Research Article Vol. 33, No. 25 / 15 Dec 2025 / Optics Express 53154
Optics EXPRESS

Cascaded LSTM with downsampling strategy for fast and accurate modeling of chirped pulse amplification

XINYI MA,† GUOQING PU,*,† HELIN JIANG,† WEISHENG HU, AND LILIN YI

State Key Laboratory of Photonics and Communications, School of Information Science and Electronic Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China
†These authors contributed equally.
*teddyghf1994@sjtu.edu.cn

Abstract: Conventional iterative numerical modeling of chirped pulse amplification (CPA) systems requires a large temporal simulation window to accommodate the heavily chirped pulses. Simultaneously, maintaining the high temporal resolution certainly improves the computational complexity, posing challenges for CPA system design and optimization based on numerical simulations. To overcome this limitation, we propose a cascaded long-short-term memory (LSTM) model with a downsampling strategy trained for efficient and accurate modeling of a multi-stage optical fiber system. This approach delivers full-field simulation of the heavily chirped pulse with 10-nm spectral bandwidth at the pulse energy reaching 14.9 μJ. Through aggressive downsampling in the time domain, the proposed framework reduces the computational complexity by 929 times and achieves a remarkable 1,564-fold speedup compared to conventional numerical simulations, while maintaining prediction errors of the pulse energy and duration below 2%. Our work provides an efficient and high-fidelity CPA systems modeling alternative, which is particularly suitable for the inverse design and optimization of CPA systems for high-energy short pulses generation.

© 2025 Optica Publishing Group under the terms of the Optica Open Access Publishing Agreement

1. Introduction

High-energy femtosecond pulses generated in CPA systems play a critical role in diverse fields [1], ranging from precision industrial manufacturing [2,3], biotechnology [4,5], and metrology [6,7]. In CPA systems, optimizing the output quality requires a thorough understanding of the dynamics of high-energy femtosecond pulses [8], which can be studied through numerical simulations. Conventional modeling of CPA systems relies on the iterative numerical approach solving the nonlinear Schrödinger equation (NLSE) and the rate equations (REs). The computational inefficiency of numerical simulations hinders the inverse design and optimization of CPA systems [9,10]. In laser optics, artificial intelligence (AI) has emerged as a powerful tool for modeling and optimizing. The propagation of high-order solitons and supercontinuum generation in fibers is accurately simulated with an LSTM model, achieving a speedup of several orders of magnitude compared to the conventional split-step Fourier method (SSFM) [11]. Subsequently, an LSTM-based approach was proposed to enable fast, accurate, and full-field modeling of mode-locked fiber lasers [12]. Further, AI is applied in designing and optimizing photonic crystal structures by building up a multi-level abstraction of massive data [13,14].

However, applying AI to model CPA systems still presents significant challenges. Compared to the nonlinear dynamics in passive fiber, the evolution dynamics in CPA are more complex due to the deeply involved interplay among gain, dispersion, and nonlinearities. Meanwhile, due to the large prior dispersion introduced by the stretcher, a ns-level temporal simulation window is necessary to prevent overlap, and a femtosecond-level temporal resolution is required to reveal

#584849 https://doi.org/10.1364/OE.584849
Journal © 2025 Received 14 Nov 2025; revised 30 Nov 2025; accepted 1 Dec 2025; published 9 Dec 2025

Optics EXPRESS

the detailed characteristics of pulses before stretching and after compression. Therefore, the number of sampling points of the simulation can be quite large. Both spatial and temporal complexity of the simulation scales as the dispersion of the stretcher becomes larger, and our previous studies circumvent the problem. For instance, we built an LSTM model to simulate a CPA system with a 300 ps time window and a 200 fs resolution, achieving an output with merely 400 nJ energy and 266 fs pulse duration [15]. For modeling CPA systems with output energy above the microjoule level, the number of sampling points must be increased to accommodate spectral broadening induced by strong nonlinear effects. Under such circumstances, the vanilla LSTM model is no longer applicable, as the large input time series leads to insufficient GPU memory. Consequently, an end-to-end model solved this problem by only modeling the pulse before the stretcher and after the compressor, which reduced the temporal window from 4 ns to 2 ps and enabled the simulation of CPA systems with a high temporal resolution of 20 fs [14]. However, the end-to-end model fails to manifest the evolution dynamics of pulses in gain fibers, functioning essentially as a complete "black box".

To address the aforementioned problem, we propose a cascaded LSTM model with the temporal downsampling strategy to reduce the computational complexity of CPA systems modeling. By separately downsampling the pulse amplitude and unwrapped phase, the redundancy in temporal resolution is reduced while ensuring ultra-high precision recovery. We validate the proposed method in simulating a 3-stage CPA system with the temporal simulation window beyond 1 ns and the temporal resolution of 30 fs. Simulation results indicate that the cascaded LSTM model with a downsampling strategy resolves the issue of insufficient memory, and the speedup can reach up to 1,564 times compared to the conventional numerical approach. Additionally, the grouped parameter optimization (GPO) strategy is employed to reduce prediction errors, resulting in an overall prediction energy error of less than 2% throughout the entire pulse evolution process. Learnable parameter encoding is applied to enhance the generalization ability of the model.

2. Principles

2.1. CPA simulation system

We construct a 3-stage amplifier CPA simulation system, with its structure depicted in Fig. 1, and its critical parameters elaborated in Table 1. Notably, the pulse duration expands to over 300 ps after the stretching process, whereas after the compression process, the output pulse duration is approximately 300 fs, and the output energy reaches up to 15 μJ.

2.2. Conventional numerical simulation method

The ultrafast pulse amplification process in the gain fiber involves linear and nonlinear effects, and photon energy level dynamics, which are described by two physical equations: the NLSE and the REs [16]. The NLSE is numerically solved using the SSFM, while the fourth-order Runge-Kutta method (RK4) is employed to solve the REs. The NLSE models the evolution of the complex electric field envelope of light in a passive fiber [17]. Dispersion and nonlinear effects, including self-phase modulation, self-steepening, and stimulated Raman scattering, are given by

$$\frac{\partial A}{\partial z}+\frac{i\beta_2}{2}\frac{\partial^2 A}{\partial t^2}-\frac{\beta_3}{6}\frac{\partial^3 A}{\partial t^3}+\frac{\alpha}{2}A = i\gamma\left[|A|^2+\frac{i}{\omega_0}\frac{\partial}{\partial t}|A|^2A-T_RA\frac{\partial|A|^2}{\partial t}\right]A. \tag{1}$$

where A is the optical field envelope, ω_0 is the central angular frequency, z denotes propagation distance, and t represents time. Parameters β_2, β_3, α, and γ correspond to the second-order dispersion, third-order dispersion, loss, and nonlinear coefficient, respectively. T_R relates to the slope of the Raman gain.

Within the gain fiber, wavelength-dependent gain arises from temporal dynamics of the energy level population. The REs account for absorption, stimulated emission, and spontaneous emission

Optics EXPRESS

Table 1. Parameters of the components in the 3-stage amplifiers CPA.

Component	Property	Value
Seed	Central wavelength	1030 nm
	Duration	3 ps
	Energy	4 nJ
	Repetition rate	25 MHz
	Bandwidth	8.5 nm
CFBG	Bandwidth	100 nm
PM Yb-SMF 6/125	Length	2 m
	Core Absorption	250.0 dB/m@975 nm
AOM	Output Repetition rate	1 MHz
PM Yb-DCF 14/125	Length	2 m
	Clad Absorption	16.6 dB/m@976 nm
PM Yb-DCF 20/130	Length	2 m
	Clad Absorption	10.20 dB/m@976 nm
Grating pairs	Line number	1000
	Angle of incidence	30°

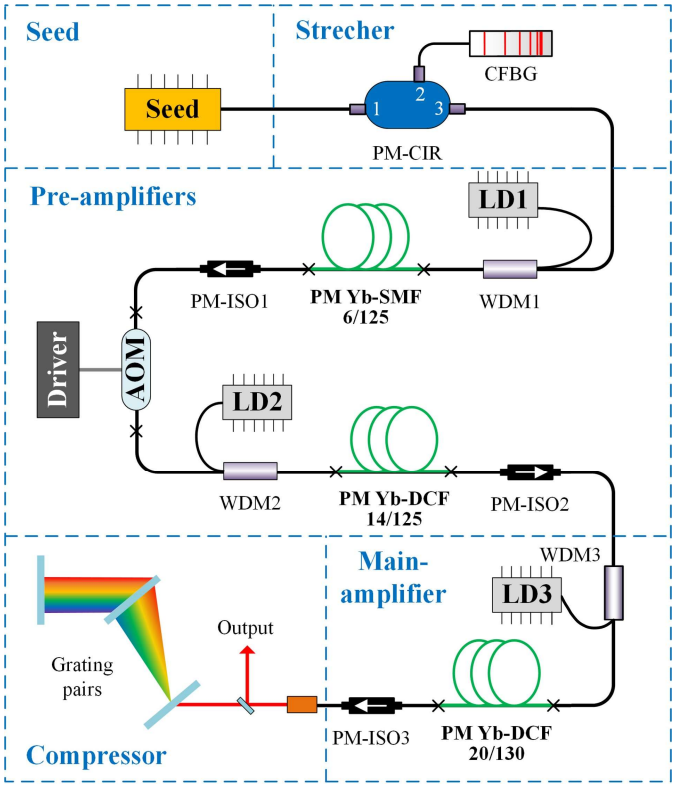

Fig. 1. Schematic setup of 3-stage amplifiers CPA system, including seed, stretcher amplifiers, and compressor. CFBG, chirped fiber Bragg grating; PM-CIR, polarization maintaining circulator; LD, laser diode; WDM, wavelength division multiplexer; Yb-SCF, Ytterbium-doped single-clad fiber; PM-ISO, polarization maintaining isolator; AOM, acousto-optic modulator; Yb-DCF, Ytterbium-doped double-clad fiber.

Optics EXPRESS

in the quasi-three-level Yb system. These equations describe the steady state of the full dynamics of the population inversion [18]:

$$N_2(z) = \frac{\frac{1}{hcA_c}\sum_{k=1}^{K}\Gamma_k\lambda_k\sigma_a(\lambda_k)N_T P_k^{+}(z)}{\frac{1}{hcA_c}\sum_{k=1}^{K}\Gamma_k\lambda_k[\sigma_a(\lambda_k)+\sigma_c(\lambda_k)]P_k^{+}(z)+\frac{1}{\tau}}. \tag{2}$$

$$\begin{aligned}\frac{dP_k^{\pm}(z)}{dz} = {} & \pm(\Gamma_k[(\sigma_c(\lambda_k)+\sigma_a(\lambda_k))N_2(z)-\sigma_a(\lambda_k)N_T]P_k^{\pm}(z)-\eta_k P_k^{\pm}(z) \\ & +\Gamma_k\sigma_e(\lambda_k)N_2(z)\frac{2hc^2\Delta\lambda}{\lambda_k^3}.\end{aligned} \tag{3}$$

$$N_T = N_1(z,t)+N_2(z,t). \tag{4}$$

z, t, and λ represent spatial position, time, and wavelength, respectively. Index k corresponds to pump and signal components. h is Planck's constant, c is the vacuum speed of light, τ is the upper-state lifetime, and N_T is the total dopant density with $N_1(z,t)$ and $N_2(z,t)$ being the populations of the lower and upper energy states. $P^{\pm}$ denotes optical power (+: forward, −: backward). σ_a and σ_e are absorption and emission cross-sections, Γ_k is the geometric overlap factor, and $\Delta\lambda$ is the wavelength resolution.

Early numerical models effectively describe continuous-wave amplification but neglect critical nonlinear effects and dispersion in ultrafast pulse regimes [19]. Other common workarounds apply NLSE, assuming wavelength-independent gain, which overlooks spatially and spectrally varying gain in fiber amplifiers [20,21]. Unlike the separate model fails to capture the dynamic interplay between gain, dispersion, and non-linearity [22], Lindberg et al. introduce a combined numerical model integrating with RE and NLSE [16]. As a result, the gain profile calculated from RE at each fiber position determines subsequent dispersion and nonlinear effects in the NLSE. The pulse propagation then updates the pulse profile, which in turn influences the subsequent gain dynamics. The training data is generated using the combined numerical model proposed in [16] with a simulation time window of 1024 ps and the temporal resolution of ~30 fs, resulting in 32,768 points per pulse.

2.3. Cascaded LSTM with downsampling strategy

The parameter disparities among various gain fibers constitute the primary factor contributing to the distinct dynamic evolution of pulses. For instance, variations in doping concentration dictate vastly different gain dynamics, while fiber radius governs the scale of nonlinear effects. Consequently, a single unified LSTM model faces challenges in characterizing a multi-stage optical fiber system, prompting the proposal of a cascaded LSTM. As illustrated in Fig. 2, each LSTM module with identical model structure is independently trained and optimized. Because the direct downsampling of complex-field pulses would cause irreversible information loss, the selection of an optimal downsampling strategy plays a decisive role in enhancing simulation efficiency for full-field CPA modeling (i.e., outputting the amplitude and phase simultaneously). The temporal pulse amplitude and unwrapped phase exhibit quasi-parabolic evolutionary characteristics due to the substantial dispersion introduced by the CFBG causes. This property enables effective signal recovery through cubic interpolation after independently down-sampling on temporal amplitude and unwrapped phase. Since the unwrapped phase value is much greater than the pulse amplitude by orders of magnitude, the initial phase with CFBG dispersion parameters set to [20 ps^2, 0.1 ps^3] is used as the phase offset. The phase of the pulse simulated at each step needs to be subtracted by this phase offset, which means that the cascaded LSTM model predicts the pulse amplitude and phase variation, enabling it to better learn fine-grained phase information.

Optics EXPRESS

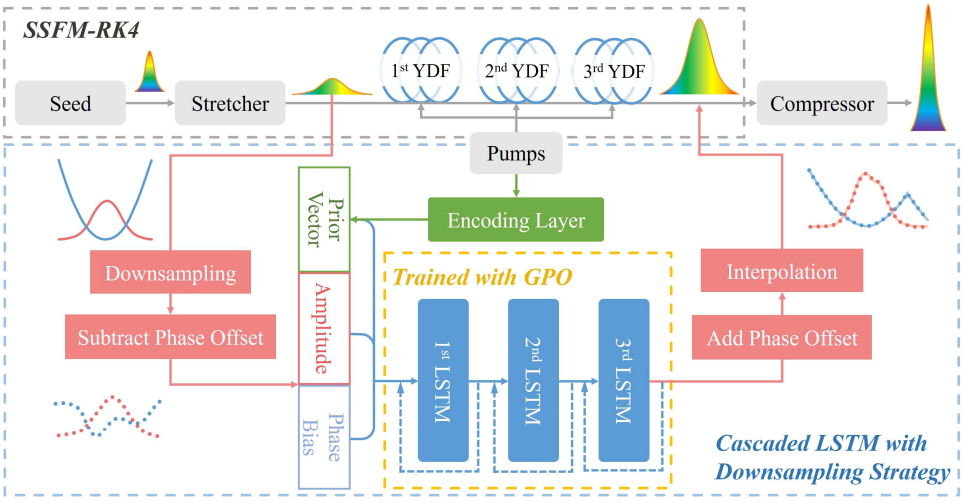

Fig. 2. The proposed cascaded LSTM for simulating the 3-stage CPA system, where three identical LSTM models correspond to three different Ytterbium-doped fiber (YDF). Temporal pulse amplitude, unwrapped phase (after phase offset subtraction), and a prior vector (from pump power via encoding layers) are concatenated as LSTM input.

The generalization ability of data-driven models is a huge challenge. For ultrafast pulse propagation, different input conditions lead to different nonlinear evolution maps. Here, the cascaded LSTM is trained with data generated under different chirp and pump conditions, as illustrated in Table 2. To enhance the generalization ability of the model, the power of 3-stage pumps is input into the encoding layers to generate the prior vector. Finally, the temporal amplitude, unwrapped phase, and the prior vector are concatenated as the input to the LSTM layers, while the model's output is the predicted pulse amplitude and phase before compression.

Table 2. System setup under pump powers and chirp parameters generalization.

Component	Parameters	Value
CFBG	Chirp parameter β_2	18~22 ps^2
	Chirp parameter β_3	0~0.2 ps^3
Pump1	Power	1~2 W
Pump2	Power	1~2 W
Pump3	Power	25~30 W

2.4. Training of the cascaded LSTM

Each LSTM is independently trained and employed to simulate temporal dynamics within different ytterbium-doped fiber (YDF), with a fixed sequence length of 20. A dense layer then consolidates the outputs from the LSTM layers and predicts the next step. During training, the input sequences are pulses obtained via the SSFM-RK4 method. During inference, the input sequence operates in a sliding-window manner, and the last step of the 20-step input is replaced by the previous prediction. Note that for predicting the second pulse, the first 20 steps are all identical to the initial pulse. Moreover, for the SSFM-RK4 method, a step size of 1 mm is adopted, with 6000 steps per evolution over three 2-meter YDF. To reduce the inference time, the simulation steps are downsampled by a factor of 20, reducing the inference steps to 300.

Optics EXPRESS

In deep learning, GPO denotes the requirement for distinct optimization strategies to be applied to parameters of different network layers. This includes adopting different learning rates for various parameter groups or employing different optimization algorithms across different layers according to their different functions [23–26]. Considering the huge functional differences between encoding layers and LSTM layers, we introduce a GPO method that incorporates layer-wise statistical gradient norms. We design a two-stage learning rate scheduling strategy: a warm-up phase where the learning rate increases linearly, and a cosine annealing phase where the learning rate decays following a cosine function. The warm-up phase prevents unstable convergence caused by excessively large initial learning rates, enabling the model to gradually adapt to parameter updates and avoid getting trapped in poor local optima, laying a foundation for stable convergence in subsequent phases. The mathematical expression of the proposed strategy is as follows,

$$LR(k) = \begin{cases} LR_0 \cdot \frac{k}{N} & , \quad 0 \le k < N \\ LR_0 \cdot s \cdot \frac{1+cos\left(\pi \cdot \frac{k-N}{T}\right)}{2} & , \quad k \ge N \end{cases}. \tag{5}$$

where $LR(k)$ denotes the learning rate at the current training epoch k, LR_0 is the target learning rate at the end of the warm-up phase, N stands for the total number of epochs for the warm-up phase, T represents the period of the cosine annealing phase. s is the scaling coefficient for different layers calculated by,

$$s = \log_{10}(1 + 1/(g_l||_2)). \tag{6}$$

where l is the index of the network layer, and g_{l2} refers to the L_2 norm of the gradient for the parameters in different sub-networks, which is calculated at the end of the warm-up stage.

2.5. Evaluation metric

To evaluate the performance of the model, a weighted MSE loss function incorporating both amplitude and unwrapped phase indicated in Eq. (7) is used during training. During inference, the model's performance was assessed by comparing the pulse width and energy along the gain process with those of the pulse generated by the SSFM-RK4 method.

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{N}\sum_{i=1}^{N}[\alpha \cdot (A^{(i)}_{\mathrm{pred}} - A^{(i)}_{\mathrm{SSFM-RK4}})^2 + \beta \cdot (\phi^{(i)}_{\mathrm{pred}} - \phi^{(i)}_{\mathrm{SSFM-RK4}})^2]. \tag{7}$$

$$\mathrm{SR} = \frac{\max(|\mathrm{A}_{\mathrm{compressed}}(\mathrm{t})|^2)}{\max(|\mathrm{A}_{\mathrm{TL}}(\mathrm{t})|^2)}. \tag{8}$$

Furthermore, to evaluate the accuracy of the final recovered pulse, three metrics were adopted: the compressed output peak power, the full width at half maximum (FWHM), and the Strehl ratio. The Strehl ratio, as defined in Eq. (8) is defined as the peak-power ratio between the compressed pulses and the TL pulses [27], which indicates how closely the compressed pulse approaches its corresponding Fourier transform-limited pulse.

3. Results

To present the prediction accuracy of the cascaded LSTM model, Fig. 3 represents the overall pulse evolution under the downsampling ratio of 512 (denoted by the symbol r), and the detailed pulse profile at concatenation points between discrete gain fibers. Figure 3(a) shows the temporal dynamics of the pulse propagation based on the proposed model, while Fig. 3(b) shows the intensity and phase of the signal located at fiber-to-fiber coupling points and the end of all amplifiers. Compared to the simulation results generated by the SSFM-RK4 model, the proposed model predicts both the temporal amplitude and temporal phase accurately. The spectral full-field

Optics EXPRESS

signal can be obtained by performing the fast Fourier transform (FFT) on the predicted temporal signal after cubic interpolation. The spectral amplitude and phase also maintain incredible consistency in Fig. 3(c). Figure 3(d) displays the spectral dynamics of the predicted pulse propagation obtained by performing FFT on the interpolated temporal pulse evolution dynamics.

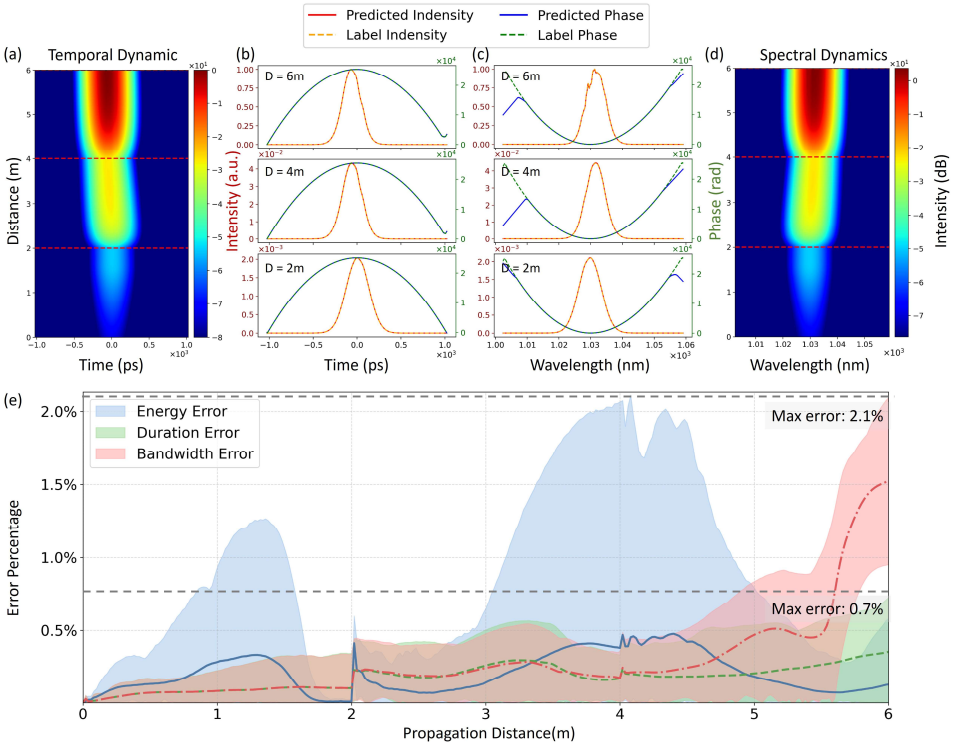

Fig. 3. Simulation results based on the cascaded LSTM under the downsampling ratio of 512 (a) the temporal dynamics and (d) the spectral dynamics of pulse propagation in 3-stages CPA system; the comparison of the intensity and phase in (b) temporal domain and (c) spectral domain locating at coupling points and output; (e) pulse energy, duration and bandwidth error within the whole system, including mean value of errors (denoted by lines) and error distributions (denoted by color blocks).

To demonstrate the generalization ability of the cascaded LSTM model, Fig. 3(e) shows pulse energy, duration, and bandwidth error distribution within gain fibers evaluated across various parameter configurations. The proposed model reveals the great ability of modeling the gain process with the error of pulse energy and bandwidth remaining below 2.1%, while the error of duration remains below 0.7%.

To further prove the validity of the cascaded LSTM model, both the interpolated predicted pulses and the label pulses are compressed with identical grating distances. The amplified pulse with a duration of 300 ps can be compressed to a pulse with a duration of less than 300 fs, and the time resolution can be reconstructed from 32 ps to ~30 fs by interpolation. To evaluate the performance of the compressed pulse, we proposed three metrics, including compressed duration, peak power, and Strehl ratio. Figure 4(a) demonstrates that the reconstruction error of these metrics remains below 2% when r is less than 1024, which means the cubic interpolation over pulses predicted by LSTM can be considered lossless. However, the cubic interpolation cannot completely restore the information of the pulse when r is 1024. The average reconstruction error of the predicted compressed duration is 4.90%, and the average errors of the peak power and

Optics EXPRESS

Strehl ratio are both greater than 20%. Figure 4(b) and Fig. 4(c) demonstrate compressed pulse intensity under different parameter configurations to prove the generalization abilities of cascaded LSTM models. As a result, when r is less than 1024, the predicted compressed pulses can be well fitted to the label.

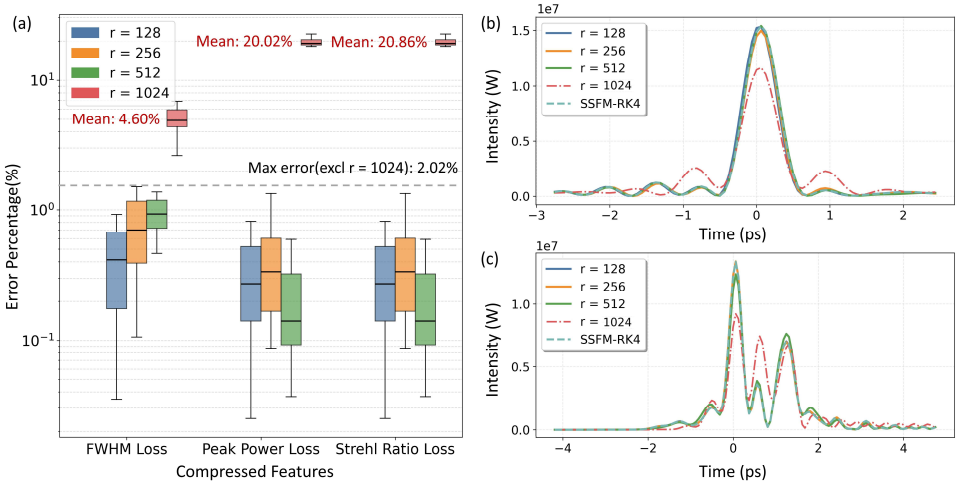

Fig. 4. (a) Reconstruction error of the compressed pulses predicted by different r; compressed pulses predicted by different r with (b) CFBG dispersion parameters of [22 ps^2, 0.0 ps^3], pump powers of [1 W, 1 W, 25 W], (c) CFBG dispersion parameters of [18 ps^2, 0.2 ps^3], pump powers of [2 W, 2 W, 30 W].

4. Discussion

4.1. Inference time and computational complexity

To investigate the performance dependence of the proposed model on r, 4 separate models with different rates from 128~1024 are trained and tested. Figure 5(a) and Fig. 5(b) show the comparison among these models in terms of the simulation time on the identical GPU-based platforms (GeForce RTX 3090) and CPU-based platforms (Intel Xeon Gold 6146 CPU @ 3.20 GHz). With the acceleration of CUDA, the inference time for different cascaded LSTM models does not exceed 0.36s, and the time required increases slowly as the batch size increases. On the other hand, the SSFM-RK4 model takes more than 1,500 times longer than LSTM, and the gap between the two methods widens as the batch size increases. When testing on an identical AMD-CPU-based platform, the cascaded LSTM models are still over 800 times faster than the SSFM-RK4 model.

Meanwhile, the number of floating-point operations (FLOPs) and the number of model parameters are utilized to characterize the computational complexity. Figure 5(c) shows the computational complexity of the cascaded LSTM model and the SSFM-RK4 model with a batch size of 50. When downsampled by 512 times, the number of FLOPs of the cascaded LSTM model is ~929 times less than the conventional SSFM-RK4 model. As r increases, the complexity decreases approximately linearly. Specifically, when r reaches 512, the FLOPs are reduced to only 2.81 G.

4.2. Validity of the cascaded LSTM

The parameter differences of various gain fibers are the primary factor leading to the distinct dynamic evolution of pulses. The core parameters include doping concentration and fiber radius,

Optics EXPRESS

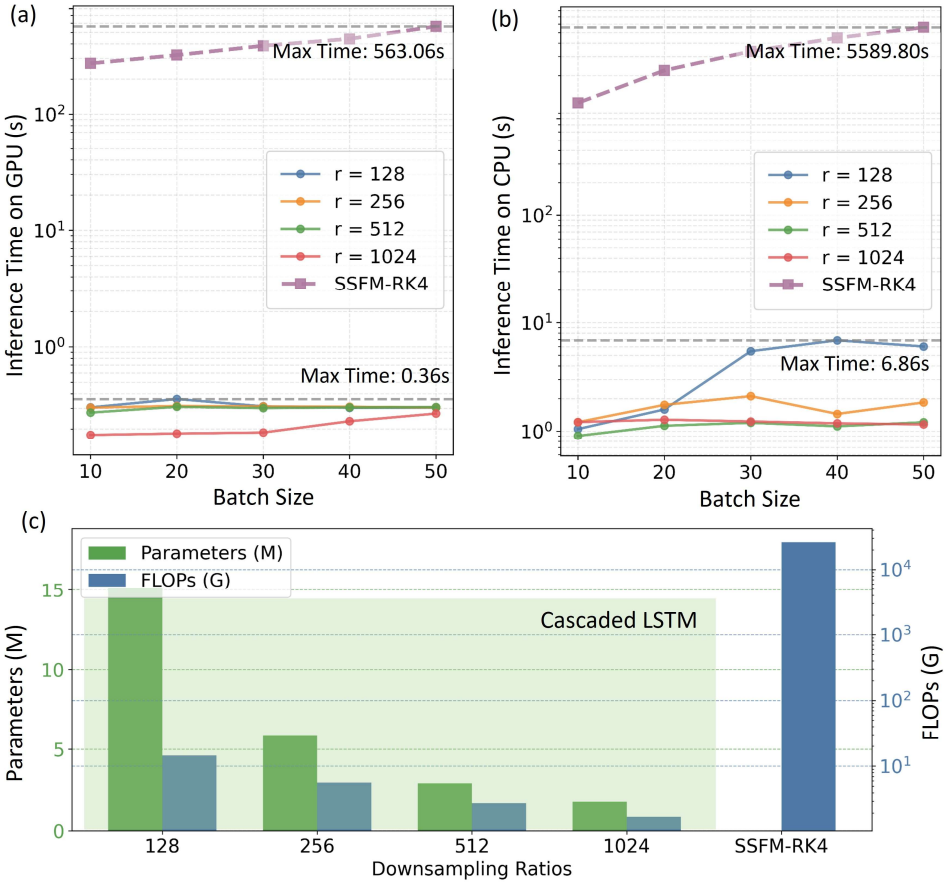

Fig. 5. Inference time of the cascaded LSTM with different downsampling ratios r and the SSFM-RK4 model on the identical (a) GPU-based platform and (b) CPU-based platform; (c) Computational complexity comparison between the cascaded LSTM with different r and the SSFM-RK4 model. FLOPs, Floating point operations.

governing the gain profile and the magnitude of nonlinear effects, respectively. To accurately simulate the pulse dynamics, two distinct training approaches are tested separately: the single LSTM refers to using a single LSTM to characterize the physical effects in different optical fibers; correspondingly, training independent LSTM models for different types of fibers is termed the cascade LSTM.

As shown in Fig. 6, the generalization capability of the single LSTM is insufficient to predict pulse dynamics among different optical fibers. When inferring, the single LSTM exhibits extremely high prediction errors. At the end of the second amplification fiber, the maximum error in pulse energy prediction reaches 298.1%. In contrast, despite the identical model structure for different fibers in cascade LSTM, merely training them separately significantly improves prediction accuracy. During the inference phase, the pulse energy prediction error of the cascade LSTM is always below 2%.

4.3. *Validity of the GPO strategy*

As proposed in Section 2.4, the GPO strategy calculates the statistical gradient norms of each sub-network during the warm-up stage to determine the appropriate learning rate for each layer.

Optics EXPRESS

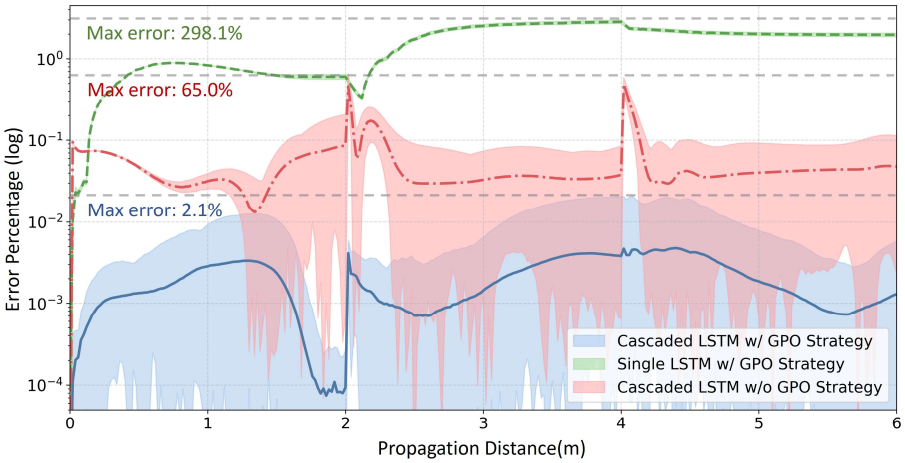

Fig. 6. Pulse energy error of the cascaded LSTM trained with GPO strategy(blue), the single LSTM trained with GPO strategy(green), and the cascaded LSTM trained without GPO strategy (red), including the mean value of errors (denoted by lines) and error distributions (denoted by color blocks).

Specifically, the L_2 norm of gradients in the LSTM network is found to be an order of magnitude larger than that in the encoding layer. Therefore, the initial learning rate of the LSTM is set larger than that of the encoding layers after warming up, to alleviate the gradient vanishing issue. Figure 6 also compares the prediction error of the pulse energy with different training strategies. The model trained with a unified learning rate witnesses the highest pulse energy error at fiber-to-fiber coupling points up to 65.03%, which is the most difficult position to predict. However, the model trained with the GPO strategy reveals better ability in the inference phase, with the prediction error remaining below 2.11%.

5. Conclusion

In summary, we present a cascaded LSTM framework with a downsampling strategy for efficient and accurate full-field simulation of a multi-amplifier CPA system. The proposed model effectively handles large-bandwidth, heavily chirped, high-energy pulse propagation and significantly reduces computational cost. Compared to the conventional SSFM-RK4 method, the proposed method delivers a 1,564-fold speedup and a 929-fold reduction in computational complexity under the downsampling factor of 512, while maintaining reconstruction errors of pulse energy and duration below 2% simultaneously. We expect this substantial improvement in computational efficiency to greatly benefit large-scale parametric studies and optimization tasks in high-power CPA system design. Furthermore, we believe the proposed architecture and training strategy can be extended to other ultrafast laser systems and serve as a general-purpose modeling tool in computational photonics.

Funding. National Natural Science Foundation of China (62205199, 62227821, 62025503).

Acknowledgment. National Natural Science Foundation of China awards: 62205199 to G. Pu, 62227821 to L. Yi, and 62025503 to L. Yi.

Disclosures. The authors declare that there are no conflicts of interest related to this article.

Data availability. 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.

Optics EXPRESS

References

1. C. L. Li, C. J. Fisher, R. Burke, *et al.*, "Orthopedics-related applications of ultrafast laser and its recent advances," Appl. Sci. **12**(8), 3957 (2022).
2. R. R. Gattass and E. Mazur, "Femtosecond laser micromachining in transparent materials," Nature Photon **2**(4), 219–225 (2008).
3. L. Shah, M. E. Fermann, J. W. Dawson, *et al.*, "Micromachining with a 50 W, 50 μJ, subpicosecond fiber laser system," Opt. Express **14**(25), 12546 (2006).
4. H. Lubatschowski and A. Heisterkamp, "Interaction with biological tissue," *in Femtosecond Technology for Technical and Medical Applications*, F. Dausinger, H. Lubatschowski, and F. Lichtner, eds., Topics in Applied Physics (Springer Berlin Heidelberg, 2004), 96, pp. 91–105.
5. Y. Xing, R. Chen, L. Zhang, *et al.*, "SLAM medical imaging enabled by a pre-chirp and gain jointly managed Yb-fiber laser," Biomed. Opt. Express **15**(2), 911 (2024).
6. V. P. Mitrokhin, A. B. Fedotov, A. A. Ivanov, *et al.*, "Coherent anti-Stokes Raman scattering microspectroscopy of silicon components with a photonic-crystal fiber frequency shifter," Opt. Lett **32**(23), 3471 (2007).
7. Y. Azamoum, G. A. Becker, S. Keppler, *et al.*, "Optical probing of ultrafast laser-induced solid-to-overdense-plasma transitions," Light Sci Appl **13**(1), 109 (2024).
8. H. Song, B. Liu, Y. Li, *et al.*, "Practical 24-fs, 1-μJ, 1-MHz Yb-fiber laser amplification system," Opt. Express **25**(7), 7559 (2017).
9. X. Liu, D. Popa, and N. Akhmediev, "Revealing the transition dynamics from Q switching to mode locking in a soliton laser," Phys. Rev. Lett **123**(9), 093901 (2019).
10. J. S. Feehan, S. R. Yoffe, E. Brunetti, *et al.*, "Computer-automated design of mode-locked fiber lasers," Opt. Express **30**(3), 3455 (2022).
11. L. Salmela, N. Tsipinakis, A. Foi, *et al.*, "Predicting ultrafast nonlinear dynamics in fiber optics with a recurrent neural network," Nat Mach Intell **3**(4), 344–354 (2021).
12. G. Pu, R. Liu, H. Yang, *et al.*, "Fast predicting the complex nonlinear dynamics of mode-locked fiber laser by a recurrent neural network with prior information feeding," Laser Photonics Rev. **17**(6), 2200363 (2023).
13. W. Ma, Z. Liu, Z. A. Kudyshev, *et al.*, "Deep learning for the design of photonic structures," Nat. Photonics **15**(2), 77–90 (2021).
14. H. Jiang, G. Pu, X. Ma, *et al.*, "Active-learning-augmented end-to-end modeling towards fast inverse design in chirped pulse amplification," (n.d.).
15. T. Chu, G. Pu, H. Yang, *et al.*, "AI-enabled fast and accurate modeling for femtosecond chirped-pulse amplification," in *2022 Asia Communications and Photonics Conference (ACP)* (IEEE, 2022), pp. 1899–1902.
16. R. Lindberg, P. Zeil, M. Malmström, *et al.*, "Accurate modeling of high-repetition rate ultrashort pulse amplification in optical fibers," Sci Rep **6**(1), 34742 (2016).
17. G. P. Agrawal, *Nonlinear Fiber Optics*, Fifth edition (Elsevier/Academic Press, 2013).
18. R. Paschotta, J. Nilsson, A. C. Tropper, *et al.*, "Ytterbium-doped fiber amplifiers," IEEE J. Quantum Electron **33**(7), 1049–1056 (1997).
19. Z. Huang, J. Wang, H. Lin, *et al.*, "Combined numerical model of laser rate equation and Ginzburg–Landau equation for ytterbium-doped fiber amplifier," J. Opt. Soc. Am. B **29**(6), 1418 (2012).
20. P. K. Mukhopadhyay, K. Ozgoren, I. L. Budunoglu, *et al.*, "All-fiber low-noise high-power femtosecond Yb-fiber amplifier system seeded by an all-normal dispersion fiber oscillator," IEEE J. Select. Topics Quantum Electron **15**(1), 145–152 (2009).
21. J. Limpert, T. Schreiber, T. Clausnitzer, *et al.*, "High-power femtosecond Yb-doped fiber amplifier," Opt. Express **10**(14), 628 (2002).
22. K. Gürel, P. Elahi, L. Budunoğlu, *et al.*, "Prediction of pulse-to-pulse intensity fluctuation characteristics of high power ultrafast fiber amplifiers," Appl. Phys. Lett. **105**(1), 011111 (2014).
23. B. Singh, S. De, Y. Zhang, *et al.*, "Layer-specific adaptive learning rates for deep networks," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* (IEEE, 2015), pp. 364–368.
24. B. Ginsburg, P. Castonguay, O. Hrinchuk, *et al.*, "Stochastic gradient methods with layer-wise adaptive moments for training of deep networks," (2020).
25. P. Sridar, A. Kumar, A. Quinton, *et al.*, "Decision fusion-based fetal ultrasound image plane classification using convolutional neural networks," Ultrasound in Medicine & Biology **45**(5), 1259–1273 (2019).
26. Z. Zhang, Q. Zhang, Z. Gao, *et al.*, "Gradient-based parameter selection for efficient fine-tuning," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2024), pp. 28566–28577.
27. S. Wang, B. Liu, M. Hu, *et al.*, "On the efficiency of parabolic self-similar pulse evolution in fiber amplifiers with gain shaping," J. Lightwave Technol **34**(13), 3023–3034 (2016).