Article

Comparing Performance of Deep Convolution Networks in Reconstructing Soliton Molecules Dynamics from Real-Time Spectral Interference

Caiyun Li , Jiangyong He, Yange Liu, Yang Yue, Luhe Zhang, Longfei Zhu, Mengjie Zhou, Congcong Liu, Kaiyan Zhu and Zhi Wang *

Tianjin Key Laboratory of Optoelectronic Sensor and Sensing Network Technology, Institute of Modern Optics, Nankai University, Tianjin 300350, China; 1120180099@mail.nankai.edu.cn (C.L.); 1120190097@mail.nankai.edu.cn (J.H.); ygliu@nankai.edu.cn (Y.L.); yueyang@nankai.edu.cn (Y.Y.); 2120180269@mail.nankai.edu.cn (L.Z.); 2120180259@mail.nankai.edu.cn (L.Z.); 2120190246@mail.nankai.edu.cn (M.Z.); 1120200110@mail.nankai.edu.cn (C.L.); 2120190247@mail.nankai.edu.cn (K.Z.)
* Correspondence: zhiwang@nankai.edu.cn

Abstract: Deep neural networks have enabled the reconstruction of optical soliton molecules with more complex structures using the real-time spectral interferences obtained by photonic time-stretch dispersive Fourier transformation (TS-DFT) technology. In this paper, we propose to use three kinds of deep convolution networks (DCNs), including VGG, ResNets, and DenseNets, for revealing internal dynamics evolution of soliton molecules based on the real-time spectral interferences. When analyzing soliton molecules with equidistant composite structures, all three models are effective. The DenseNets with layers of 48 perform the best for extracting the dynamic information of complex five-soliton molecules from TS-DFT data. The mean Pearson correlation coefficient (MPCC) between the predicted results and the real results is about 0.9975. Further, the ResNets in which the MPCC achieves 0.9906 also has the better ability of phase extraction than VGG which the MPCC is about 0.9739. The general applicability is demonstrated for extracting internal information from complex soliton molecule structures with high accuracy. The presented DCNs-based techniques can be employed to explore undiscovered mechanisms underlying the distribution and evolution of large numbers of solitons in dissipative systems in experimental research.

Keywords: fiber nonlinearities; deep learning (DL); artificial intelligence (AI)

1. Introduction

Soliton molecules are localized soliton bound states formed by self-organized dissipative soliton through subtle interaction mechanisms [1]. The potential of soliton molecules to expand the transmission capacity in optical communication systems has drawn much research attention and has become an attractive topic for nonlinear optical fibers in recent decades [2–10]. In addition to predicting the dynamic evolution of soliton molecules theoretically [3], the dynamic evolution of soliton molecules is also proved experimentally [8–12], which extends the degrees of freedom toward internal dynamics. The internal dynamics of soliton molecules is difficult to analyze when only the change of the pulse energy is considered in the oscilloscope traces. Recently, the photonic time-stretch dispersive Fourier transformation (TS-DFT) technology has been used to real-time monitor the internal dynamics of soliton molecules in passive mode-locked lasers (PMLs). Concretely, TS-DFT observe various rare events and transient phenomena including soliton buildup [6,7], soliton pulsation [13,14], soliton explosion [15,16], and soliton molecules [2–4,9,10]. It appears tremendous potential in simulating dynamic process of various complex molecules. The structure of soliton molecules encompasses simple two-soliton and three-soliton molecules [4,5,10–12], 2+2 soliton molecular complexes [9], composite patterns in both

global and local ranges [14], and supramolecular arrangements that mimic various many-body biochemical and biological systems [8]. To restructure the internal dynamics of the soliton molecules from TS-DFT spectra, a autocorrelation method is usually employed [9]. In this method, a discrete Fourier transform is performed on the interference fringes to obtain the single-shot autocorrelation traces for retrieving the soliton separation and the relative phase in the soliton molecules.

However, the autocorrelation method cannot further quantitatively analyze all the dynamic evolution processes in complex molecular structures, such as relative phase of each soliton. When multisoliton molecules and soliton pairs with near equal spacing happen [11], it is almost impossible to obtain relative phase differences (PDs) evolution [17]. Therefore, the autocorrelation method is suitable for analyzing simple soliton molecule structures consisting of soliton pairs, unequally spacing three solitons [4], etc. Recent years have seen the rapid growth and development of the field of ultrafast photonics, where artificial intelligence algorithms are being applied in exploring complex dynamical processes of soliton molecules in PMLFLs [17], the extreme events in optical fibre modulation instability [18], and the generation and characterization of light pulses [19,20]. In order to solve the internal dynamics of complex soliton molecules, we introduced artificial intelligence combining with TS-DFT. Although the residual networks (ResNets) [21] have been used for exploring complex dynamical processes in soliton molecules experimentally and numerically based on TS-DFT in passive mode-locked lasers (PMLs), emerging models continue to push the limits of what can be achieved. It also has proved that the data generated based on theory can be used to analyze experimental data [17]. It is necessary to consider whether the network structures outside the ResNets are more accurate and effective to analyze the internal dynamics of multisoliton molecules.

Recently, deep convolution networks (DCNs) have demonstrated a powerful ability to apply in mode-locked lasers [17,22], decompose the modes in few-mode fibers [23], recognize orbital angular momentum modes with fractional topological charges [24], mitigate fiber nonlinearity in optical communication [25], and the characterization and control of ultrafast propagation dynamics [26]. It is well known that convolutional neural networks (CNNs) have dominated machine-learning landscape in data-rich applications, such as VGG (Visual Geometry Group) [27], Residual Networks (ResNets) [21], Dense Convolutional Networks (DenseNets) [28], and other models. Theoretical and empirical evidences indicate that the depth of neural networks is crucial for its accuracy and/or performance [29]. The core of DenseNets and ResNets models is to establish "shortcuts, Skip Connection" between the front and back layers, which will facilitate shortcuts and skip connections during training and enable deeper CNN networks to be trained and achieve higher accuracy. The difference in DenseNets model is that each layer can directly obtain the gradients from the loss function and the original input signal, thus forming an implicit form of deep supervision [30,31]. This makes the feature reuse through the connection of features across the channel for faster error converge. Considering the representativeness of VGG, ResNets, and DenseNets models and their characteristics of easily deepening the network layers, the three kinds of models are chosen to compare the ability in extracting internal dynamics evolution of soliton molecules.

Here, we propose and demonstrate, theoretically, the analysis the internal dynamics of bound states of complex dissipative solitons by employing DCNs. We implement VGG, ResNets, and DenseNets which are able to extract the phase evolution information of more complex soliton molecules from TS-DFT spectra data by modifying the network structure. Comparing the performance of the three DCNs by numerical analysis, the ResNets and DenseNets represent lower complexity than VGG and can easily enjoy accuracy gains from greatly increased layers. The DenseNets we used have better parameter efficiency and more lower error than ResNets in the test data. Thus, DenseNets have been demonstrated to achieve superior performance in comparison to other two models by almost any meaningful metric.

2. Methods

2.1. Generate Simulated TS-DFT Data of Soliton Molecules

The generation of simulated TS-DFT data of soliton bound state is considering factors such as bandwidth, sampling, and noise, which has been proven to be used for deep learning data sets [17]. The complex amplitude of the slowly varying envelope of soliton molecules is described by the superposition of solitons, which is given by [32]

$$U_M(T) = \sum_{k=1}^{M} u_k(T - \tau_k)e_k^{-i\varphi_k}, \tag{1}$$

where T is the relative reference time of the pulse, M is the number of solitons, and u_k, τ_k, and φ_k represent the slowly varying envelope, relative temporal delay, and relative phase of the k-th soliton, respectively. When the bandwidth and sampling speed of the electronic devices are matching with experiment, for example, the parameters of a real-time oscilloscope are 59 GHz and 200 GSa/s, the TS-DFT spectrum with resolution of 2.8626 ps is calculated first with high temporal resolution (0.01 ps) and then filtered by a fourth-order Butterworth lowpass filter and downsampled. Thus, the simulated TS-DFT dataset for the soliton molecules can be acquired based on a series of relative temporal delay τ_k and phase φ_k is given. All the TS-DFT data are superimposed white noise. The TS-DFT system, we used here, has a dispersion-compensating fiber (DCF, -134 ps2/km) with length of 1.5 km. We assume that the solitons in soliton molecules are hyperbolic secant pulses with a central wavelength of 1560 nm. As shown in Figure 1, when multisoliton molecules are considered, the TS-DFT dataset is generated with random PDs. The TS-DFT dataset is filtered and divided into a training set and a verification set proportionally (8:2). All the TS-DFTs are converted to bitmap for the inputs of DCNs. After the training via DCNs, the simulated testing dataset, with noise, is used to predict the PDs of the soliton dynamics.
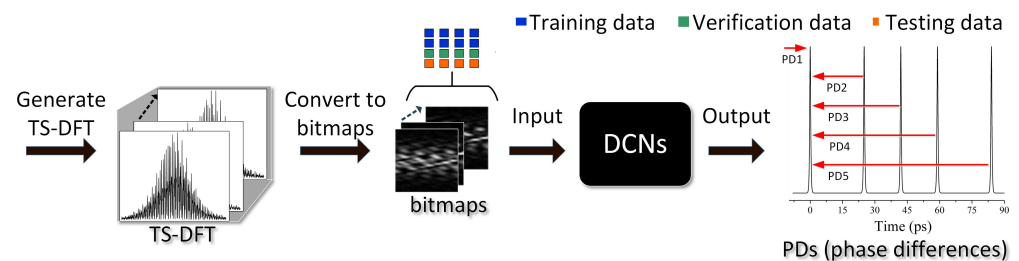

Figure 1. Processing flow for TS-DFT data for the soliton molecules based on the neural networks.

2.2. Structures of Deep Convolution Networks (DCNs)

The based architectures of three DCNs, namely VGG, ResNets, and DenseNets, are ref. [28,33,34]. We made some modification in these three models, including the number of layers of the network, the size of convolution kernel, and the structure of subblock. Especially, in Figure 2a, a batch normalization (BN) is added before each convolution block unlike VGG nets in ref [35]. Meanwhile a regularization $L2$-norm is used in each convolutional layer. The convolutional layers have the same convolution kernel (K_i) in one convolution block. With the stack of convolution blocks, the number of convolution kernels increases or is the same as the previous block. The main parts of ResNets/DenseNets are made up of their ResBlocks/Dense Blocks as shown in Figure 2b,c. The number of subblocks for each ResBlock/Dense Block is set respectively. Their structure of the subblocks are displayed in the box pointed to by the arrow. In addition, all the convolutional layers with regularization $L2$-norm are employed and batch normalization is applied among the layers. The activation function, which uses the rectified linear unit (ReLU) [36] and the Batch-Normalization [37], regularization $L2$-norm, and pooling, used in our three DCNs, can prevent overfitting. The regularization $L2$-norm makes the objective function easy to converge to the global optimal solution. The weights of the DCNs are optimized

during the training process through backpropagation. The optimizer we used is Adam [38], a variant of stochastic gradient descent that has individual adaptive learning rates for different parameters, which are calculated from estimates of the first and second moments of the gradients. Moreover, the mean absolute error (MAE) is chosen here because DCNs implement regression problems. The function of the optimizer is to reduce the gap between the predicted value and the sample label value. The DCNs' models are implemented using the Tensorflow framework [39].
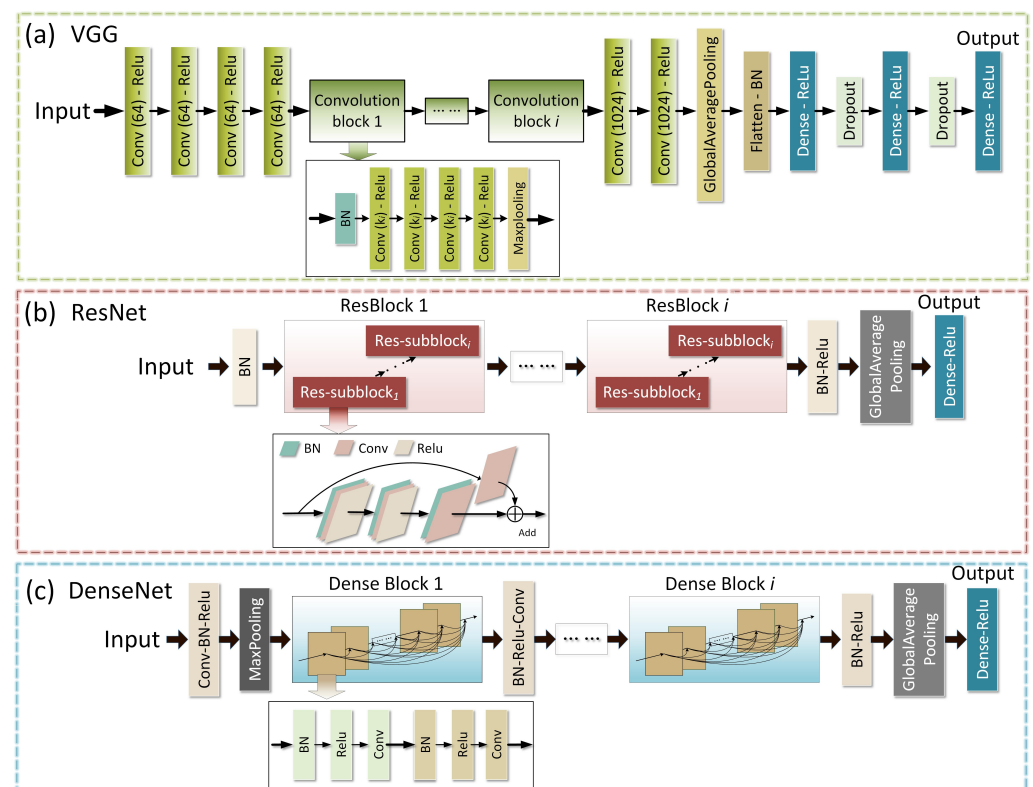

Figure 2. The structure of the DCNs (deep convolution networks). (**a**) The VGG Net (Visual Geometry Group Networks). (**b**) The ResNet (Residual Networks). (**c**) The DenseNet (Densely Connected Convolutional Networks).

3. Results and Discussion

3.1. Soliton Molecular Structure of Test Set

A complex soliton molecular structure with five solitons, which is exhibited in Figure 3, is used to test the ability of the three DCNs in extracting relative phase differences (PDs). In particular, the internal phase evolution of the soliton molecules contain oscillating and the diverging sliding phase [4,5]. The test set includes both phases and the equal temporal separations so it is impossible to extract internal phase evolution of each soliton by autocorrelation method. The temporal trace of simulated dataset is shown in Figure 3a. The temporal separations of the five solitons contain two kinds of equal spacing 17 and 42 picoseconds (ps). As presented in Figure 3a, a phasor representation is constructed to picture the five-soliton molecules constituted. We defined the leftmost soliton as the first pulse which is set as the reference with a fixed pointing direction. Then, the PD from the following pulse to the first pulse are defined as PD2, PD3, etc., denoted by the variables (φ). Figure 3e lists two PDs as representatives containing oscillating and the diverging sliding phase [40]. The TS-DFT of five-soliton molecules with given phases as the simulated testing dataset show in Figure 3b. Because there are soliton pairs with almost equal separation within the soliton molecule, their corresponding autocorrelation peaks are coherently superposed. The autocorrelation trajectories are flickering as shown in Figure 3c. Specifically, two roundtrips (580 and 704 roundtrips) of autocorrelation curves are drawn

in Figure 3d. It is obvious that the intensity varies greatly at the autocorrelation peaks for the interaction of isometric soliton molecules. This complex molecular structure as a test set involves the difficulties mentioned above and has the ability to evaluate the merits and demerits of the DCNs.
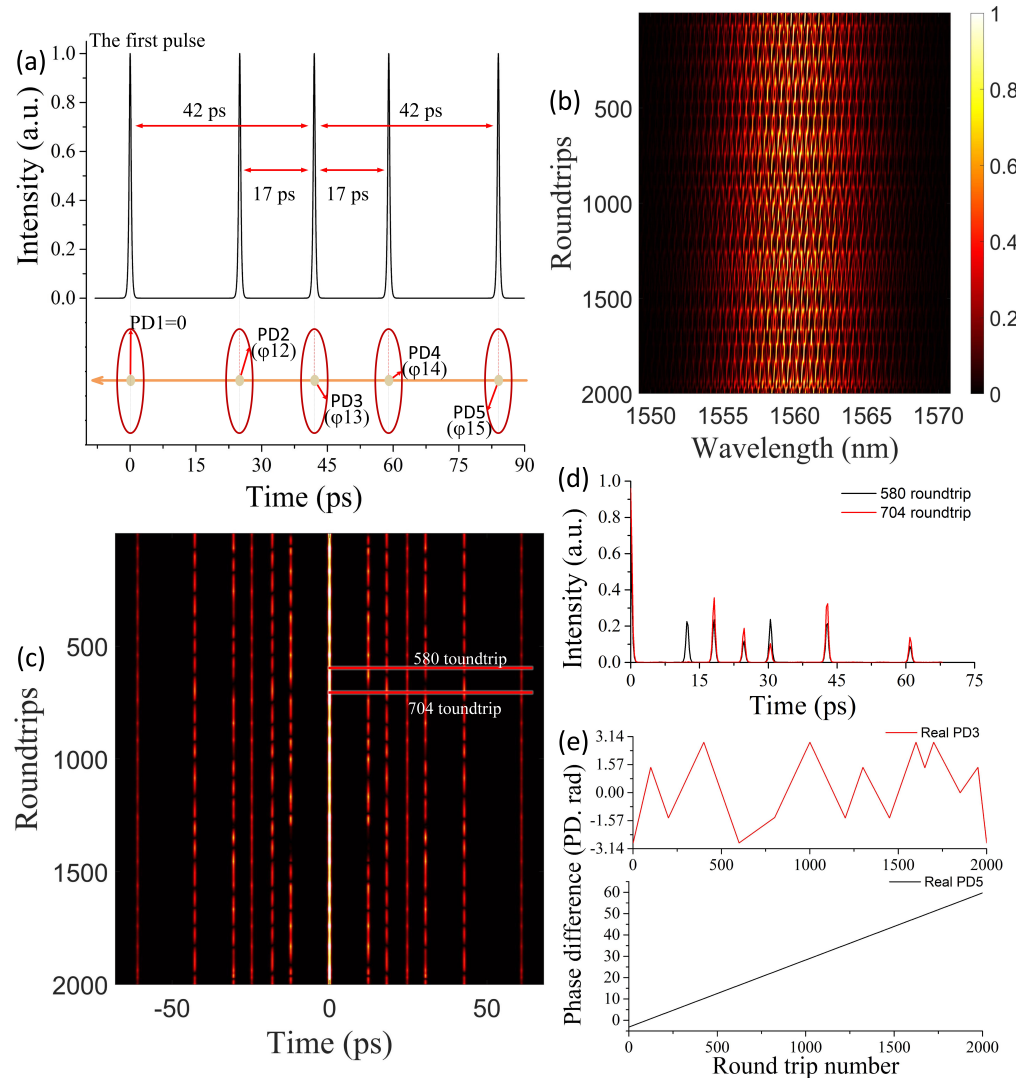

Figure 3. (**a**) Graphical representation of five-soliton molecules. (**b**) The testing TS-DFT (time-stretch dispersive Fourier transformation). (**c**) The autocorrelation trajectories of TS-DFT. (**d**) Comparison of two autocorrelation trajectories. (**e**) Two kinds of phase differences (PDs) evolution as representatives containing oscillating and the diverging sliding phases corresponding Real PD3 and PD5.

3.2. Perform Three DCNs on TS-DFT Datasets of Five-Soliton Molecules

The TS-DFT dataset, with 39 × 39 pixels each, put into three DCNs for training. We add three callback functions to control the program. These include dynamic adjustment of learning rate (LR) which is multiplied by 0.6 to decrease value if the error of lose function does not decrease after 5 iterations. The Early-Stop function is to terminate the program when the error of lose function does not decrease after 20 iterations. The Best-Model function saves optimal parameter model when the error is less than previous error. The training results are shown in Figure 4. The convergence speed and error of different networks are diverse because of the number of layers. As shown in Figure 4a, we considered for VGG of four network layers of 17, 21, 25, and 29. ResNet of three network layers of 65 (k = 512), 77 (k = 515), and 65 (k = 1024) are in Figure 4b. DenseNet of four network layers of 121 (k = 32), 161 (k = 32), 161 (k = 48), 169 (k = 32) and 169 (k = 48) are in Figure 4c.

Table 1 lists the depth of networks, the size of parameter model, the number of iterations, the verification errors and test errors of different model structures for TS-DFT of five-soliton molecules. Thereinto, the DenseNet of 161($k = 48$) has the best testing results with smallest error 2.2355 and faster convergence rate on the comprehensive. Because overfitting cannot be avoided completely and different networks have different inhibitory overfitting effects. Thus, the trends of verification error and testing error have a little inconsistency. From Figure 4, the error trend remains the same: the lower the verification error, the lower the testing error. Here we evaluate the accuracy of the networks mainly based on the error of the test data. It can be seen from Table 1 and Figure 4 that VGG networks have the worst effect for phase extraction. Its minimum testing error is high, 5.2528. DenseNet, with minimum testing error 2.2355, has a slightly smaller advantage over ResNet whose value is 2.6260. By comparing the verification errors of the optimal results in each DCNs, as shown in Figure 4d, we can still conclude that the VGG shows the worst convergence and the optimal one is the DenseNet, where the networks with shortcut connection can suppress gradient explosion better than the common convolutional network.
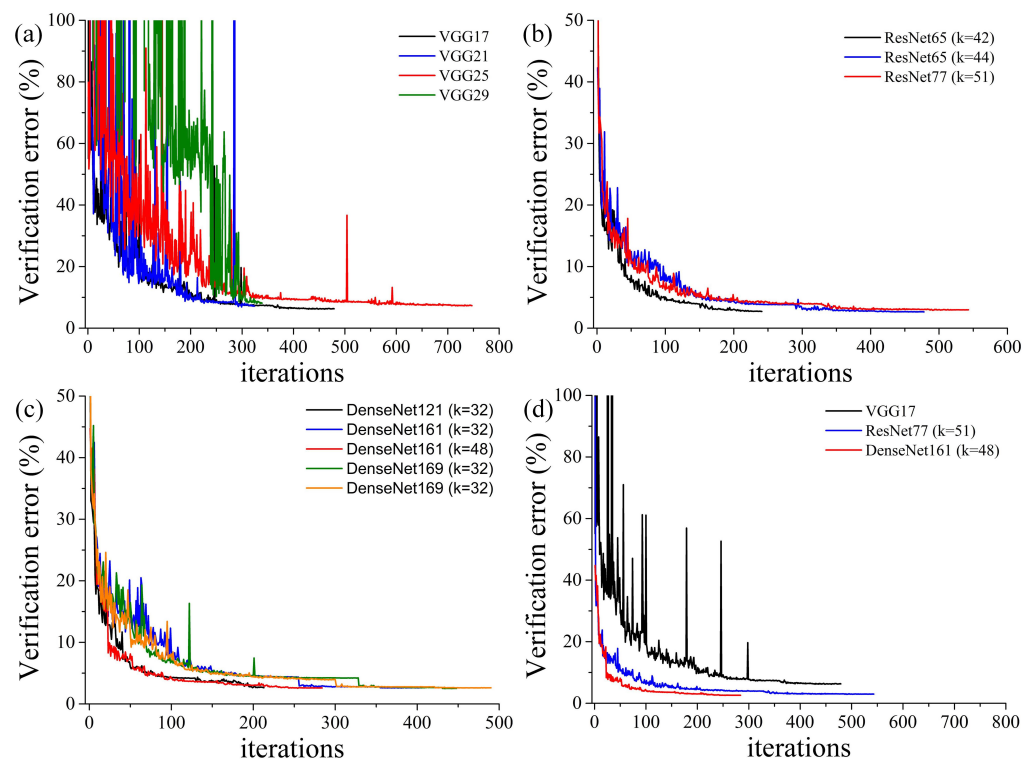

Figure 4. Verification error of DCNs. (**a**) Four VGG networks. (**b**) Three ResNets. (**c**) Five DenseNets. (**d**) Three DCNs with optimal test results.

3.3. Pearson Correlation Analysis of Real and Predicted Values

Next, we compare the real relative PDs (black lines) with the extraction results (red lines) from the optimal model in each DCN. The left column in Figure 5a is the PDs extracted from VGG-17 with a minimum error of 5.2528. Figure 5b plot the PDs extracted from ResNet-77 with a minimum error of 2.6260. In addition, the PDs extracted from DenseNet-161 ($k = 48$) with a minimum error of 2.2355 in Figure 5c. The correlation between the real value and the extracted value is analyzed by Pearson Correlation Coefficient (PCC). The mean Pearson correlation coefficients (MPCC) of each group of PDs are 0.9739, 0.9906, 0.9975 which correspond to DCNs of VGG-PDs, ResNet-PDs, and DenseNet-PDs, respectively. After comparing the VGG, ResNet, and DenseNet, the ResNets and DenseNet represent fewer smaller error and lower complexity than VGG and can easily enjoy accuracy gains from greatly increased layers. It is worth noticing that extremely deep nets with shortcut paths are easy to optimize, but simply stack layers exhibit higher testing

error when the depth increases [21]. Because short paths in the network have a strong regularizing effect and reduce overfitting on smaller training sets [30]. Besides, DenseNets we used have better parameter efficiency and more lower error than ResNets in the test data. It has been reported that DenseNets are easier to train due to their improved information flow and gradients throughout the network [30,31]. On these, the DenseNets have the best testing results with smallest testing error and superior parameter efficiency on the comprehensive. They tend to require far fewer parameters when compared against alternative algorithms with comparable accuracy. Consequently, we infer that the DCNs model have the potential to analyze the dynamics of more complex soliton molecules and DenseNets performs best.

Table 1. Error rates (%) of single-model results on the TS-DFT interference spectra of five-soliton molecules datasets.

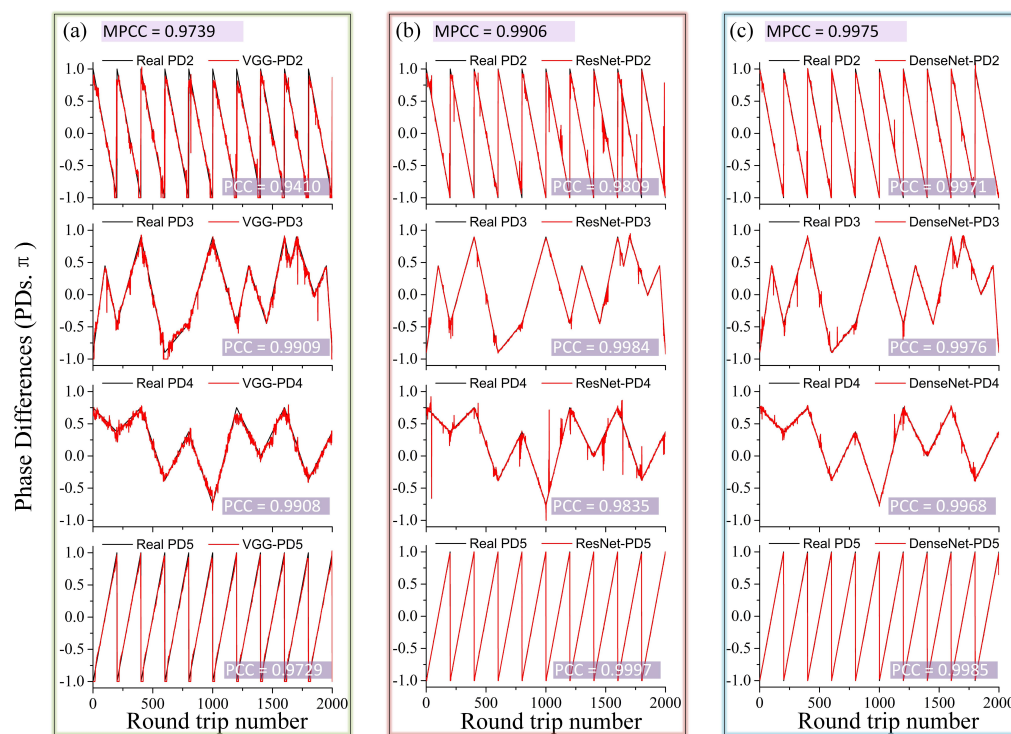| Model-Layers	Params	Iterations	Verification Error (%)	Testing Error (%)
VGG17	268 M	479	**6.2891**	**5.2528**
VGG21	272 M	324	7.1810	6.5479
VGG25	320 M	747	7.3101	6.8600
VGG29	332 M	339	8.1265	7.3815
ResNet65 ($k = 42$)	122 M	241	2.7159	2.9438
ResNet65 ($k = 44$)	426 M	543	**2.6445**	2.8491
ResNet77 ($k = 51$)	187 M	478	2.9573	**2.6260**
DenseNet121 ($k = 32$)	68.1 M	213	2.6361	2.6155
DenseNet161 ($k = 32$)	112 M	405	2.6057	2.5037
DenseNet161 ($k = 48$)	246 M	284	2.5917	**2.2355**
DenseNet169 ($k = 32$)	126 M	448	**2.5088**	2.7286
DenseNet169 ($k = 48$)	278 M	490	2.6103	2.8331

Figure 5. Relative phase difference (PD2-PD5). (**a**) The real PDs and VGG-PDs. (**b**) The real PDs and ResNet-PDs. (**c**) The real PDs and DenseNet-PDs.

4. Conclusions

The methods based on DCNs can solve the situation of more solitons and existence of equidistant soliton pairs where the autocorrelation method is limited. Comparing the

VGG, ResNet, and DenseNet models, we demonstrate their effectiveness on TS-DFT inter-ference spectra of more complex five-soliton molecules datasets with equal spacing pairs. The DenseNets outperform VGG and ResNets in extracting the internal information from complex five-soliton molecules, where the second best is the ResNets whether considering parameter efficiency or testing error. The investigation on the soliton molecule in the PMLs would contribute to understanding the complex nonlinear dynamics of pulse propaga-tion in PMLs and benefit the potential applications of telecommunications and fiber laser sources. This provides the possibility of simulating the dynamic behaviors of complex chemical molecules and other multibody systems based on soliton molecules in PMLs optically. We expect that our method can promote simulating the dynamic behaviors of complex chemical molecules and other multibody systems based on soliton molecules in PMLs optically and explore the potential mechanism of the distribution and evolution of a large numbers of solitons in a dissipative system.

Author Contributions: Conceptualization, C.L. (Caiyun Li), J.H. and Z.W.; methodology, C.L. (Caiyun Li); software, C.L. (Caiyun Li); validation, L.Z. (Luhe Zhang), L.Z. (Longfei Zhu), M.Z., C.L. (Congcong Liu) and K.Z.; formal analysis, C.L. (Caiyun Li) and Z.W.; investigation, C.L. (Con-gcong Liu); resources, Z.W., Y.Y. and Y.L.; data curation, C.L. (Caiyun Li); writing—original draft preparation, C.L. (Caiyun Li) and Z.W.; writing—review and editing, C.L. (Caiyun Li); visualization, C.L. (Caiyun Li) and Z.W.; supervision, Z.W.; project administration, Z.W., Y.Y. and Y.L.; fund-ing acquisition, Z.W., Y.Y. and Y.L. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request. The data processing and simulation codes that were used to generate the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: This manuscript has not been published or presented elsewhere in part or in entirety and is not under consideration by another journal. We have read and understood your journal's policies, and we believe that neither the manuscript nor the study violates any of these. There are no conflicts of interest to declare.

Abbreviations

The following abbreviations are used in this manuscript:

TS-DFT	time-stretch dispersive Fourier transformation
DCNs	deep convolution networks
MPCC	mean Pearson correlation coefficient
PDs	relative phase differences

References

1. Grelu, P.; Akhmediev, N. Dissipative solitons for mode-locked lasers. *Nat. Photonics* **2012**, *6*, 84–92. [CrossRef]
2. Stratmann, M.; Pagel, T.; Mitschke, F. Experimental observation of temporal soliton molecules. *Phys. Rev. Lett.* **2005**, *95*, 143902. [CrossRef] [PubMed]
3. Zavyalov, A.; Iliew, R.; Egorov, O.; Lederer, F. Dissipative soliton molecules with independently evolving or flipping phases in mode-locked fiber lasers. *Phys. Rev. A* **2009**, *80*, 043829. [CrossRef]
4. Herink, G.; Kurtz, F.; Jalali, B.; Solli, D.R.; Ropers, C. Real-time spectral interferometry probes the internal dynamics of femtosecond soliton molecules. *Science* **2017**, *356*, 50–53. [CrossRef] [PubMed]
5. Krupa, K.; Nithyanandan, K.; Andral, U.; Tchofo-Dinda, P.; Grelu, P. Real-Time Observation of Internal Motion within Ultrafast Dissipative Optical Soliton Molecules. *Phy. Rev. Lett.* **2017**, *118*, 243901. [CrossRef] [PubMed]
6. Liu, X.; Yao, X.; Cui, Y. Real-Time Observation of the Buildup of Soliton Molecules. *Phy. Rev. Lett.* **2018**, *121*, 023905. [CrossRef] [PubMed]
7. Peng, J.; Zeng, H. Build-Up of Dissipative Optical Soliton Molecules via Diverse Soliton Interactions. *Laser Photonics Rev.* **2018**, *12*, 1800009. [CrossRef]

8. He, W.; Pang, M.; Yeh, D.H.; Huang, J.; Menyuk, C.R.; Russell, P.S.J. Formation of optical supramolecular structures in a fibre laser by tailoring long-range soliton interactions. *Nat. Commun.* **2019**, *10*, 5756. [CrossRef] [PubMed]

9. Wang, Z.Q.; Nithyanandan, K.; Coillet, A.; Tchofo-Dinda, P.; Grelu, P. Optical soliton molecular complexes in a passively mode-locked fibre laser. *Nat. Commun.* **2019**, *10*, 830. [CrossRef]

10. Kurtz, F.; Ropers, C.; Herink, G. Resonant excitation and all-optical switching of femtosecond soliton molecules. *Nat. Photonics* **2020**, *14*, 9–13. [CrossRef]

11. Luo, Y.; Xia, R.; Shum, P.; Ni, W.; Ys, L.; Lam, H.; Sun, Q.; Tang, X.; Zhao, L. Real-time dynamics of soliton triplets in fiber lasers. *Photonics Res.* **2020**, *8*, 884–891. [CrossRef]

12. Liang, H.; Zhao, X.; Liu, B.; Yu, J.; Liu, Y.; He, R.; He, J.; Li, H.; Wang, Z. Real-time dynamics of soliton collision in a bound-state soliton fiber laser. *Nanophotonics* **2020**, *9*, 1921–1929. [CrossRef]

13. Peng, J.; Boscolo, S.; Zhao, Z.; Zeng, H. Breathing dissipative solitons in mode-locked fiber lasers. *Sci. Adv.* **2019**, *5*, eaax1110. [CrossRef]

14. Wang, Z.; Wang, Z.; Liu, Y.; He, R.; Zhao, J.; Wang, G.; Yang, G. Self-organized compound pattern and pulsation of dissipative solitons in a passively mode-locked fiber laser. *Opt. Lett.* **2018**, *43*, 478–481. [CrossRef]

15. Runge, A.F.J.; Broderick, N.G.R.; Erkintalo, M. Observation of soliton explosions in a passively mode-locked fiber laser. *Optica* **2015**, *2*, 36–39. [CrossRef]

16. Wang, X.; Liu, Y.G.; Wang, Z.; Yue, Y.; He, J.; Mao, B.; He, R.; Hu, J. Transient behaviors of pure soliton pulsations and soliton explosion in an L-band normal-dispersion mode-locked fiber laser. *Opt. Express* **2019**, *27*, 17729–17742. [CrossRef] [PubMed]

17. Li, C.; He, J.; He, R.; Liu, Y.; Yue, Y.; Liu, W.; Zhang, L.; Zhu, L.; Zhou, M.; Zhu, K.; et al. Analysis of real-time spectral interference using a deep neural network to reconstruct multi-soliton dynamics in mode-locked lasers. *APL Photonics* **2020**, *5*, 116101. [CrossRef]

18. Närhi, M.; Salmela, L.; Toivonen, J.; Billet, C.; Dudley, J.M.; Genty, G. Machine learning analysis of extreme events in optical fibre modulation instability. *Nat. Commun.* **2018**, *9*, 4923. [CrossRef]

19. Boscolo, S.; Finot, C. Artificial neural networks for nonlinear pulse shaping in optical fibers. *Opt. Laser Technol.* **2020**, *131*, 106439. [CrossRef]

20. Kokhanovskiy, A.; Bednyakova, A.; Kuprikov, E.; Ivanenko, A.; Dyatlov, M.; Lotkov, D.; Kobtsev, S.; Turitsyn, S. Machine learning-based pulse characterization in figure-eight mode-locked lasers. *Opt. Lett.* **2019**, *44*, 3410–3413. [CrossRef]

21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 27–30 June 2016; pp. 770–778. [CrossRef]

22. Baumeister, T.; Brunton, S.L.; Kutz, J.N. Deep learning and model predictive control for self-tuning mode-locked lasers. *J. Opt. Soc. Am. B Opt. Phys.* **2018**, *35*, 617–626. [CrossRef]

23. An, Y.; Huang, L.; Li, J.; Leng, J.; Yang, L.; Zhou, P. Learning to decompose the modes in few-mode fibers with deep convolutional neural network. *Opt. Express* **2019**, *27*, 10127–10137. [CrossRef] [PubMed]

24. Liu, Z.; Yan, S.; Liu, H.; Chen, X. Superhigh-Resolution Recognition of Optical Vortex Modes Assisted by a Deep-Learning Method. *Phys. Rev. Lett.* **2019**, *123*, 183902. [CrossRef] [PubMed]

25. Zibar, D.; Piels, M.; Jones, R.; Schaeeffer, C.G. Machine Learning Techniques in Optical Communication. *J. Lightwave Technol.* **2016**, *34*, 1442–1452. [CrossRef]

26. Genty, G.; Salmela, L.; Dudley, J.M.; Brunner, D.; Kokhanovskiy, A.; Kobtsev, S.; Turitsyn, S.K. Machine learning and applications in ultrafast photonics. *Nat. Photonics* **2020**, *15*, 91–101. [CrossRef]

27. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J.Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]

28. Huang, G.; Liu, Z.; Van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. *arXiv* **2017**, arXiv:1608.06993.

29. Yu, D.; Seltzer, M.L.; Li, J.; Huang, J.T.; Seide, F. Feature Learning in Deep Neural Networks—Studies on Speech Recognition Tasks. *arXiv* **2013**, arXiv:1301.3605.

30. Huang, G.; Liu, Z.; Pleiss, G.; van der Maaten, L.; Weinberger, K.Q. Convolutional Networks with Dense Connectivity. *arXiv* **2020**, arXiv:2001.02394.

31. Lee, C.Y.; Xie, S.; Gallagher, P.W.; Zhang, Z.; Tu, Z. Deeply-Supervised Nets. In Proceedings of the Artificial Intelligence and Statistics (AISTATS), San Diego, CA, USA, 9–12 May 2015; Volume 38, pp. 562–570.

32. Wang, Z.; Wang, Z.; Liu, Y.; He, R.; Wang, G.; Yang, G.; Han, S. Generation and time jitter of the loose soliton bunch in a passively mode-locked fiber laser. *Chin. Opt. Lett.* **2017**, *15*, 080605. [CrossRef]

33. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity Mappings in Deep Residual Networks. In Proceedings of the 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; Leibe, B., Matas, J., Sebe, N., Welling, M., Ed.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016; Volume 9908, pp. 630–645. [CrossRef]

34. Machrisaa, C. tensorflow-vgg: VGG19 and VGG16 on Tensorflow. Available online: https://github.com/machrisaa/tensorflow-vgg (accessed on 10 February 2021).

35. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

36. Li, Y.; Yuan, Y. Convergence Analysis of Two-layer Neural Networks with ReLU Activation. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 597–607.

37. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.
38. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
39. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for large-scale machine learning. In Proceedings of the OSDI'16: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
40. Meng, F.; Lapre, C.; Billet, C.; Genty, G.; Dudley, J.M. Instabilities in a dissipative soliton-similariton laser using a scalar iterative map. *Opt. Lett.* **2020**, *45*, 1232–1235. [CrossRef]