Titre: Title: Engineering spectro-temporal light states with physics-embedded deep learning. Supplément

Auteurs: Authors: Shilong Liu, Stéphane Virally, Gabriel Demontigny, Patrick Cusson, & Denis Seletskiy

Date: 2025

Type: Article de revue / Article

Référence: Citation: Liu, S., Virally, S., Demontigny, G., Cusson, P., & Seletskiy, D. (2025). Engineering spectro-temporal light states with physics-embedded deep learning. Ultrafast Science, 5, 14 pages. https://doi.org/10.34133/ultrafastscience.0107

Titre de la revue: Journal Title:	Ultrafast Science (vol. 5)
Maison d'édition: Publisher:	American Association for the Advancement of Science
URL officiel: Official URL:	https://doi.org/10.34133/ultrafastscience.0107
Mention légale: Legal notice:	

Supplementary Material for: Engineering spectro-temporal light states with physics-embedded deep learning

Contents

S1 Parameter space and implementations

S1.1 The $4f$ pulse shaper system

The $4f$ pulse shaper system consists of two diffraction gratings, a spatial light modulator (SLM), and two lenses. The input laser pulse, originating from the mode-locked fiber laser and represented as $\tilde{A}_{\text{in}}(\omega)$, is first dispersed into its spectral components by the first diffraction grating. The dispersed beam is then collimated and focused by the input lens.

At the *Fourier plane* of the input lens, the electric field is transformed from the spectral-temporal domain into the spatial-temporal domain, and is expressed as $\tilde{A}_{\text{in}}(x_\omega)$, where x_ω represents the spatial position along the horizontal direction. The spectral-spatial dispersion relationship at the Fourier plane can be derived as follows [1, 2]:

$$\frac{\delta\omega}{\delta x} = \frac{2\pi c d\cos(\theta_d)}{\lambda_0^2 f_0}, \tag{S1}$$

where c is the speed of light in vacuum, d is the grating's spatial period, λ_0 is the central wavelength of the input pulse, and θ_d is the first-order diffraction angle.

Due to diffraction, each spectral component acquires a finite size δx in the Fourier plane [1]:

$$\delta x = \sqrt{2\ln(2)}\frac{\cos(\theta_i) f_0\lambda_0}{\cos(\theta_d)\pi w_{\text{in}}}, \tag{S2}$$

where f_0 is the focal length of the lens, w_{in} is the waist of the input beam, and θ_i and θ_d are the input and first-order diffraction angles of the optical grating, respectively.

Using Supplementary Equations (S1) and (S2), the temporal window T_w for the $4f$ pulse shaper system is obtained as:

$$T_w = \frac{4\ln(2)}{\delta\omega}. \tag{S3}$$

The temporal window T_w plays an important role in achieving high-quality pulse shaping and determines the range of the parameter space used in the subsequent subsection. Based on the parameters listed in Table S1, the calculated temporal window is approximately $8.5\,\text{ps}$.

Table S1: Specific parameters of the $4f$ pulse shaper system.

Parameter	**Value**	**Units**
Central wavelength, λ_0	1550	nm
Grooves of optical grating	940	grooves/mm
Focal length of the lens, f_0	150	mm
Resolution of SLM	1440 × 1050	pixels
Input beam waist, w_{in}	1	mm
Diffraction angle on the grating, θ_d	47.5	°

S1.2 The Parameter Space

The parameter space defines the ranges of values for the relevant dispersion parameters. In the current pulse shaper system, we focus on shaping the spectral phase to alter various temporal profiles. A general spectral phase ϕ_{SLM}, applied to the hologram of the SLM, can be expressed as a superposition of regular and fractional group velocity dispersions (GVD), written as:

$$\phi_{\mathrm{SLM}} = \sum_{k=2,3,\ldots} \frac{\beta_k}{k!}\omega^k X_k + \sum_{\alpha=0,0.1,\ldots}^{2} \frac{D_\alpha}{2}|\omega|^\alpha X_\alpha, \tag{S4}$$

where β_k is the k-th regular GVD coefficient, D_α is the fractional GVD coefficient, α is the Lévy index (LI), and X represents the dispersion distance. The inclusion of fractional dispersions significantly expands the parameter space, enabling more diverse waveform shaping [3, 4].

In this study, the parameter space incorporates both regular and fractional dispersions. Specifically, we found that fractional dispersions with $\alpha = \{0.2, 0.5, 1\}$ and regular dispersions with $k = \{2, 3, 4\}$ are most effective for shaping the desired waveforms. This results in a complex superposed spectral phase:

$$\exp\left[i\left(\frac{\beta_2 X_2}{2!}\omega^2 + \frac{\beta_3 X_3}{3!}\omega^3 + \frac{\beta_4 X_4}{4!}\omega^4 + \frac{D_{0.2} X_5}{2!}|\omega|^{0.2} + \frac{D_{0.5} X_6}{2!}|\omega|^{0.5} + \frac{D_1 X_7}{2!}|\omega|^{1}\right)\right]. \tag{S5}$$

The values of the dispersion coefficients β_k and D_α are fixed as follows:

$$\begin{aligned}
\beta_2 &= -21 \times 10^{-3}\,\mathrm{ps}^2/\mathrm{m},\\
\beta_3 &= 0.1 \times 10^{-3}\,\mathrm{ps}^3/\mathrm{m},\\
\beta_4 &= -0.01 \times 10^{-3}\,\mathrm{ps}^4/\mathrm{m},\\
D_{0.2} &= 21 \times 10^{-3}\,\mathrm{ps}^{0.2}/\mathrm{m},\\
D_{0.5} &= 21 \times 10^{-3}\,\mathrm{ps}^{0.5}/\mathrm{m},\\
D_1 &= 21 \times 10^{-3}\,\mathrm{ps}^{1}/\mathrm{m}.
\end{aligned} \tag{S6}$$

While the dispersion coefficients remain fixed, the dispersion parameters $\{X_n\}$, related to the dispersion length, are scanned as part of the parameter space. To avoid significant distortions in the pulse profile due to the finite temporal window of the pulse shaping system (Eq. S3), boundaries for $\{X_n\}$ are imposed, as detailed below:

$$\begin{aligned}
X_2 &= [-10, 6]\,\mathrm{m},\\
X_3 &= [-800, 800]\,\mathrm{m},\\
X_4 &= [-2000, 1500]\,\mathrm{m},\\
X_5 &= [-150, 150]\,\mathrm{m},\\
X_6 &= [-80, 60]\,\mathrm{m},\\
X_7 &= [-30, 30]\,\mathrm{m}.
\end{aligned} \tag{S7}$$

By loading these phase profiles onto the SLM, the system can effectively perform pulse shaping. To minimize crosstalk and improve the signal-to-noise ratio, an additional spatial optical grating phase is applied along the vertical direction along the SLM' surface to generate the phase pattern in the first-order diffraction [5, 6, 7, 2]. Using a second-harmonic generation Frequency-Resolved Optical Gating (SHG-FROG) system, we calibrated the pulse shaping setup, including offset group delay dispersion (GDD), as described in the previous work [8].

S2 The structure of networks and simulations

S2.1 Strategy for Numerical Simulations

We employ Latin Hypercube Sampling (LHS) [9] to systematically sample the dispersion parameters $\{X_n\}$, leading to a series of complex phases as defined in Eq. (S5). Given the spectral amplitude, the initial temporal field can be estimated via the inverse Fourier transform:

$$A_{\text{in}}(t) = \mathcal{F}^{-1}\left[|\tilde{A}(\omega)| \exp\left(i\phi_{\text{SLM}}(\omega)\right)\right], \tag{S8}$$

where $|\tilde{A}(\omega)|$ is the spectral amplitude and $\phi_{\text{SLM}}(\omega)$ is the spectral phase. Once the initial temporal field is obtained, the supercontinuum (SC) can be simulated using the nonlinear Schrödinger equation (NLSE) or the generalized nonlinear Schrödinger equation (GNLSE) described in the main text.

To numerically simulate the NLSE and GNLSE, we employ the Split-Step Fourier Method, which takes the general form:

$$i\frac{\partial U}{\partial z} = \hat{L}\{U\} + \hat{N}\{U\}, \tag{S9}$$

where $U \equiv A(t,z)$ represents the field envelope. The standard Split-Step Fourier transform (FFT) method approximates the propagation over a small step δz as [10]:

$$U(t, z+\delta z) = \exp\left(-i\hat{L}\frac{\delta z}{2}\right) \cdot \exp(-i\hat{N}\delta z) \cdot \exp\left(-i\hat{L}\frac{\delta z}{2}\right) \cdot U(t,z). \tag{S10}$$

The dispersion operation $\hat{L}$ is applied in the frequency domain, while the nonlinear effect $\hat{N}$ is applied in the time domain. For the NLSE, only the Kerr nonlinearity is considered. In the case of GNLSE, the presence of the Raman response function (RRF) introduces additional complexity. The nonlinear step can be expressed as [11]:

$$i\frac{\partial U}{\partial z} = \gamma\left(1 - i\tau_{\text{shock}}\frac{\partial}{\partial t}\right) UR(t) * |U|^2, \tag{S11}$$

where $*$ represents the convolution operator:

$$R * |U|^2 = \int R(t-t')|U(t',z)|^2\, dt' \tag{S12}$$

over the interval $[z_0, z_0 + \delta z]$.

The fourth-order Runge-Kutta (RK) method can be used to solve the nonlinear step. Following the formulation in [12], the nonlinear term can be expressed a new formula:

$$V(t,z) = U(t,z)\exp\left[-i(z-z_0)\gamma R(t) * |U_0|^2\right], \tag{S13}$$

where $U_0 = V_0 = U(t, z = z_0)$. The first-order derivative of V can be written as:

$$\frac{\partial V}{\partial z} = i\gamma V R(t) * \left(|V_0|^2 + |V|^2\right) - \gamma\frac{1}{\omega_0}\frac{\partial}{\partial t}\left(VR(t) * |V|^2\right). \tag{S14}$$

The convolution integrals in the time domain can be expressed as products in the Fourier domain:

$$\frac{\partial V}{\partial z} = i\gamma V \mathcal{F}^{-1}\left[R(\omega)\mathcal{F}(|V_0|^2 + |V|^2)\right] - \gamma \frac{1}{\omega_0}\frac{\partial}{\partial t}\left(V\mathcal{F}^{-1}\left[R(\omega)\mathcal{F}(|V|^2)\right]\right). \tag{S15}$$

This equation can be solved using the standard fourth-order Runge-Kutta algorithm [13]. Here:

$$\begin{aligned} &\tfrac{\partial y}{\partial x} = F(x,y), \quad y = V(t), \quad x = z, \\ &F(x,y) = i\gamma V\mathcal{F}^{-1}\left[R(\omega)\mathcal{F}(|V_0|^2 + |V|^2)\right] - \gamma \tfrac{1}{\omega_0}\tfrac{\partial}{\partial t}\left(V\mathcal{F}^{-1}\left[R(\omega)\mathcal{F}(|V|^2)\right]\right). \end{aligned} \tag{S16}$$

In this equation, the Raman response function $R(\omega)$ or $R(t)$ is required. $R(t)$ is related to the Raman gain spectrum $g_R(\omega)$ as [10]:

$$g_R(\omega) = \frac{\omega_0}{cn_0(\omega_0)} f_R \kappa^{(3)} \mathrm{Im}[R(\omega)]. \tag{S17}$$

Here, the Raman gain spectrum $g_R(\omega)$ is accessible by experiments, and $R(\omega)$ is the transfer function obtained via the Fourier transform (FT) of $R(t)$. The real part of $R(\omega)$ can be derived from its imaginary part using the Kramers-Kronig relation [14]. Once $R(\omega)$ is determined, the Raman response function $R(t)$ can be recovered in the time domain.

In our simulation, the used Raman gain spectrum $g_R(\omega)$ comes from the data for the fused silica [15]. By iteratively applying the split-step Fourier method and solving the nonlinear step using the RK algorithm, we obtain the final supercontinuum output, such as the spectral intensity $I_{\mathrm{SC}}(\omega)$.

S2.2 Feed-Forward Neural Networks (FNN)

Using Latin Hypercube Sampling (LHS) [9], we systematically sample the dispersion parameters $\{X_n\}$, which lead to a series of complex phases $\Phi(\omega)$ as defined in Eq. (S5). In the simulation, we set the boundary for the dispersion parameters $\{X_n\}$ to be:

$$\begin{aligned} X_2 &= [-4, 4]\,\mathrm{m}, \\ X_3 &= [-500, 500]\,\mathrm{m}, \\ X_4 &= [-1000, 1000]\,\mathrm{m}, \\ X_5 &= [-100, 100]\,\mathrm{m}, \\ X_6 &= [-60, 60]\,\mathrm{m}, \\ X_7 &= [-20, 20]\,\mathrm{m}. \end{aligned} \tag{S18}$$

These ranges of parameters are different from the case of experiment as shown in Eq. S15. By incorporating the spectral amplitude, the initial temporal field $A_{\mathrm{in}}(t)$ is obtained using Eq. (S8). Numerical simulations then generate the final output supercontinuum spectra $I_{SC}(\omega)$, with the length of the highly nonlinear fiber fixed at $3.5\,\mathrm{cm}$.

To enrich the input features for the feed-forward neural network (FNN), we use both the spectral phase $\Phi(\omega)$ and the temporal intensity $A_{\mathrm{in}}(t)$ as a complex input. Here, the spectral phase $\Phi(\omega)$ is normalized to the range [0, 1], while $A_{\mathrm{in}}(t)$ is normalized to its maximum value within the training set. An example input is shown in Fig. S1(a), where the left part depicts the normalized spectral phase $\Phi(\omega)$, and the right curve shows the corresponding temporal intensity $A_{\mathrm{in}}(t)$.

The structure of the feed-forward neural network is shown in Fig. S1(d). The network consists of three hidden layers with 200, 200, and 100 neurons, respectively. The activation function for each hidden layer is set to a hyperbolic tangent sigmoid transfer function ('tansig'), mathematically expressed as $\tanh(N)$ [16], and the output is mapped to the interval $[-1, 1]$. The final output layer employs a linear transfer function ('purelin') to ensure a direct mapping to the output spectrum. To mitigate statistical overfitting, a regularization parameter of 0.125 is introduced, and all parameters are optimized for the FNN model.

The total dataset consists of 2500 samples, divided into 70% for training, 15% for validation, and 15% for testing. Figure S1(b) shows the loss curves for both training and validation processes. The minimum training loss achieves approximately $-40\,\mathrm{dB}$, demonstrating effective learning for the training data.

Figure S1(c) compares the predicted spectrum with the ground truth generated by the GNLSE. The input parameters correspond to those shown in Fig. S1(a). The blue curve represents the spectrum calculated using the GNLSE, while the red curve shows the output predicted by the FNN. Minor discrepancies between the two spectra are observed, particularly near low-amplitude regions. Notably, the FNN occasionally outputs negative values for some test inputs, which are subsequently normalized for practical applications.

S2.3 Convolutional Neural Network (CNN)

Using the same dataset employed for training the FNN, we train a physics-embedded convolutional neural network (P-CNN), as illustrated in Fig. S1(e). For the P-CNN, the input is the Wigner function derived from the initial pulse field. The original Wigner function matrix, with a size of 8192×8192, is computationally prohibitive for training due to its large size. To optimize computational efficiency, we apply both cropping and downsampling, reducing the matrix size to 241×81. The time window is restricted to $[-1.76, 1.76]\,\mathrm{ps}$, and the frequency axis ω spans from -85 to $85\,\mathrm{ps}^{-1}$. An example of the reduced Wigner function, corresponding to the initial condition in Fig. S1(a), is shown in Fig. S1(f).

The structure and explanation of the CNN is also detailed in the Methods-2 section of the main text. For training, we employ the 'adam' optimizer with an initial learning rate of 0.00015 and an L_2-regularization parameter of 0.001. These hyperparameters are carefully optimized for the simulated dataset to achieve a balance between convergence speed and generalization.

The training process is shown in Fig. S1(g), where the loss (in dB) is plotted against the number of epochs. After approximately 2000 epochs, the loss stabilizes around $-24.5\,\mathrm{dB}$ (or 0.0035), indicating good convergence and effective training.

To evaluate the predictive performance of the trained CNN, we feed the downsampled Wigner function as input and obtain the corresponding output spectrum. Figure S1(h) presents the predicted spectrum, demonstrating strong agreement with the reference solution generated by the GNLSE.

S2.4 Performance Comparison and Noise Evaluation of FNN and P-CNN

The performance of the networks is evaluated using the loss function defined as the half mean squared error (MSE), which is commonly used for regression tasks:

$$\mathrm{Loss} = \frac{1}{2N} \sum_{j=1}^{M} \left(Y_{\mathrm{Pre},j} - Y_{\mathrm{Tar},j}\right)^2 , \tag{S19}$$

where $Y_{\mathrm{Pre},j}$ is the predicted output, $Y_{\mathrm{Tar},j}$ is the target output, M is the total number of responses, and N is the number of observations.

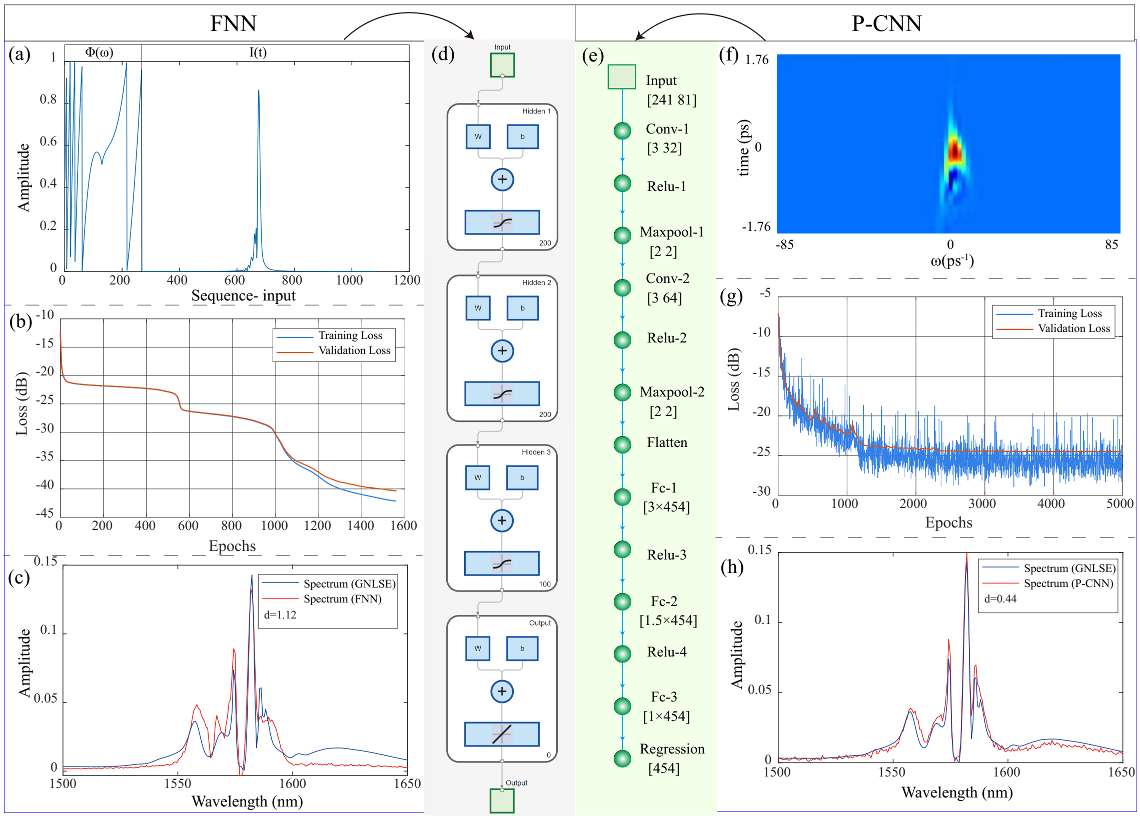

Figure S1: Comparison of feed-forward neural network (FNN) and convolutional neural network (CNN) for supercontinuum prediction. (a) Input parameters: spectral phase $\Phi(\omega)$ (left) and temporal intensity $A_{\text{in}}(t)$ (right). (b) Loss curves for FNN training and validation, reaching a minimum loss of approximately $-40\,\text{dB}$. (c) Comparison between GNLSE-calculated spectrum (blue) and FNN-predicted spectrum (red). (d) FNN architecture with three hidden layers containing 200, 200, and 100 neurons, respectively. (e) CNN architecture used for the P-CNN model. (f) Example of a Wigner function used as input to the CNN, corresponding to the initial conditions in panel (a). (g) CNN training loss curve, converging to $-24.5\,\text{dB}$ after 2000 epochs. (h) Comparison of the output spectrum predicted by the CNN (red) and the reference spectrum generated using the GNLSE (blue).

Figure S1(g) shows the training loss for the P-CNN, which stabilizes after several epochs, demonstrating excellent convergence behavior. In contrast, as shown in Fig. S1(b), the FNN does not achieve stable convergence despite reaching a lower minimum loss of approximately $-40\,\mathrm{dB}$. This behavior suggests that while the FNN can achieve smaller loss values, the P-CNN offers higher overall robustness and stability. Predictions of the spectral output further highlight the P-CNN's superior performance, as it captures finer spectral structures with smaller Euclidean distance errors, as shown in Fig. S1(c) and (h).

To further compare the performance of the FNN and P-CNN under noisy conditions, we introduce Gaussian noise into the input parameter space:

$$\{X_n'\} = \{X_n\} + \{\Delta X_n\} \cdot A_{\text{noise}} \cdot R(n), \tag{S20}$$

where $\{X_n\}$ represents a parameter set from the test samples, and $\{X_n'\}$ is the perturbed parameter set after adding noise. Here, $R(n)$ is a normally distributed random variable, generated using MATLAB's random number function, while A_{noise} controls the noise amplitude (set to 0.01). The boundary matrix $\{\Delta X_n\}$ is defined as in Eq. (S7), which denotes the range for the corresponding dispersion parameter.

Once the noisy parameters $\{X_n'\}$ are generated, we compute the perturbed spectral phase, temporal intensity, and corresponding Wigner function for the initial pulse. This allows us to analyze the noise sensitivity of predictions from the FNN and P-CNN.

To quantify the noise characteristics in the input parameter space, we calculate the normalized deviation as:

$$\text{Noise}_{\text{stat}} = \sum \frac{\{X_n'\} - \{X_n\}}{\{\Delta X_n\}}. \tag{S21}$$

The histogram of this noise statistic, shown in the inset of Fig. 2(c) in the main text, is well-fitted by a Gaussian distribution with a width of approximately 0.01.

We then evaluate the noise dependence of the trained networks by generating 1000 perturbed initial pulses. For each pulse, the output predictions from the FNN and P-CNN are obtained. To quantify the prediction errors under noise, we calculate the Euclidean distance between the noisy predictions and the noiseless target outputs.

The histogram of the Euclidean distances for the 1000 predictions is presented in Fig. 2(d) in the main text. The results demonstrate that the P-CNN outperforms the FNN, exhibiting smaller deviations and better robustness to input noise.

S2.5 Performance and Noise Evaluation of P-FNN

Although the P-CNN demonstrates superior performance compared to the conventional FNN, it is worth exploring a hybrid case where a fully connected neural network (FNN) is provided with physics-driven input—specifically, the Wigner distribution. To ensure a fair comparison using the same network architecture, we reshape the two-dimensional Wigner function (241×81), already calculated in the P-CNN section, into a one-dimensional vector.

Figure S2(a) shows an example of this 1D input vector, derived from the Wigner function shown in Fig. S1(f). In this process, margin regions with near-zero values are removed, and the data is downsampled to a size of 1×1098, matching the input dimension used in the FNN.

Figure S2(b) presents the training loss of the corresponding P-FNN, which achieves a lower loss compared to the baseline FNN shown in Fig. S1(b). Figure S2(d) displays the gradient (top panel) and validation checks (bottom panel) during the training process. The gradient decreases from approximately 1 to 1.259×10^{-5} at epoch 2754, indicating a clear convergence.

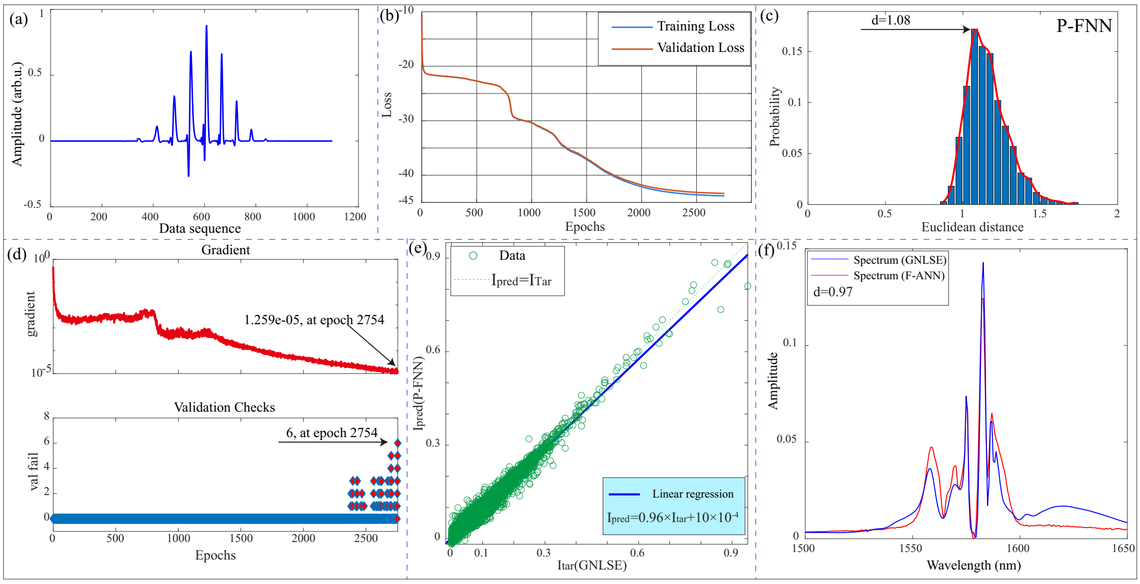

Figure S2: Performance and noise evaluation of the P-FNN model. (a) Example of the input data: a one-dimensional vector obtained by reshaping and downsampling the Wigner function. (b) Training and validation loss curves for the P-FNN model (blue: training, red: validation). (c) Statistical analysis of the Euclidean distance d for noise robustness evaluation. (d) Gradient (top) and validation checks (bottom) during training; early stopping is triggered after six failed validations. (e) Linear regression between the GNLSE-simulated outputs and the P-FNN predictions on the test set. (f) Comparison of the output spectrum predicted by the P-FNN (red) with the reference spectrum generated using the GNLSE (blue).

In the bottom panel, blue dots represent epochs with successful validation. Early stopping was triggered at epoch 2754 after 6 consecutive validation failures, as defined by the training validation-fail parameter 6, effectively monitor the possible overfitting. Figure S2(e) shows the linear regression between the predictions of P-FNN and the GNLSE results, yielding a regression coefficient of 0.96—an improvement over the 0.91 achieved by the purely data-driven FNN.

We also performed the similar noise robustness tests using the trained P-FNN under the same Gaussian noise conditions. The Euclidean distance d was computed as shown in Fig. S2(c), with a peak value of 1.08—lower than the FNN's value of approximately 1.17. Figure S2(f) displays one example of the spectral intensity prediction, where the blue curve represents the target spectrum and the red curve shows the P-FNN prediction. The corresponding Euclidean distance is 0.97, again smaller than the FNN's value of $\sim$1.12 in Fig. S1(c). While the P-FNN shows notable improvement over the standard FNN, the P-CNN still outperforms both, owing to its enhanced capacity to extract rich spectral-temporal features via its deep convolutional architecture.

S2.6 Generalization Performance of FNN, P-FNN, and P-CNN

To evaluate the generalization capability of the trained networks, we perform out-of-distribution tests by probing their response to input parameters beyond the original training space. Specifically, we select one of the phase parameters $\{X_n\}$ and apply a linear scan beyond its defined boundary, while keeping all other parameters fixed at zero. For example, we vary X_2—which

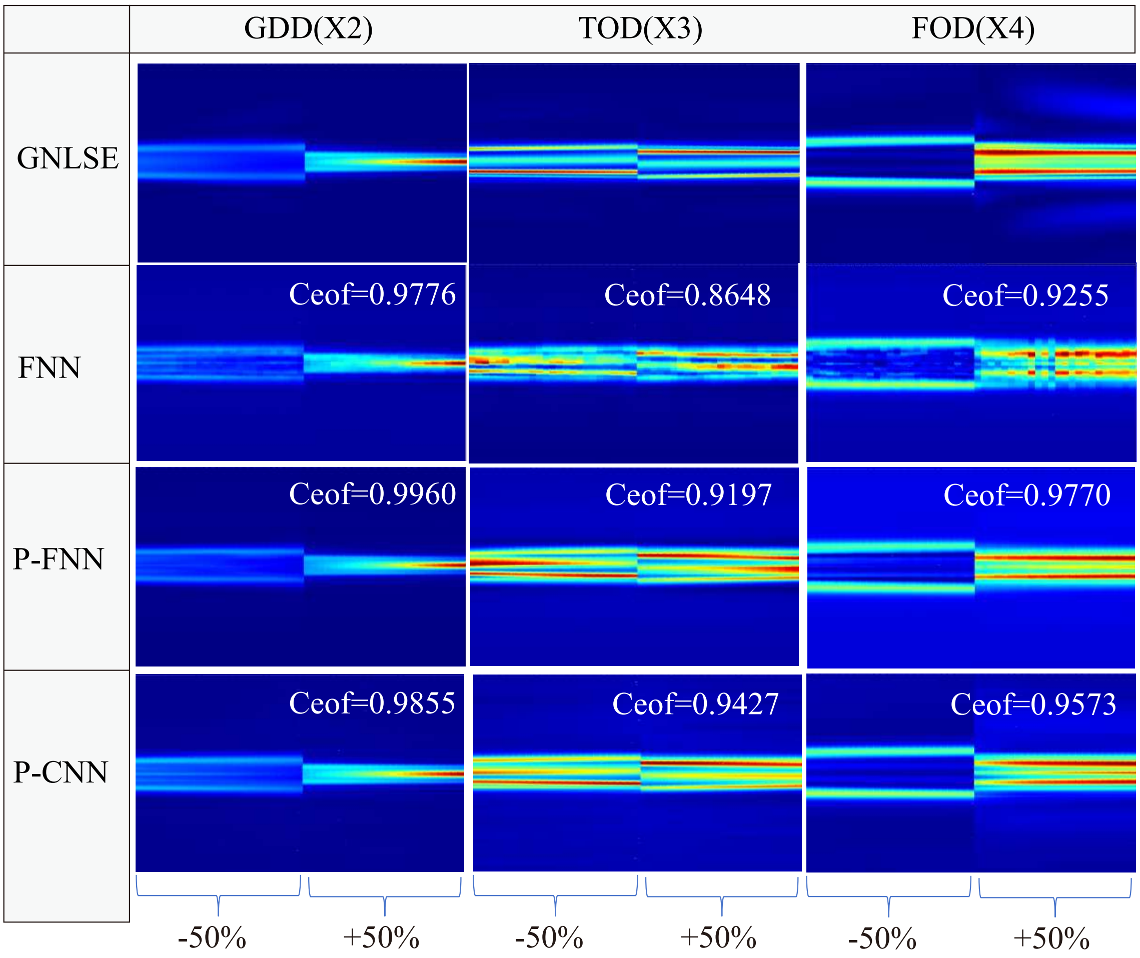

Figure S3: The simulations for spectra under GDD, TOD, and FOD scanning out of parameter space. The four rows in each column show the spectra generated by the GNLSE, FNN, P-FNN, and P-CNN, respectively. For each of panel, the x-axis represents the sampling sequence of the scanned parameter (50 samples), while the y-axis corresponds to wavelength, ranging from 1480 nm to 1640 nm.

governs the Group Delay Dispersion (GDD)—across an extended range:

$$X_2' = [(-150\%, -100\%); (100\%, 150\%)] \times \max(X_2), \tag{S22}$$

where the maximum value of X_2 is given in Eq. S18.

For each sampled value in this extended GDD range, we generate the corresponding spectrum using the GNLSE, which serves as the reference, and compare it with predictions from the trained FNN, P-FNN, and P-CNN models.

Figure S3 presents the spectral predictions under GDD, TOD, and FOD scanning. The x-axis represents the sampling sequence of the scanned parameter (50 samples), while the y-axis corresponds to wavelength, ranging from 1480 nm to 1640 nm. The four rows in each column show the spectra generated by the GNLSE (ground truth), FNN, P-FNN, and P-CNN, respectively. The first column corresponds to the GDD scan, while the second and third columns extend the same evaluation to TOD and FOD parameters.

While all three models produce reasonable predictions, the standard FNN introduces noticeable spectral noise, particularly in the TOD and FOD scans. In contrast, both the P-FNN and P-CNN demonstrate significantly improved agreement with the GNLSE spectra, benefited by the physics addressed Wigner function.

To quantify prediction accuracy under generalization, we calculate the Pearson correlation coefficient between the predicted spectra and the GNLSE reference. For two vectors A and B each of length N, the coefficient is defined as [17]:

$$\mathrm{Coef}(A,B) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{A_i-\mu_A}{\sigma_A}\right)\left(\frac{B_i-\mu_B}{\sigma_B}\right), \tag{S23}$$

where $\mu_{A,B}$ and $\sigma_{A,B}$ denote the mean and standard deviation of vectors A and B, respectively.

Here, A represents the reference spectrum from the GNLSE, and B corresponds to the network prediction. The calculated correlation coefficients are annotated within the figure panels. The results confirm that both P-FNN and P-CNN exhibit superior generalization performance compared to the baseline FNN, in particular, P-CNN gives $> 94\%$ crossing over all three cases.

S3 Searching strategy using P-CNN

S3.1 Searching strategy: P-CNN

Supplementary Fig. S4(a) outlines the P-CNN assisted approach for targeting specific physical states in supercontinuum generation, such as high-order solitons, depicted in the inset ‘Target state’. The process involves five key steps:

The P-CNN assistance for supercontinnum engineering

- 1. **Parameter Space Configuration**: Define the boundaries of the parameter space $\{X_n\}$ and create a two-dimensional image by calculating the Wigner function, utilizing reconstructions from the FROG system.
- 2. **P-CNN Training**: Train the Physics-embedded Convolutional Neural Network (P-CNN) within the defined parameter space, employing the Latin Hypercube Sampling (LHS) method to generate training data, that is $Y = \text{P-CNN}[X]$.
- 3. **Target Function Definition**: Specify the desired physical state as the target function, denoted as I_{tar}.
- 4. **Global Searching**: Identify several global solutions by minimizing the error function between the target and the predictions from the trained network, expressed as $|I_{\text{tar}} - Y|$.
- 5. **Local Searching**: Refine these global solutions within the nonlinear system using iterative methods, specifically the gradient descent algorithm, to achieve the optimal solution.

Supplementary Fig. S4(b) illustrates the nonlinear complex system under consideration and its training framework. Once the network is trained, the inverse searching algorithm is used to identify the minimum solution. However, to prevent potential divergences, we actively search for the best solution directly within the real system. Initially, three solutions in proximity to the minimum, referred to as ‘global searching’ (step 4 in the box), are selected based on the desired target state. These solutions are then tested in the actual system to determine the optimal one, based on error metrics, employing the gradient descent algorithm (step 5 in the box) [18]. This step is crucial to circumvent specific solutions that may not converge on the trained network, ensuring the reliability and accuracy of the results.

Supplementary Fig. S4 (c) displays 20 spectra derived from simulations using the General Nonlinear Schrödinger Equation (GNLSE) alongside predictions made by the trained P-CNN shown in the last section, focusing specifically on the soliton wave range. The comparison reveals a high degree of similarity between the simulated and predicted spectra, underscoring the effectiveness of the P-CNN in modeling complex nonlinear dynamics within this specific spectral range.

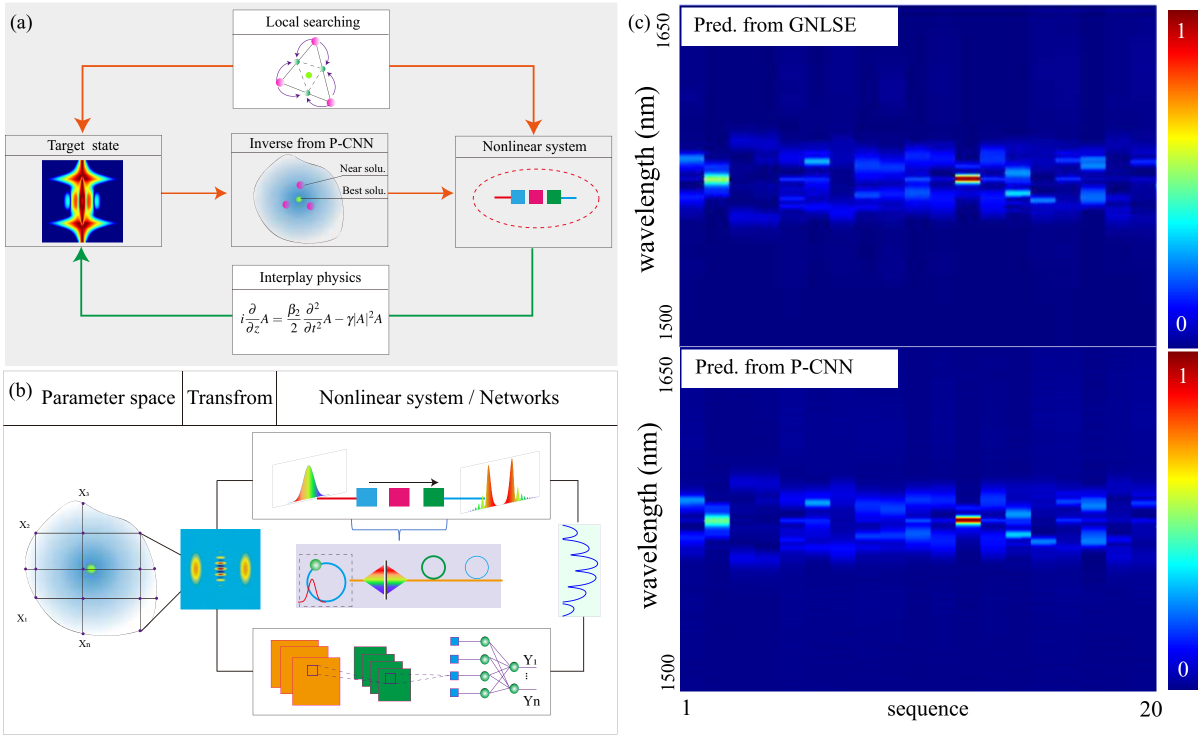

Figure S4: Schematic for P-CNN assistance in Supercontinuum Engineering. (a) Overall Process for Target State Identification: Illustrates the method of target physical states within a nonlinear system using the P-CNN. 'Inverse from P-CNN' refers to the inverse solutions derived from the trained network. 'Local searching' denotes the iterative refinement process within the optical setup, and 'interplay physics' refers to the dynamics governed by the Nonlinear Schrödinger Equation. (b) Framework for the Nonlinear System and P-CNN: Outlines the components involved in the analysis, including the parameter space (dispersion), transformation process (Wigner function), and the nonlinear complex system characterized by CNN. (c) Spectrum Comparisons: Displays spectra generated from simulations using the General Nonlinear Schrödinger Equation (GNLSE) alongside those predicted by the trained P-CNN.

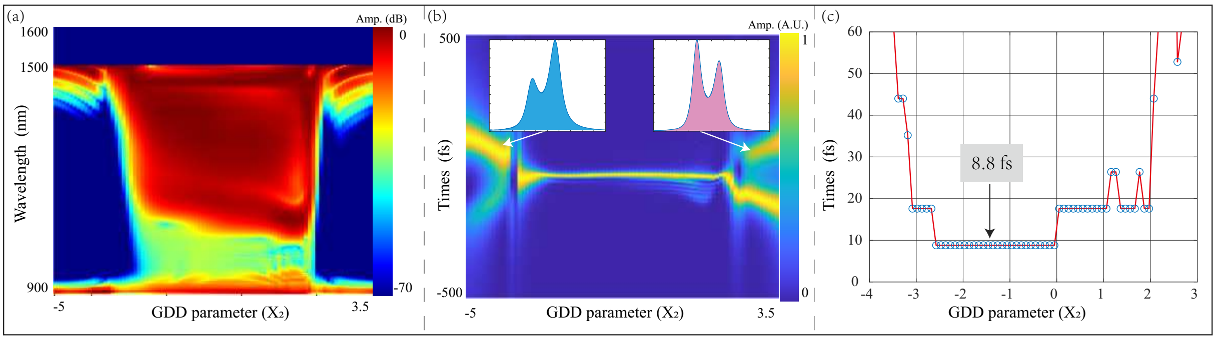

Figure S5: (a) Displays the spectrum obtained from simulations for scanning the Group Delay Dispersion (GDD) phase. The Group Velocity Dispersion (β_2) is fixed at $-21\times10^{-3}\mathrm{ps}^2/\mathrm{m}$, and the dispersion parameter X_2 is varied from -5 to 3.5 m. (b) Shows the corresponding temporal profiles obtained by performing a Fourier transform on the calculated spectrum, limited to the wavelength range of 900 to 1500 nm. (c) Illustrates the range of extracted pulse durations as the GDD dispersion parameter varies from -4 to 3 m.

S4 Simulations for the few-cycle pulse engineering

Using the theoretical model described in **Method-1** and the section of S2, we can examine the effect of the dispersion, such as Group Velocity Dispersion (GVD), on the pulse duration of the temporal pulse. To align with the experimental setup, a short-pass filter is used to exclude wavelengths longer than 1500 nm.

Supplementary Fig. S5(a) illustrates the spectrum in simulation for scanning the Group Delay Dispersion (GDD) phase. We fix β_2 at -21×10^{-3} ps^2/m, and scan X_2 (defined in Eq. S5) from -5 to 3.5 m, while keeping other parameters fixed at zero. Parts of the spectrum exceeding 1500 nm are removed as previously mentioned. By performing a Fourier transform on these spectral fields, we obtain the temporal profiles displayed in Fig. S5(b). Notably, the pulse typically exhibits a single lobe at the center of the GDD dispersion parameter X_2. However, for larger absolute values of X_2, two lobes appear, consisting of one larger and one smaller lobe in amplitude. The smaller lobe is located on the left and right sides of the time center 0, for negative and positive X_2 values, respectively. This asymmetry arises from the presence of fringes in the spectrum, as depicted in Fig. S5(a). Consequently, the temporal profiles create a systematic effect akin to a ‘temporal inversion’, consistent with experimental observations.

The range of extracted pulse durations from GDD dispersion parameter scanning from -4 to 3 m is shown in Supplementary Fig. S5(c). The minimum pulse duration can reach as low as 8.8 fs, while the maximum pulse durations are around 60 fs. These simulation results align well with the observations, highlighting the symmetries also in the pulse durations.

S5 Pulse reconstructions by Frequency-Resolved Optical Gating (FROG)

S5.1 Collinear FROG: soliton pulse

To observe the temporal profiles, such as temporal intensity for the soliton pulse, we conducted Second Harmonic Generation Frequency-Resolved Optical Gating (SHG-FROG) measurements on the corresponding pulses. Given the low pulse energy output from the Mode-Locked Fiber Laser (MLF), we employed an efficient collinear FROG setup. This setup utilizes a type-II BBO (Beta Barium Borate) crystal (o+e→e) with a length of 1 mm to enhance measurement sensitivity and accuracy [19, 8].

Supplementary Fig. S6 (a) presents the measurement results for the soliton pulse generated by the custom MLF [7]. This figure includes five panels arranged from left to right, showcasing the recorded spectrum, measured and reconstructed FROG traces, reconstructed spectral phase, and the pulse temporal profile. The measured spectral width (half-width intensity) is approximately 4.16 nm, and the reconstructed pulse duration is about 625.5 fs.

Supplementary Fig. S6 (b) details the outcomes for the laser pulse post-amplification through EDFA-1, as shown in Fig. 5. This stage serves two primary functions: firstly, to boost the pulse energy, and secondly, to broaden the pulse spectrum through a process akin to a self-similar amplifier [20], primarily facilitated by self-phase modulation. Following this amplification, the measured spectral bandwidth reaches approximately 6.8 nm, and the reconstructed pulse duration is about 400.8 fs.

Supplementary Fig. S6 (c) shows the characteristics of the pulse after further processing by the second EDFA, optimized with the pulse shaper and a section of SMF, L_3 from Fig. 5. At this point, the measured pulse width is around 69.5 fs. Additionally, the average pulse power at this stage is approximately 180 mW with a repetition frequency of 34 MHz, corresponding to a peak power of about 76 kW and a pulse energy of roughly 5.3 nJ.

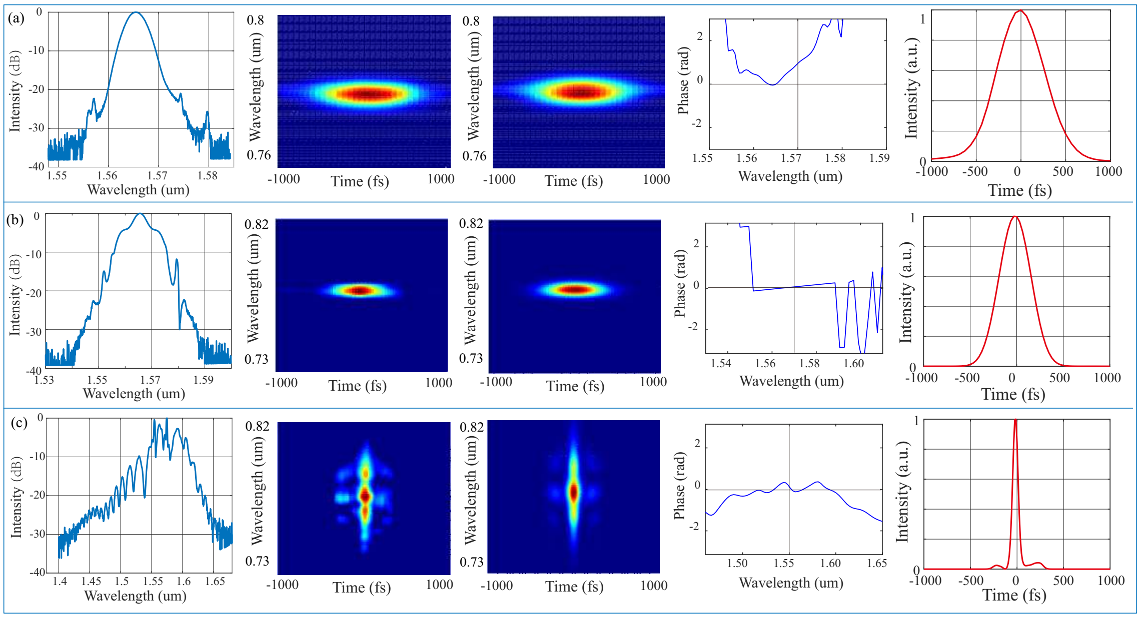

Figure S6: The measured and reconstructed Spectrum and Temporal Profile Using FROG System. (a) Output Soliton Pulse from Mode-Locked Fiber Laser: Displays the measured results. The first panel shows the recorded spectrum. The second and third panels depict the measured and reconstructed FROG traces, respectively. The fourth panel presents the reconstructed spectral phase. The final panel illustrates the normalized pulse profile, with a pulse width of approximately 625.5 fs. (b) Soliton Pulse Post-EDFA-1 (before the pulse shaper): Represents the stage before the pulse shaper in Fig. 5. The final panel shows the reconstructed pulse width of approximately 400.8 fs. (c) Laser Pulse at Position A4 (before the high nonlinear fiber) in Fig. 5: Displays the measured results with a pulse width of around 69.5 fs.

S5.2 Non-collinear FROG: few-cycle pulse

Supplementary Fig. S7 (a) depicts the setup of the SHG-FROG system in a non-collinear geometry, as opposed to the one mentioned in the last section of S5-1, which was set in a collinear geometry. This system incorporates a short-pass filter (3.2 mm thickness) to exclude wavelengths beyond 1500 nm, two beam splitters (BS) (1 mm thickness each) in an interferometer setup to split and recombine the incoming pulses. The beam is then focused by a lens ($f = 100$ mm, 2 mm thickness) into a nonlinear BBO crystal (type-I, $\theta = 19.8°$, 1 mm thickness). To enhance signal-to-noise ratio and nonlinear efficiency, the angle between the two beams is set to approximately $5°$. Additional lenses focus the Second Harmonic Generation (SHG) pulse beam into a spectrometer (Ocean Optics HR 2000) for spectral measurements. A movable electric translation stage on one arm of the interferometer allows for precise adjustments of the relative delay between the two beams before reaching the BBO crystal.

Considering the total optical path, the combined thickness amounts to approximately 8.2 mm, corresponding to a GDD of about 120 fs^2, with a second-order dispersion coefficient $\beta_2 \approx 15\mathrm{fs}^2/\mathrm{mm}$. The reconstructed GDD from the FROG traces for the pulse with the minimum duration approximates a second-order distribution, with a fitted GDD of about 100 fs^2.

Supplementary Fig. S7 (b)-(c) present the measurement and reconstruction results for the few-cycle laser pulse under various GDD dispersion phases (X_2) of 2 and -2 m, respectively. Each panels in (c) or (d) displays the measured and reconstructed FROG traces, reconstructed spectrum and phase, and temporal profiles. The reconstructed pulse durations are approximately 20 fs for both positive and negative GDD settings, and about 12 fs when X_2 is set to zero, marking the minimum pulse duration observed. The spectral phase and profile in the third panel show similar intensity profiles, but with inverted chirp characteristics based on the spectral phase profiles. During these reconstructions, a standard generalized-projections (GP) iterative algorithm is used to retrieve the spectral phase and amplitude [21].

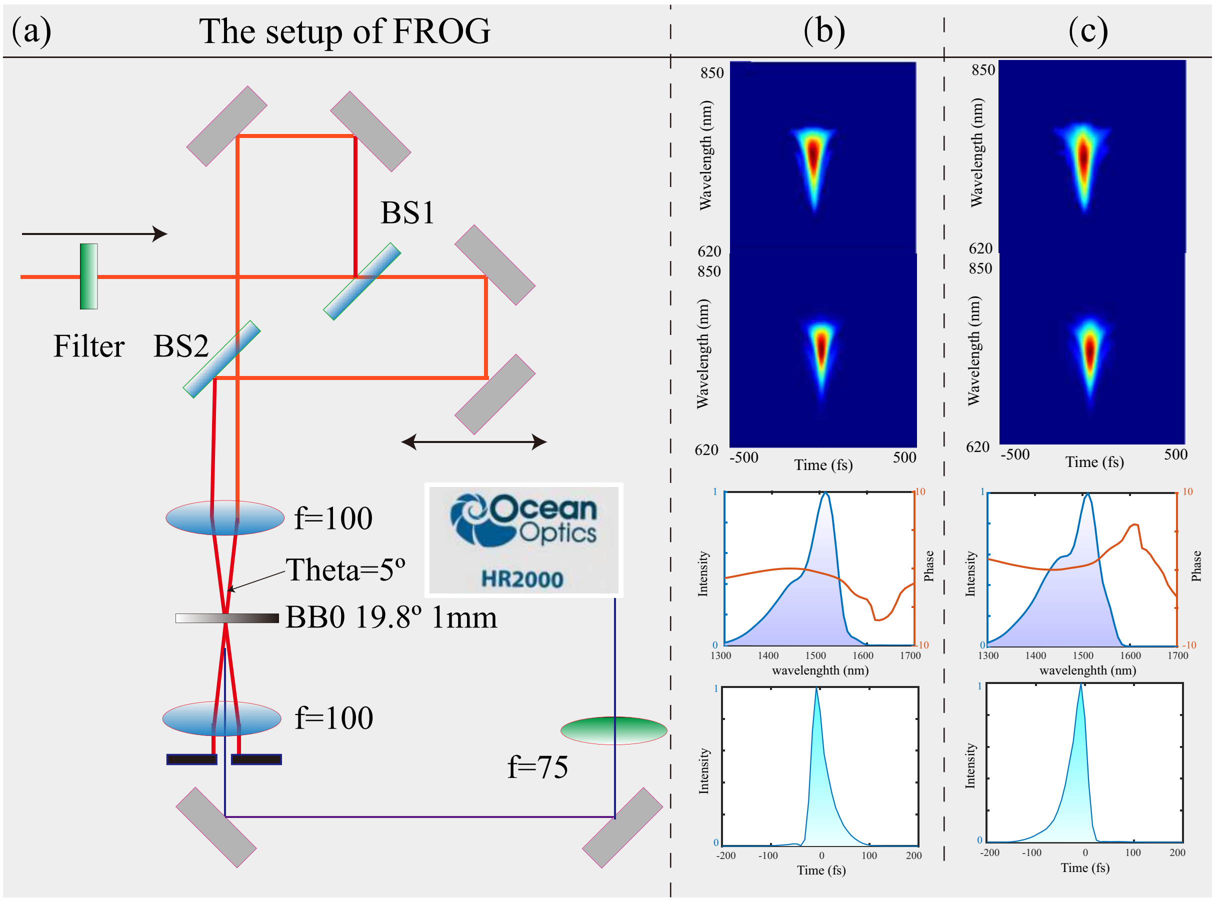

Figure S7: (a) The experimental setup of SHG-FROG measurement. (b)-(c) The measured and reconstructed traces for the few cycle pulse when the pulse shaper carries the GDD dispersion parameter X_2 of 2 and -2 m, respectively. In (b)-(c), four panels show the measured, reconstructed FROG trace, reconstructed spectrum and phase, as well as the temporal profile, respectively.

S6 The theoretical models and simulations for high-order soliton

The temporal pulse dynamics through a nonlinear medium such as optical fiber can be effectively modeled by the Nonlinear Schrödinger Equation (NSE), which accounts for both dispersion (β_2) and nonlinearity (γ):

$$-i\frac{\partial E(t)}{\partial z}=\frac{\beta_2}{2}\frac{\partial^2}{\partial t^2}E(t)+\gamma|E(t)|^2E(t). \tag{S24}$$

High-order solitons emerge when there is an imbalance between dispersion and nonlinearity, particularly at the initial pulse with high pulse energies. To quantify this, we define several key parameters [10]:

$$\begin{aligned} L_\mathrm{d}&=\frac{T_0^2}{|\beta_2|}, && \text{(Dispersion Length)}\\ L_\mathrm{n}&=\frac{1}{\gamma P_\mathrm{k}}, && \text{(Nonlinear Length)}\\ z_0&=\frac{\pi L_\mathrm{d}}{2}, && \text{(Soliton Period)}\\ N^2&=\frac{L_\mathrm{d}}{L_\mathrm{n}}, && \text{(Soliton Number)} \end{aligned} \tag{S25}$$

where P_k is the peak power of the initial pulse, and T_0 is the pulse duration. The initial field is set as $E_0(t)=\sqrt{P_\mathrm{k}}\,\mathrm{sech}(T/T_0)$. By carefully adjusting P_k to align with a specific soliton number N, the dynamics of N-order solitons can be observed within the framework of the NSE.

Supplementary Fig. S8(a) captures the spectral dynamics for $N=3$ over 1.5 soliton periods, showcasing the characteristic 'breathing' phenomenon along the propagation distance z.

To consistently produce high-order solitons within the same fiber length, one method involves modulating the soliton period z_0 by adjusting the initial pulse duration T_0, which directly alters the dispersion length L_d. However, simultaneously, the peak power P_k and the nonlinear length L_n must also be adjusted to maintain a consistent soliton number N, which complicates experimental setups where only pulse duration is varied. Nevertheless, this challenge can be effectively addressed using deep learning networks in current experiments.

To examine the spectral dynamics of high-order solitons from a theoretical perspective, both the pulse duration and peak power are varied to fine-tune the dispersion and nonlinear lengths, respectively, ensuring the soliton number remains constant at $N=3$. Supplementary Fig. S8(b) demonstrates a periodic scanning of dispersion and nonlinear lengths, maintaining a soliton number of $N=3$ as depicted in Supplementary Fig. S8(d). This configuration results in the output spectrum exhibiting characteristic 'breather' behavior at consistent propagation lengths.

Supplementary Fig. S8(c) displays the dynamics over 90 sequence sampling points, where the observed dynamics align closely with the predictions shown in Supplementary Fig. S8(a), confirming the model's accuracy and the efficacy of the deep learning network in controlling complex soliton dynamics.

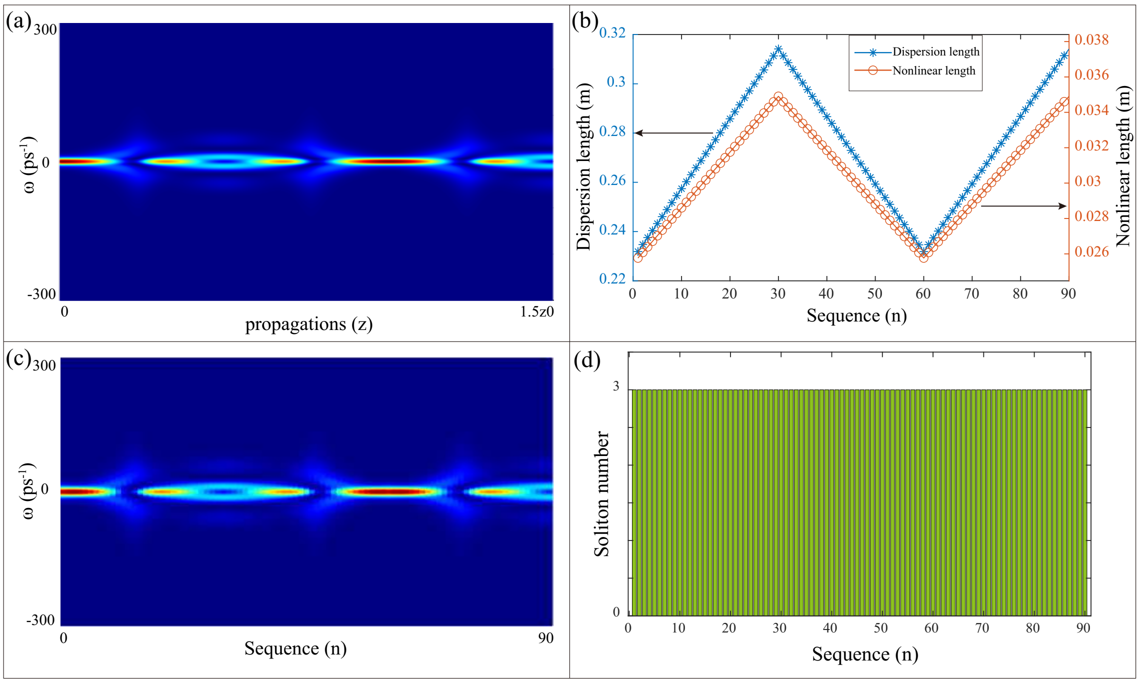

Figure S8: (a) Displays the dynamics in spectrum of a third-order soliton ($N = 3$) over 1.5 soliton periods, highlighting characteristic ‘breathing’ patterns. (b) Illustrates variations in dispersion length (L_d) and nonlinear length (L_n), tailored to maintain consistent soliton number. (c) Shows the dynamics of the third-order soliton’s spectrum across 90 sequential sampling points. (d) Depicts the soliton number (N) changing in dispersion and nonlinear lengths at the same time.

References

[1] Antoine Monmayrant, Sébastien J Weber, and Béatrice Chatel. A newcomer's guide to ultrashort pulse shaping and characterization. *J. Phys. B*, 43(10):103001, 2010.

[2] Andrew M Weiner. Ultrafast optical pulse shaping: A tutorial review. *Opt. Commun.*, 284(15):3669–3692, 2011.

[3] Boris A Malomed. Optical solitons and vortices in fractional media: A mini-review of recent results. In *Photonics*, volume 8, page 353. Multidisciplinary Digital Publishing Institute, 2021.

[4] Shilong Liu, Yingwen Zhang, Boris A Malomed, and Ebrahim Karimi. Experimental realisations of the fractional schrödinger equation in the temporal domain. *Nature Communications*, 14(1):222, 2023.

[5] Eliot Bolduc, Nicolas Bent, Enrico Santamato, Ebrahim Karimi, and Robert W Boyd. Exact solution to simultaneous intensity and phase encryption with a single phase-only hologram. *Opt. Lett.*, 38(18):3546–3549, 2013.

[6] Shi-Long Liu, Qiang Zhou, Shi-Kai Liu, Yan Li, Yin-Hai Li, Zhi-Yuan Zhou, Guang-Can Guo, and Bao-Sen Shi. Classical analogy of a cat state using vortex light. *Communications Physics*, 2(1):1–9, 2019.

[7] Shilong Liu, Yudong Cui, Ebrahim Karimi, and Boris A Malomed. On-demand harnessing of photonic soliton molecules. *Optica*, 9(2):240–250, 2022.

[8] Shilong Liu, Yingwen Zhang, Stéphane Virally, Ebrahim Karimi, Boris A Malomed, and Denis V Seletskiy. Observation of the spectral bifurcation in the fractional nonlinear schr\"{o} dinger equation. *arXiv preprint arXiv:2311.15150*, 2023.

[9] Michael D McKay, Richard J Beckman, and William J Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1):55–61, 2000.

[10] Govind P Agrawal. Nonlinear fiber optics. In *Nonlinear Science at the Dawn of the 21st Century*, pages 195–211. Springer, 2000.

[11] Keith J Blow and David Wood. Theoretical description of transient stimulated raman scattering in optical fibers. *IEEE Journal of Quantum Electronics*, 25(12):2665–2673, 1989.

[12] Ilaria Cristiani, Riccardo Tediosi, Luca Tartara, and Vittorio Degiorgio. Dispersive wave generation by solitons in microstructured optical fibers. *Opt. Express*, 12(1):124–135, 2004.

[13] Aly-Khan Kassam and Lloyd N Trefethen. Fourth-order time-stepping for stiff pdes. *SIAM Journal on Scientific Computing*, 26(4):1214–1233, 2005.

[14] Valerio Lucarini, Jarkko J Saarinen, Kai-Erik Peiponen, and Erik M Vartiainen. *Kramers-Kronig relations in optical materials research*, volume 110. Springer Science & Business Media, 2005.

[15] Govind P Agrawal and MJ Potasek. Nonlinear pulse distortion in single-mode optical fibers at the zero-dispersion wavelength. *Phys. Rev. A*, 33(3):1765, 1986.

[16] Thomas P Vogl, JK Mangis, AK Rigler, WT Zink, and DL Alkon. Accelerating the convergence of the back-propagation method. *Biological cybernetics*, 59:257–263, 1988.

[17] Maurice George Kendall. The advanced theory of statistics. 1948.

[18] Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[19] Ian A Walmsley and Christophe Dorrer. Characterization of ultrashort electromagnetic pulses. *Adv. Opt. Photonics*, 1(2):308–437, 2009.

[20] FÖ Ilday, JR Buckley, WG Clark, and FW Wise. Self-similar evolution of parabolic pulses in a laser. *Phys. Rev. Lett.*, 92(21):213902, 2004.

[21] Rick Trebino, Kenneth W DeLong, David N Fittinghoff, John N Sweetser, Marco A Krumbügel, Bruce A Richman, and Daniel J Kane. Measuring ultrashort laser pulses in the time-frequency domain using frequency-resolved optical gating. *Rev. Sci. Instrum.*, 68(9):3277–3295, 1997.