Optical Fiber Transmission Systems Using Perturbation-Based Neural Network. *IEEE J. Light. Technol.* **2022**, *40*, 1250–1257. [\[CrossRef\]](#)

18. Tao, Z.; Dou, L.; Yan, W.; Li, L.; Hoshida, T.; Rasmussen, J.C. Multiplierfree intrachannel nonlinearity compensating algorithm operating at symbol rate. *IEEE J. Light. Technol.* **2011**, *29*, 2570–2576. [[CrossRef](#)]
19. Bassey, J.; Qian, L.; Li, X. A Survey of Complex-Valued Neural Networks. *arXiv* **2021**, arXiv:2101.12249.
20. Yadav, S.; Jerripothula, K.R. CCNs: Fully Complex-valued Convolutional Networks using Complex-valued Color Model and Loss Function. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2–3 October 2023.
21. Popa, C.A. Complex-valued convolutional neural networks for real-valued image classification. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 14–19.
22. Liao, T.; Xue, L.; Huang, L.; Hu, W.; Yi, L. Training data generation and validation for a neural network-based equalizer. *Opt. Lett.* **2020**, *45*, 5113–5116. [[CrossRef](#)] [[PubMed](#)]
23. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1026–1034.
24. Oleg Sidelnikov, A.R.; Sygletos, S. Equalization performance and complexity analysis of dynamic deep neural networks in long haul transmission systems. *Opt. Express* **2018**, *26*, 32765–32776. [[CrossRef](#)]
25. Stavros, D.; Charis, M.; Adonis, B. Performance and Complexity Analysis of i-directional Recurrent Neural Network Models vs. Volterra Nonlinear Equalizers in Digital Coherent Systems. *J. Light. Technol.* **2021**, *39*, 5791–5798.
26. Freire, P.J. Deep Neural Network-Aided Soft-Demapping in Coherent Optical Systems: Regression Versus Classification. *IEEE Trans. Commun.* **2022**, *70*, 7973–7988. [[CrossRef](#)]
27. Neskorniuk, V. End-to-end deep learning of long-haul coherent optical fiber communications via regular perturbation model. In Proceedings of the European Conference on Optical Communication (ECOC), Bordeaux, France, 13–16 September 2021.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.