Machine Learning-Based Direct Solver for One-To-Many Problems on Temporal Shaping of Electron Beams

Jinyu Wan
Institute of High Energy Physics
Yi Jiao (jiaoyi@ihep.ac.cn)
Institute of High Energy Physics
Juhao Wu
SLAC National Accelerator Laboratory

Research Article

Keywords: Machine learning, CGAN, electron beams, optima

Posted Date: May 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-524222/v1

Machine learning-based direct solver for one-to-many problems on temporal shaping of electron beams

Jinyu Wan,[1,2] Yi Jiao,[1,2,*] and Juhao Wu[3,†]

[1] *Key Laboratory of Particle Acceleration Physics and Technology, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China*
[2] *University of Chinese Academy of Sciences, Beijing 100049, China*
[3] *SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, United States*

(Dated: May 14, 2021)

To control the temporal profile of an electron beam to meet requirements of various advanced scientific applications, a widely-used technique is to manipulate the dispersion terms which turns out to be one-to-many problems. Due to their intrinsic one-to-many property, current popular stochastic optimization approaches on temporal shaping are not very effective, for being trapped into local optima or suggesting only one solution. Here we propose a real-time solver for one-to-many problems of temporal shaping, with the aid of a semi-supervised machine learning method, the conditional generative adversarial network (CGAN). We demonstrate that the CGAN solver can learn the one-to-many dynamics and is able to accurately and quickly predict the required dispersion terms for different custom temporal profiles. This machine learning-based solver overcomes the limitation of the stochastic optimization methods and is expected to have the potential for wide applications to one-to-many problems in other scientific fields.

INTRODUCTION

Much of the interest in advanced scientific applications of particle accelerator user facilities such as free-electron lasers (FELs) [1, 2], terahertz radiation [3, 4] and plasma wakefield accelerators (PWFAs) [5] has grown in the past decades. FELs and terahertz radiation are emerging as powerful imaging tools in various fields [6–9] like physics, chemistry, biology and material science. PWFAs are potential to provide an accelerating gradient up to multi-GV/m level for high energy physics and photon science [10]. Such a gradient is much higher than that obtained with conventional radio-frequency-based accelerators. To realize these advanced accelerator applications, a critical issue is to provide electron beams of particular temporal shapes to improve quality of the electron or photon beam, namely temporal shaping of electron beams [11–16]. For example, a linearly ramped beam is required by PWFAs to supply a high transformer ratio of wakefield [11, 16], and beams with flat-top temporal profiles are desired in FELs to obtain high radiation performance [12, 13].

One widely-used temporal shaping method is to let an electron bunch with inhomogeneous energy distribution pass through an energy dispersion section, like a so-called bunch compressor consisting of several bending magnets to realize various temporal profiles [17]. For beams with small transverse emittances and large energy spread, the contribution of geometrical terms in the transfer map, which relate to the transverse emittances, can be neglected and the dispersion terms that connect to the energy spread will tend to dominate the process [18]. Hence, for an initial beam having an inhomogeneous energy chirp, a desired custom temporal profile can be achieved by finding an appropriate combination of dispersion terms (see details in Methods section). Due to the highly nonlinear process, different combinations of dispersion terms may realize the same target profile, i.e. the temporal shaping can intrinsically turn into a one-to-many problem. The exploration of multiple solutions instead of only one is beneficial for various scientific applications, e.g. for beams with the same temporal profile, but with opposite energy chirp along the electron beam can provide different benefits in FELs [19, 20]. In addition, it should be noted that not all potential solutions are equally feasible to implement in practical scientific experiments, e.g. one may require unfeasibly stronger magnets than those in other solutions. Therefore, it is important to derive multiple, if not all the, potential solutions that can result in almost the same target profile.

Grid scan [21] was first used to solve such a temporal shaping problem to realize a ramped profile with a second-order approximation. However, due to the highly nonlinear process, the second-order approximation can be insufficient. Yet, when contributions of the higher-order longitudinal dispersion terms are included in the scan, the time cost of grid scan can exponentially grow.

Later, it has been shown [22–25] that such one-to-many problems can be solved more efficiently with stochastic optimization methods like genetic algorithm (GA) [26, 27], particle swarm optimization (PSO) [28] and extremum seeking (ES) [29]. Nevertheless, the stochastic optimization process is still indirect and some challenging problems still remain. One challenge of using these stochastic optimization methods is to avoid being trapped into local optima [30–32]. In addition, for one-to-many problems that have multiple potential solutions, the optimization process usually stops when the first solution is found, resulting in the omission of other potential solutions that may be more feasible. The finally obtained results can be highly dependent on the choice of the initial solution. So a warm start is critical for most

* corresponding author: jiaoyi@ihep.ac.cn
† corresponding author: jhwu@slac.stanford.edu

of these optimization algorithms. Although one can implement additional constrains to see whether an obtained solution is feasible or not, the constrained optimization can be more difficult to solve and more computationally expensive [33].

In recent years, machine learning (ML) has attracted increasing interests of accelerator experts as a powerful tool to reveal the complicated correlations between various accelerator parameters [34–42]. With the aid of ML, the majority of the efforts, e.g. Ref. [35, 37, 38, 41, 42], has been made to improve the performance of stochastic optimization methods. It is noticed that, however, most of these ML applications are based on supervised ML models, which are only powerful to capture the map of one-to-one problems where one feature vector X has only one definite label vector Y. When a supervised ML model is trained with data samples of a one-to-many problem that has different labels for the same input feature vector, it tends to output the mean label value of the samples rather than the respective label value of each sample itself. For example, a supervised ML model is able to predict the temporal profile of an electron bunch with known accelerator settings [36]. While for the inverse problem, namely, predicting the accelerator settings for a desired custom temporal profile, the supervised ML methods like multilayer perceptron may fail to give the right answer because multiple solutions possibly exist.

To overcome above limitations, here we introduce another ML method, the generative adversarial networks (GANs) [43], that has potential to handle one-to-many problems [44, 45]. The GANs are emerging techniques of unsupervised and semisupervised ML that have become one of the state-of-the-art techniques to solve difficulties in image synthesis [46], style transfer [47], and image superresolution [48]. Instead of minimizing the mean error as in most supervised ML methods, GANs take a different approach to learn correlations hidden in the training data, via the competition of a pair of neural networks, called the generator (hereafter referred to as G) and discriminator (hereafter referred to as D). G is trained to create fake data samples as authentic as possible to fool D, and D is trained to distinguish between the fake and real samples. Considering that the generative process in original GANs is only determined by the noise fed to G and is difficult to control, the GANs are extended by Mirza et al. [49] to a conditional framework by making both G and D class-conditional. With an additional label added to the input data, both G and D networks receive additional information about the correlation of the samples and the given label so that the networks can synthesize samples with user specified content. This extended approach, called conditional GAN (CGAN), that allows one to direct the data generation process, has proven to be effective to create images with a target class [50, 51].

In this paper, we propose a CGAN solver for one-to-many problems of temporal shaping. Two common temporal profiles, the cusp-shaped profile and the double-horn profile, as well as two additional profiles that have greater scientific merit, the flat-top profile and linearly-ramped profile, are used to test the performance of the CGAN solver. We demonstrate that for the complex processes where multiple local optima exist, three widely-used stochastic optimization methods, the PSO, GA and ES are easy to fall into local optima, while the CGAN solver remains effective to accurately and quickly predict the required longitudinal dispersion terms to realize the custom temporal profiles. In addition, the CGAN solver can find many, if not all the, potential solutions, breaking the limitation of stochastic optimization methods that normally lead to one solution.

RESULTS

Temporal shaping scheme based on CGAN solver. In our temporal shaping scheme (see Fig. 1 and Methods section), the first-, second- and third-order longitudinal dispersion terms, i.e. R_{56}, T_{566} and U_{5666}, labeled with the corresponding temporal profile are taken as training data. Here we consider the third-order approximation because the third-order longitudinal dispersion term was proven to be important in previous studies and it is rarely necessary to consider even higher-order terms for single-pass transport [52]. A generator G is trained as a solver that can produce fake samples of longitudinal dispersion terms to realize new temporal profiles. Meanwhile, a discriminator D is trained to compete with G to force the fake samples generated by G to satisfy the distribution learned from the training data. Details of the data preparation and the training settings are described in Methods section. By feeding the trained solver with the target temporal profile and the noise component, the solver is expected to predict different potential solutions to realize the target when multiple solutions exist.

In this study, we use an initial electron beam (see Fig. 1) of Gaussian charge-density distribution having an inhomogeneous energy chirp, with the same beam parameters as in Ref. [13]. The initial beam is sent to a chicane-type magnetic compressor that have specific R_{56}, T_{566} and U_{5666} values, and the custom desired temporal profiles are expected to be obtained at the exit of the chicanes. The layout of the chicane is shown in Fig. 1. The layout of chicane is chosen to such that the R_{56}, T_{566} and U_{5666} of this chicane can be adjusted in a large range by tuning strengths of the magnets with a direct search method [53].

How the CGAN solver overcomes the limitations of stochastic optimization. Two temporal profiles that are common in bunch compression, i.e. a cusp-shaped profile and a double-horn profile (see Fig. 2(c, d)) are used as target profiles to test the performance of our CGAN solver. To look into the details of how the CGAN solver solves the temporal shaping problems, grid scan is first performed in the R_{56}, T_{566} and U_{5666} space to find all potential solutions for the two test target profiles, respectively. According to the $100\times100\times100$ grid scan within a empirically large range (see Fig. 2 (a,

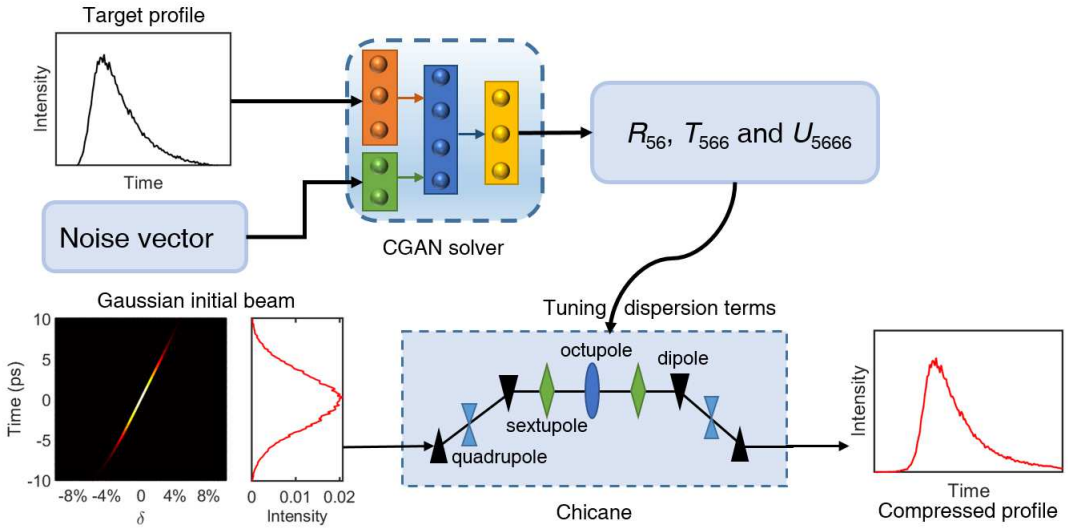

FIG. 1. Schematic diagram of the CGAN solver in this study. The color of the initial beam from black to white represents the charge-density from low to high. The compressed temporal profile is obtained by letting the initial beam pass through a chicane having specific R_{56}, T_{566} and U_{5666}. By feeding the custom temporal profiles and noise components to the trained generator of the CGAN, the CGAN solver is able to predict the required dispersion terms to realize the input target temporal profiles.

b)), at least two potential solutions that have almost the same highest objective performance are found for each target profiles, which suggests that both temporal shaping problems are one-to-many. Besides the potential solutions, a plenty of local optima are also observed in the variable space of the double-horn profile, indicating that the bunch compression process of the double-horn profile has a stronger nonlinearity.

Before implementing the CGAN solver, we also test current state-of-the-art approaches for temporal shaping, the stochastic optimization methods, to realize the two test target profiles. Three stochastic optimization methods widely-used in the accelerator filed, namely the PSO, GA and ES, are used to solve such two temporal shaping problems. The parameter settings for the three methods are described in Methods section. The evolutionary trajectories in the variable space of the three stochastic optimization methods are shown in Fig. 2 (a, b), and the final temporal profiles obtained with these stochastic optimization methods are plotted in Fig. 2 (c, d).

The results in Fig. 2(a) appears that for the cusp-shaped profile, only the PSO can find one of the two potential solutions to realize the target profile with high fitness, and other two methods fail to find any potential solutions. While for the double-horn profile that is more complicated, the evolution of PSO is also trapped in a cluster of local optima instead of the global. The results indicate that for a highly nonlinear process where many local optima exist, the stochastic optimization methods can easily lead to a local optimum instead of the global, resulting in obvious mismatch compared to the target profiles. In addition, for a one-to-many problem, the stochastic optimization methods tend to find one potential solution while with others missed, and there is no guarantee that the obtained solution is the most feasible one.

Then the CGAN solver is implemented to solve the two test problems. The target profiles and the noise components are simultaneously fed to the trained solver which finally results in multiple sets of fake R_{56}, T_{566} and U_{5666} samples. It is found that the fake samples converge to several points in the phase space, which represent multiple solutions of a temporal shaping problem. For each of the cusp-shaped profile and double-horn profile, two separate solutions obtained with the CGAN solver are shown in Fig. 2(a, b) in the variable space for comparison.

For the cusp-shaped profile, the CGAN solver not only results in almost the same solution as the best solution obtained with PSO, but also finds another potential solution with slightly lower objective performance, which is missed with the tested stochastic optimization methods. For the double-horn profile, the CGAN predictions are close to the highest peaks in the grid scan map, showing significantly higher objective performance than those obtained by the three stochastic optimization methods. The longitudinal phase space distribution resulted from the CGAN predictions are shown in Fig. 3. It is found that the beams in Fig. 3(b, d) are over compressed, i.e. the head and tail of the beam are reversed. Nevertheless, the over compressed beam finally results in almost the same temporal profiles as the under compressed beam in Fig. 3(a, c), with a high determination coefficient close to 1.

The results in Fig. 2 and Fig. 3 indicate that the CGAN solver can predict the longitudinal dispersion terms to realize the custom desire temporal profiles with high accuracy, without the risk of being trapped in local optima as the stochastic optimization methods do. Fur-

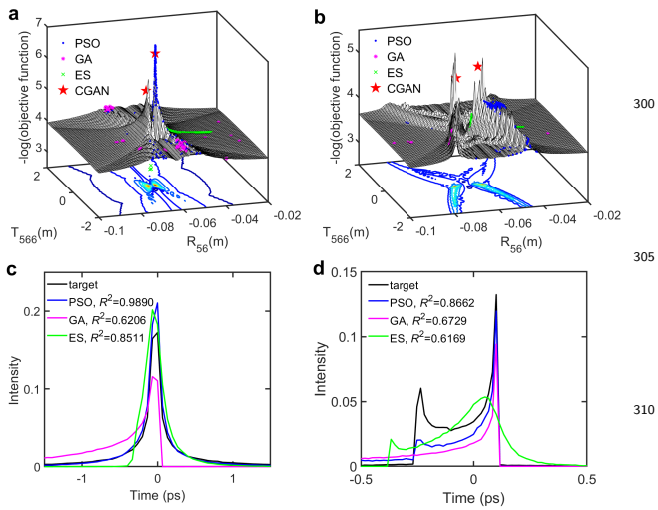

FIG. 2. Using stochastic optimization methods to solve temporal shaping problems. (a) and (b) show the grid scan results, and the evolutionary trajectory of three stochastic optimization methods in the variable space for a cusp-shaped profile and a double-horn profile, respectively. The z-axis in (a) and (b) is –log(*objective function*) representing the objective performance. The contour maps of the grid scan results are plotted at the bottom, where the color from blue to red represent the objective performance from low to high. For each R_{56}-T_{566} grid in (a) and (b), only one U_{5666} with the highest objective performance is plotted. Two separate predictions of the CGAN solver are also shown in (a) and (b) for comparison. (c) and (d) are the final temporal profiles obtained with three stochastic optimization methods for the cusp-shaped profile and double-horn profile, respectively. R^2 is the determination coefficient to the target temporal profile.

thermore, the CGAN solver is able to give different solutions for the same input temporal profile when multiple solutions exist, which breaks the limitation of stochastic optimization methods that normally lead to one solution. The acquirement of the multiple solutions in Fig. 3 is crucial for temporal shaping since an under compressed beam and an over compressed beam can provide different benefits in scientific applications. For example, an over compressed scheme has the potential to provide larger bandwidth FEL radiation [19]. Compared to the under compressed beam, however, the over compressed beam also leads to a significant coherent synchrotron radiation (CSR) that can reduce the slice alignment and spoil the transverse emittance of the electron beam [20]. Besides, it is possible to select one from the obtained multiple solutions that is more feasible to implement in practical scientific experiments by experienced operators or an additional evaluator. For instance, it is found that to realize the cusp-shaped profile, the octupole strength required to achieve the longitudinal dispersion terms of the PSO solution is significantly higher than that of another solution obtained with the CGAN solver ($-15.7\mathrm{m}^{-3}$ and $0.4\mathrm{m}^{-3}$ respectively). The strong octupoles are not trivial to build and will bring high sensitivity during the beamline optimization and operation. Fig. 4 shows the average solving time for different methods. Compared to the stochastic optimization methods, the CGAN solver can be several orders of magnitude faster because once the CGAN is trained, it only needs little time (fractions of one second) to directly predict the longitudinal terms of a new temporal profile.

Realization of two additional temporal profiles with scientific merit. In addition to the above two common temporal profiles, two other temporal profiles, namely the flat-top profile and the linearly-ramped profile (see Fig. 5), are also studied. These two additional temporal profiles are frequently-used in various scientific applications and have greater scientific merit. The flat-top temporal profile is desired in FELs to reduce the third-order curvature in the time-energy correlation due to wakefield, so as to obtain better FEL performance with improved pulse energy, peak power and bandwidth control [12]. The linearly-ramped temporal profile is treated as the optimal shape of the drive beam in PWFAs to supply a high transformer ratio because it maximizes the energy that can be gained by a trailing particle accelerated in its wakefield [11].

The two additional test temporal profiles are also fed to the trained CGAN solver to predict multiple fake samples of longitudinal dispersion terms. The final temporal profiles resulted from two randomly selected fake samples for each custom profile are illustrated in Fig. 5. For the flat-top profile, horns occur at the head and the tail of the obtained bunch, which cannot be completely eliminated due to the nature of bunch compression with a single chicane compressor. The horns may be further flattened with an additional bunch compressor [54] that is, however, beyond the scope of this study. Nevertheless, the FWHM of the horns is very narrow compared with the flat part of the bunch. For the cusp-shaped and ramped profiles, the temporal profiles resulted from the predictions of the CGAN solver fit well to the target profiles with a determination coefficient close to 1.

DISCUSSION

We have proposed a CGAN solver for one-to-many problems of temporal shaping of electron beams. By learning from the stochastically generated data, a trained CGAN solver can quickly and accurately predict available combinations of the dispersion terms up to the 3rd order to realize desired custom temporal profiles. The CGAN solver remains effective in highly nonlinear process where three widely-used stochastic optimization methods, the PSO, GA and ES are trapped in local optima. In addition, for one-to-many problems, the CGAN can predict many, if not all the, potential solutions simultaneously, which overcomes the limitation of stochastic optimization methods that find one solution with others missed.

This method can be easily transferable to other similar

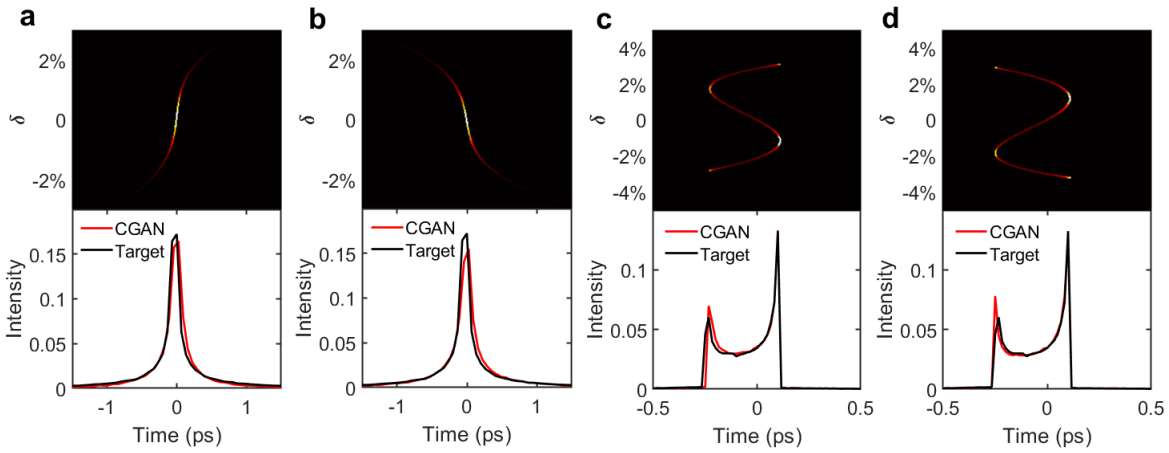

FIG. 3. Longitudinal phase space distribution and temporal profiles of two separate CGAN predictions. The left two columns represent the cusp-shaped profile, and the right two columns represent the double-horn profile, respectively. R^2 is the determination coefficient to the target temporal profile.

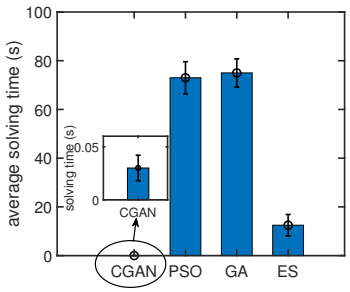

FIG. 4. The average solving time of different methods for five repeated tests.

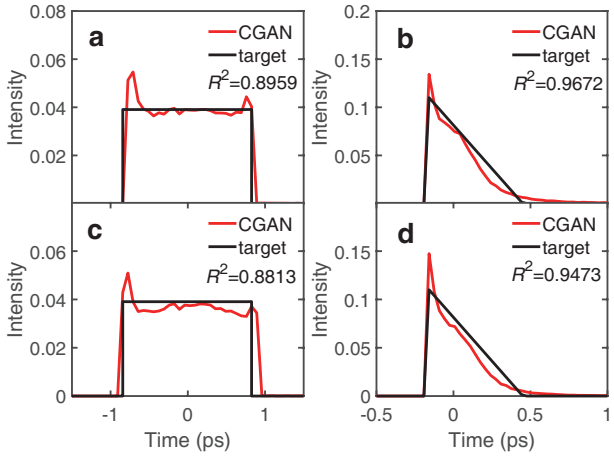

FIG. 5. The temporal profiles obtained with the CGAN solver for the flat-top profile (left) and the linearly-ramped profile (right). The two rows represent two different CGAN predictions for each temporal profile. R^2 is the determination coefficient to the target profile.

problems, for instance, photon pulse shaping and transverse phase space manipulation of an electron bunch. We expect that the CGAN solver can serve as a direct and real-time method to solve one-to-many problems in more scientific applications.

METHODS

Bunch compression mechanism. In a system of magnetic elements, a transfer map M describes the relation of initial conditions ζ^i and final conditions ζ^f, which can be symbolically written in the form $\zeta^f = M\zeta^i$. A Taylor map [18] that represents the final conditions as a Taylor series of the initial conditions can be described as

$$\zeta_j^f = \sum_k R_{jk}\zeta_k^i + \sum_{kl} T_{jkl}\zeta_k^i\zeta_l^i + \sum_{klm} U_{jklm}\zeta_k^i\zeta_l^i\zeta_m^i + \ldots \quad (1)$$

where R, T and U are the first-, second- and third-order transfer matrices, and j, k, l and m are the element indices of the coordinate.

It is empirically found that the CGAN is more effective in dealing with low dimensional problems. Therefor the transfer map here is needed to be sparse. Fortunately, for beams with small transverse emittance and large energy spread, the contribution of geometrical terms can be neglected and the dispersion terms will tend to dominate. A Taylor map that presents the final longitudinal coordinate $q_{z,f}$ of a charged particle as a Taylor series of

the initial coordinates can be simplified as

$$q_{z,f} = q_{z,0} + R_{56}\delta(q_{z,0}) + T_{566}\delta(q_{z,0})^2 + U_{5666}\delta(q_{z,0})^3 + ... \quad (2)$$

$$\delta(q_{z,0}) = h_1 q_{z,0} + h_2 q_{z,0}^2 + h_3 q_{z,0}^3 + ... \quad (3)$$

where $q_{z,0}$ is the initial longitudinal coordinate with respect to the bunch center, R_{56}, T_{566} and U_{5666} are the first-, second- and third-order longitudinal dispersion terms respectively, δ=$\Delta E/E_0$ represents the energy deviation relative to the nominal beam energy, and h_1, h_2, h_3... are the first-, second- and third-order energy chirps respectively. From Eq. (2), for the initial beam with a specific $\delta(q_{z,0})$, a desired custom temporal profile can be achieved at the exit of the bunch compressor by using an appropriate combination of longitudinal dispersion terms.

Machine learning technique. The CGAN is used in this study to solve the one-to-many problems of temporal shaping. A CGAN consists of a pair of neural networks, the generator G and the discriminator D. The input of G is a noise component p_z and a label y, and the output $G(z|y)$ is a fake sample referred to as x_G. The input of D is either a real sample x_R from the training data pool or a fake sample x_G generated from G conditioned with its label y. The output of D is a scalar $D(x|y)$, which represents the probability that the input sample comes from the training data rather than being generated from G. The training of G and D is both evaluated with a value function $V(G, D)$ that depends on both G and D. The training of a CGAN can be summarized as [49]

$$\min_G \max_D V(G,D) = \mathbb{E}_{x\sim P_{data}(x)}[\log D(x|y)] + \mathbb{E}_{z\sim P_z(z)}[\log(1 - D(G(z|y)))]. \quad (4)$$

Data preparation and training. To obtain the training data for the CGAN, we choose a chicane-type magnetic compressor layout as shown in Fig. 1. Within an empirically large enough range, 10000 sets of R_{56}, T_{566} and U_{5666} data samples are stochastically generated. The initial beam is sent to the chicanes that have different stochastic R_{56}, T_{566} and U_{5666} values, and the final longitudinal coordinates of the beam at the exit of the chicanes are calculated with an accelerator simulation code *Accelerator Toolbox* [55]. The final temporal profile is then obtained and converted to a vector of length 200. The dispersion terms and corresponding temporal profile are treated as feature and label of the real data samples, respectively. A uniform distribution with dimensionality 100 is defined, from which a noise component is randomly selected and fed to G. Besides the noise component, a temporal profile y_G is also fed to G, which then generate fake dispersion terms x_G. The fake dispersion terms and the input temporal profiles are combined to be fake samples. The real and fake samples are then fed to D to determine whether it is real or fake. The loss function of D is defined as the cross entropy loss $\mathcal{L}$ between the output of D and the ground truth for an input sample,

$$\mathcal{L} = -\sum_{i=1}^{n} D(x_i|y_i) log(GT_i) \quad (5)$$

where $D(x_i|y_i)$ is the output of D for an input sample $\{x_i,y_i\}$, GT is the ground truth of the input samples that is equal to 1 and 0 for the real and fake samples, respectively, and n is the number of training samples in a training batch. The loss function of G is defined to be 1-$\mathcal{L}(x_G|x_G)$, representing the ability of G to generate fake samples to fool D.

The training of a CGAN is difficult because one has to ensure the balance between the G and D. When the loss of any one of the G and D converges quickly to zero while the other doesn't, it will cause the training to fail [56]. For this particular case, it is found that the training of D is much simpler than the training of G. To suppress the training of D, the learning rate of D is set to be one tenth of that of G, i.e. 0.0001 and 0.001 respectively. The training of the networks is implemented with Adam optimizer [57] for 10000 epochs on an open source ML library *tensorflow* [58]. The training time of the networks is about 20 minutes on a personal computer with a GeForce RTX 2070 Super Graphics Card.

Stochastic optimization algorithms. Three stochastic optimization algorithms, namely the PSO, GA and ES, are tested to realize a cusp-shaped and double-horn temporal profiles for comparison with the CGAN solver. The optimization function for the three tested stochastic optimization algorithms is to minimize the mean square error with respect to the target profile and the optimized variables are the R_{56}, T_{566} and U_{5666} of the bunch compressor. All the optimized variables are normalized to a range of [0, 1]. The initial population for the PSO and GA are 30 solutions randomly generated from a uniform distribution within the variable range. The crossover and mutation rate of the GA are set to be 0.9 and 0.1, respectively. For the PSO, the velocity weight factor is set to be 0.4. The acceleration coefficients of the group best experience and the personal experience are both set to be 1. For the ES, the initial setting is 0.5 for each optimized variable. To reduce the influence of randomness in the optimization, the three stochastic optimization algorithms are repeatedly performed for five times and only the best solutions obtained among the repeated optimizations are selected for comparison.

[1] Emma, P. et al. First lasing and operation of an ångstrom-wavelength free-electron laser. *Nature Photonics* **4,** 641–647 (2010).

[2] Allaria, E. et al. Two-stage seeded soft-X-ray free-electron laser. *Nature Photonics* **7,** 913–918 (2013).

[3] Byrd, J. M. et al. Tailored terahertz pulses from a laser-modulated electron beam. *Phys. Rev. Lett.* **96,** 164801 (2006).

[4] Bielawski, S. et al. Tunable narrowband terahertz emission from mastered laser–electron beam interaction. *Nature Physics* **4,** 390-393 (2008).

[5] Gschwendtner, E. & Muggli P. Plasma wakefield accelerators. *Nature Reviews Physics* **1,** 246–248 (2019).

[6] Cole, B. E., Williams, J. B., King, B. T., Sherwin, M. S. & Stanley, C. R. Coherent manipulation of semiconductor quantum bits with terahertz radiation. *Nature* **410,** 60–63(2001).

[7] Hau-Riege, S. P., London, R. A., Bionta, R. M., McKernan, M. A. & Baker, S. L. Damage threshold of inorganic solids under free-electron-laser irradiation at 32.5nm wavelength. *Appl. Phys. Lett.* **90,** 173128 (2007).

[8] Wilmink, G. J. & Grundt, J. E. Invited review article: current state of research on biological effects of terahertz radiation. *Journal of Infrared, Millimeter, and Terahertz Waves* **32,** 1074-1122 (2011).

[9] Grünbein, M. L. et al. Megahertz data collection from protein microcrystals at an X-ray free-electron laser. *Nature communications* **9,** 1-9 (2018).

[10] Litos, M. et al. High-efficiency acceleration of an electron beam in a plasma wakefield accelerator. *Nature* **515,** 92–95 (2014).

[11] England, R. J., Rosenzweig, J. B. & Travish, G. Generation and measurement of relativistic electron bunches characterized by a linearly ramped current profile. *Phys. Rev. Lett.* **100,** 214802 (2008).

[12] Ding, Y. et al. Beam shaping to improve the free-electron laser performance at the Linac Coherent Light Source. *Phys. Rev. Accel. Beams* **19,** 100703 (2016).

[13] Charles, T. K., Paganin, D. M., Boland, M. J. & Dowd, R. T. Beam by design: current pulse shaping through longitudinal dispersion control. in *Proc. 8th International Particle Accelerator Conference* (2017) `https://doi.org/10.18429/JACoW-IPAC2017-MOPIK055`.

[14] Sudar, N., Musumeci, P., Gadjev, I., Sakai, Y. & Fabbri, S. Demonstration of cascaded modulator-chicane microbunching of a relativistic electron beam. *Phys. Rev. Lett.* **120,** 114802 (2018).

[15] Shpakov, V., Anania, M. P., Bellaveglia, M., Biagioni, A., Bisesto, F. & Cardelli, F. Longitudinal Phase-Space Manipulation with Beam-Driven Plasma Wakefields. *Phys. Rev. Lett.* **122,** 114801 (2019).

[16] Bane, K. L. F., Chen P. & Wilson, P. B. On collinear wake field acceleration. *IEEE Transactions on Nuclear Science* **32,** 3524-3526 (2007).

[17] Saldin, E. L., Schneidmiller, E. A. & Yurkov, M. V. An analytical description of longitudinal phase space distortions in magnetic bunch compressors. *Nucl. Instrum. Methods Phys. Res., Sect. A* **483,** 516-520 (2002).

[18] Brown, K. L. A first- and second-order matrix theory for the design of beam transport systems and charged particle spectrometers. *Advances in Particle Physics* **1,** 71 (1968).

[19] Prat, E., Dijkstal, P., Ferrari, E. & Reiche, S. Demonstration of large bandwidth hard X-ray free-electron laser pulses at SwissFEL. *Phys. Rev. Lett.* **124,** 074801 (2020).

[20] Guetg, M. W., Beutner, B., Prat, E. & Reiche, S. Optimization of free electron laser performance by dispersion-based beam-tilt correction. *Phys. Rev. ST Accel. Beams* **18,** 030701 (2015).

[21] Piot, P. et al. Generation and characterization of electron bunches with ramped current profiles in a dual-frequency superconducting linear accelerator. Phys. Rev. Lett. **108**, 034801 (2012).

[22] Charles, T. K., Paganin, D. M. & Dowd, R. T. Caustic-based approach to understanding bunching dynamics and current spike formation in particle bunches. *Phys. Rev. Accel. Beams* **19,** 104402 (2016).

[23] Dunning, D. J., Jones, J. K., Castañeda Cortés, H. M., & Thompson, N. R. Multi-objective FEL design optimisation using genetic algorithms, in *Proc. 39th Free Electron Laser Conference* (2019) `http://doi.org/10.18429/JACoW-FEL2019-THP065`.

[24] Ding, Y., Bane, K. L. F. & Nosochkov, Y. M. Beam shaping for high-repetition-rate X-ray FELs. in *Proc. 39th Free Electron Laser Conference* (2019) `http://doi.org/10.18429/JACoW-FEL2019-THP035`.

[25] Mayet, F., Assmann, R., & Lemery, F. Longitudinal phase space synthesis with tailored 3D-printable dielectric-lined waveguides. *Phys. Rev. Accel. Beams* **23,** 121302 (2020).

[26] Yang, L., Robin, D., Sannibale, F., Steier, C. & Wan, W. Global optimization of an accelerator lattice using multiobjective genetic algorithms. *Nucl. Instrum. Methods Phys. Res., Sect. A* **609,** 50-57 (2009).

[27] Wu, J. Multi-dimensional optimization of a terawatt seeded tapered free electron laser with a multiobjective genetic algorithm. *Nucl. Instrum. Methods Phys. Res., Sect. A* **846,** 56 (2017).

[28] Huang, X. & Safranek, J. Nonlinear dynamics optimization with particle swarm and genetic algorithms for SPEAR3 emittance upgrade. *Nucl. Instrum. Methods Phys. Res., Sect. A* **757,** 48-53 (2014).

[29] Bruchon, N. et al. Free-electron laser spectrum evaluation and automatic optimization. *Nucl. Instrum. Methods Phys. Res., Sect. A* **871,** 20 (2017).

[30] Tana, Y., Nešić, D., Mareels, I. M. Y. & Astolfi A. On global extremum seeking in the presence of local extrema. *Automatica* **45,** 245–251 (2009).

[31] Schmitt, M. & Wanka, R. Particle swarm optimization almost surely finds local optima. *Theoretical Computer Science* **561,** 57-72 (2015).

[32] Dang, D. C. et al. Escaping local optima using crossover with emergent diversity. *IEEE Transactions on Evolutionary Computation* **22,** 484-497 (2017).

[33] Takahama, T. & Sakai, S. Constrained optimization by the ϵ constrained differential evolution with gradient-based mutation and feasible elites. in *Proc. 2006 IEEE international conference on evolutionary computation* (2006) `http://doi.org/10.1109/CEC.2006.1688283`.

[34] Wielgosz, M., Skoczeń, A. & Mertik, M. Using LSTM recurrent neural networks for monitoring the LHC superconducting magnets. *Nucl. Instrum. Methods Phys. Res., Sect. A* **867,** 40–50 (2017).

[35] Li, Y., Cheng, W., Li, H. Y. & Rainer, R. Genetic algorithm enhanced by machine learning in dynamic aperture optimization. *Phys. Rev. Accel. Beams* **21,** 054601 (2018).

[36] Emma, C. et al. Machine learning-based longitudinal phase space prediction of particle accelerators. *Phys. Rev. Accel. Beams* **21,** 112802 (2018).

[37] Scheinker, A., Edelen, A., Bohler, D., Emma, C. & Lutman, A. Demonstration of model-independent control of the longitudinal phase space of electron beams in the

Linac-coherent light source with Femtosecond resolution. *Phys. Rev. Lett.* **121,** 044801 (2018).

[38] Wan, J., Chu, P., Jiao, Y. & Li, Y. Improvement of machine learning enhanced genetic algorithm for nonlinear beam dynamics optimization. *Nucl. Instrum. Methods Phys. Res., Sect. A* **946,** 162683 (2019).

[39] Leemann, S. C. & Liu, S. Demonstration of Machine Learning-Based Model-Independent Stabilization of Source Properties in Synchrotron Light Sources. *Phys. Rev. Lett.* **123,** 194801 (2019).

[40] Xu, X., Zhou, Y. & Leng, Y. Machine learning based image processing technology application in bunch longitudinal phase information extraction. *Phys. Rev. Accel. Beams* **23,** 032805 (2020).

[41] Edelen, A. et al. Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems. *Phys. Rev. Accel. Beams* **23,** 044601 (2020).

[42] Wan, J., Chu, P. & Jiao, Y. Neural network-based multi-objective optimization algorithm for nonlinear beam dynamics. *Phys. Rev. Accel. Beams* **23,** 081601 (2020).

[43] Goodfellow, I. et al. Generative adversarial nets. *arXiv* `https://arxiv.org/abs/1406.2661` (2014).

[44] Tuan, Y. L. & Lee, H. Y. Improving conditional sequence generative adversarial networks by stepwise evaluation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **27,** 788-798 (2019).

[45] Sandfort, V., Yan, K., Pickhardt, P. J. & Summers, R. M. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Scientific Reports* **9,** 1-9 (2019).

[46] Denton, E. L., Chintala, S., Szlam, A. & Fergus, R. Deep generative image models using a Laplacian pyramid of adversarial networks. *Advances in neural information processing systems* **28,** 1486-1494 (2015).

[47] Zhu, J. Y., Park, T., Isola, P. & Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. in *Proc. IEEE international conference on computer vision* (2017) `https://doi.org/10.1109/ICCV.2017.244`.

[48] Ledig, C. et al. Photo-realistic single image super-resolution using a generative adversarial network. in *Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition* (2017) `https://doi.org/10.1109/CVPR.2017.19`.

[49] Mirza, M. & Osindero, S. Conditional generative adversarial nets. *arXiv* `https://arxiv.org/abs/1411.1784` (2014).

[50] Gurumurthy, S., Sarvadevabhatla, R. K., & Radhakrishnan, V. B. Deligan: Generative adversarial networks for diverse and limited data. in *Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition* (2017) `https://doi.org/10.1109/CVPR.2017.525`.

[51] Isola, P., Zhu, J. Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. in *Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition* (2017) `https://doi.org/10.1109/CVPR.2017.632`.

[52] England, R. J. et al. Sextupole correction of the longitudinal transport of relativistic beams in dispersionless translating sections. *Phys. Rev. ST Accel. Beams* **8,** 012801 (2005).

[53] Lagarias, J. C., Reeds, J. A., Wright, M. H. & Wright, P. E. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal of Optimization* **9,** 112–147 (1998).

[54] Charles, T. K., Paganin, D. M., Latina, A., Boland, M. J. & Dowd, R. T. Current-horn suppression for reduced coherent-synchrotron-radiationinduced emittance growth in strong bunch compression. *Phys. Rev. Accel. Beams* **20,** 030705 (2017).

[55] Terebilo, A. Accelerator modeling with MATLAB accelerator toolbox. in *Proc. the 2001 Particle Accelerator Conference* (2001) `https://doi.org/10.1109/PAC.2001.988056`.

[56] Rizzo, G. & Van, T. H. M. Adversarial text generation with context adapted global knowledge and a self-attentive discriminator. *Information Processing & Management* **57,** 102217 (2020).

[57] Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv* `https://arxiv.org/abs/1412.6980` (2014).

[58] Abadi, M. et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv* `https://arxiv.org/abs/1603.04467` (2016).

ACKNOWLEDGEMENT

This work is supported by National Natural Science Foundation of China (No. 11922512), Youth Innovation Promotion Association of Chinese Academy of Sciences (No. Y201904) and National Key R&D Program of China (No. 2016YFA0401900). The work of J. Wu is partially supported by the US Dependent of Energy Office of Science Early Career Research Program grant (FWP-2013-SLAC-100164).

AUTHOR CONTRIBUTIONS

This project was proposed by Jinyu Wan and Yi Jiao. The particle tracking simulation and machine learning was performed by Jinyu Wan. This manuscript was written by Jinyu Wan with contributions of Yi Jiao and Juhao Wu. The whole study was supervised by Juhao Wu and Yi Jiao.

COMPETING INTERESTS

The authors declare no competing interests.

DATA AVAILABILITY

The data that support the findings of this study are available in the Methods Section. Additional information is available from the corresponding author upon request.

Figures

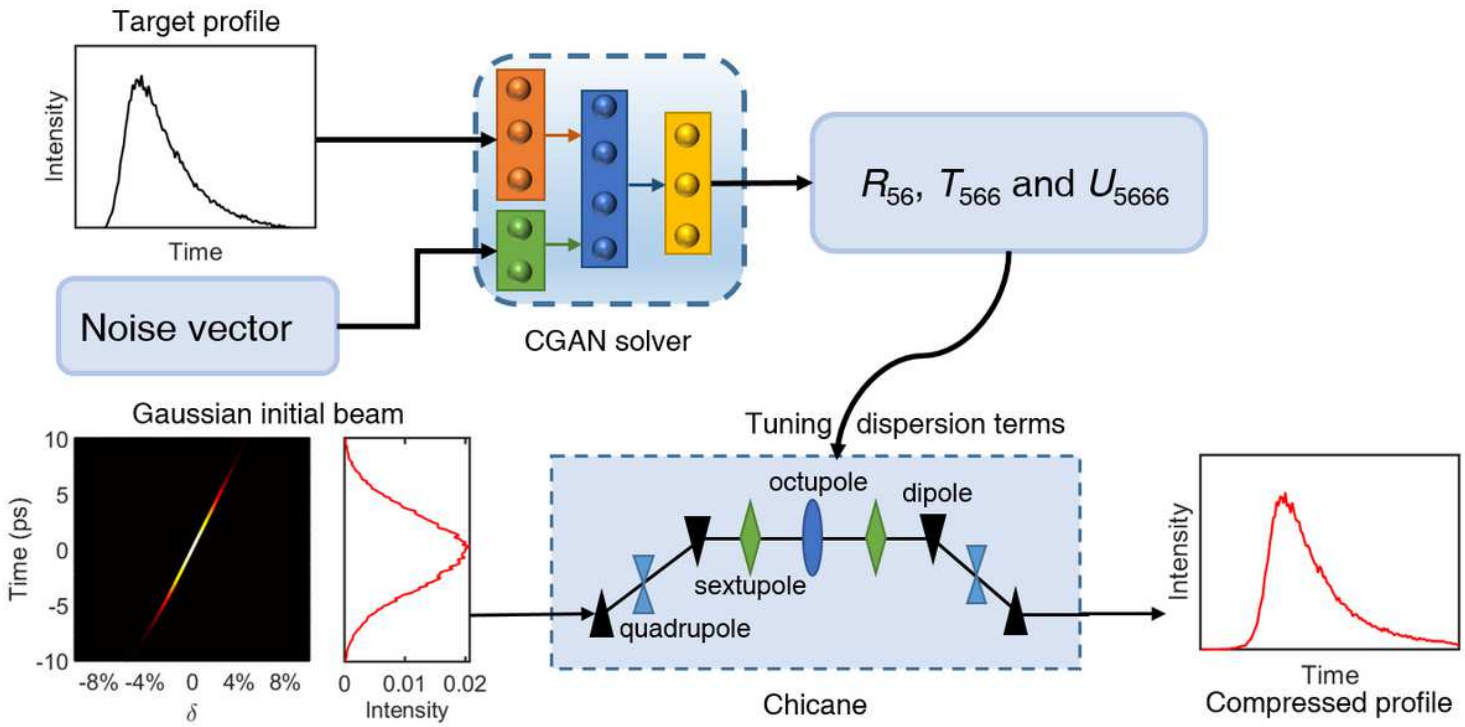

Figure 1

"Please see the Manuscript PDF file for the complete figure caption".

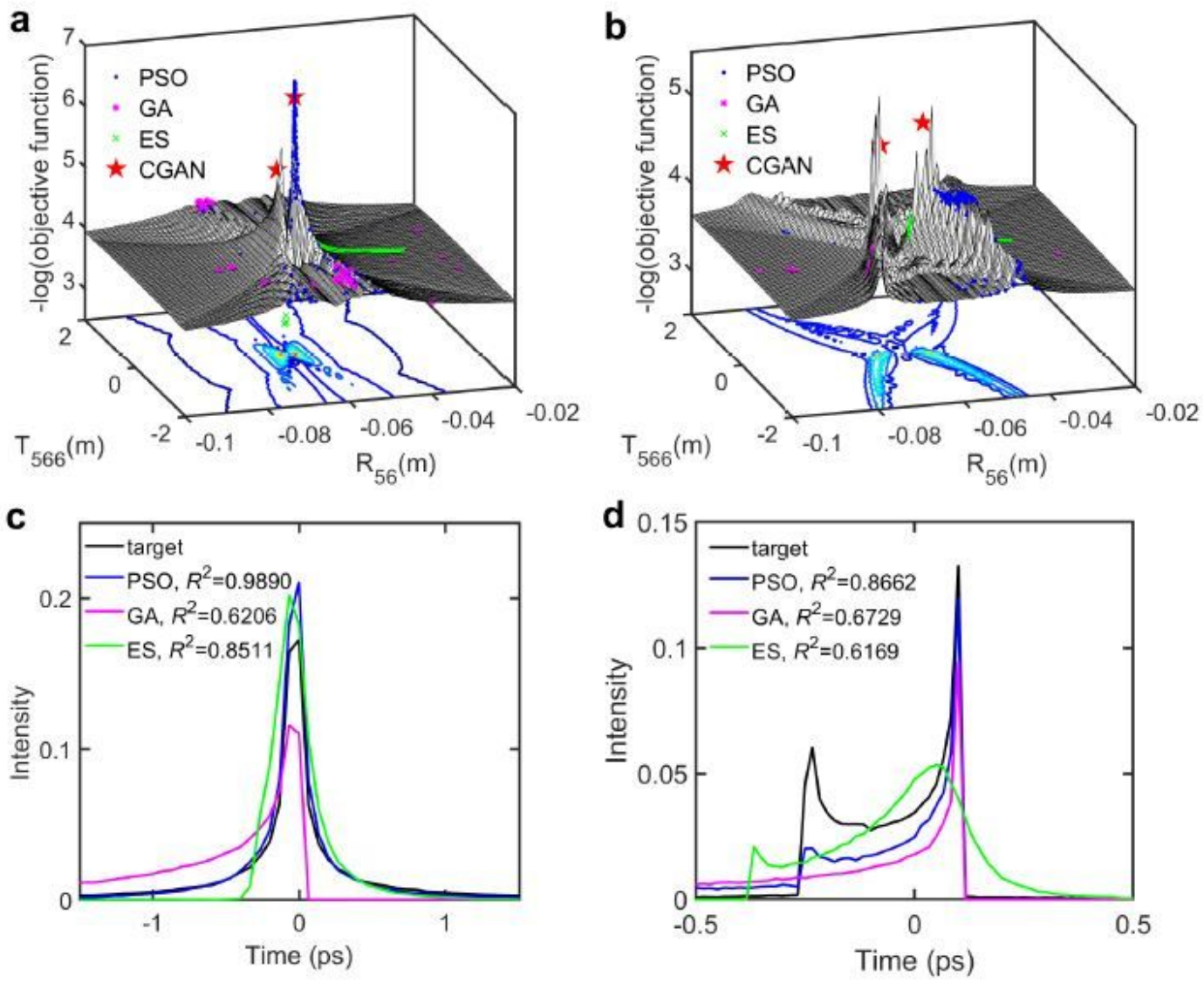

Figure 2

"Please see the Manuscript PDF file for the complete figure caption".

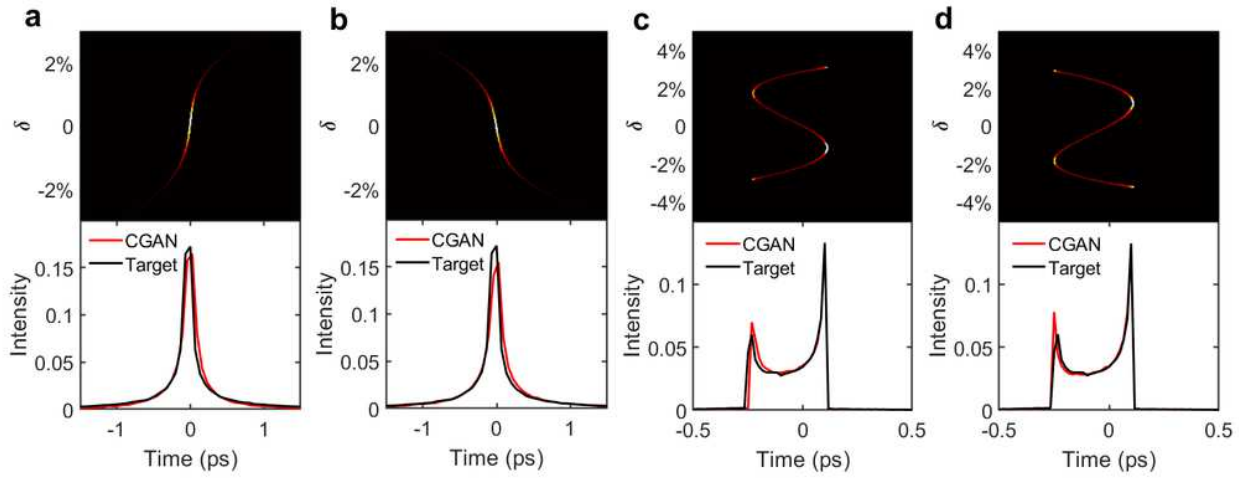

Figure 3

"Please see the Manuscript PDF file for the complete figure caption".

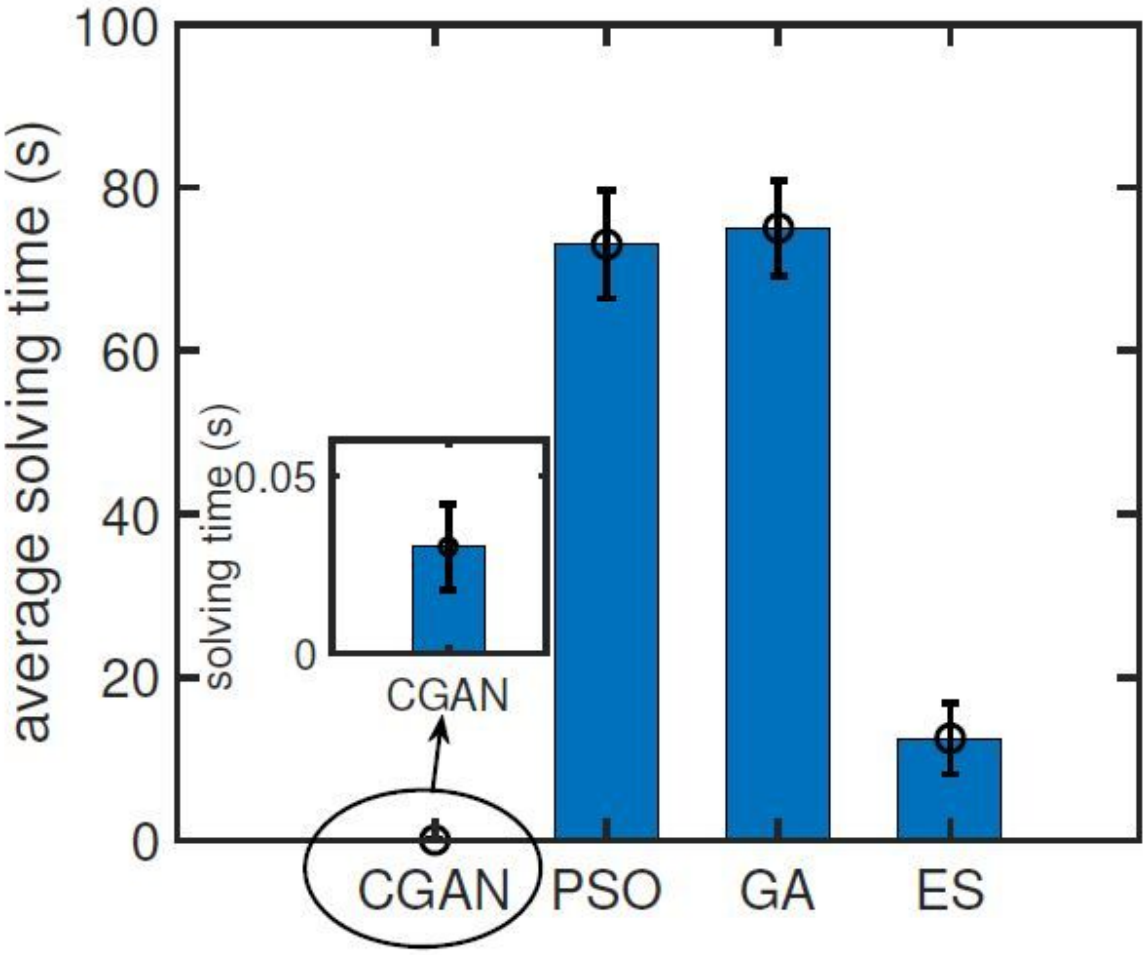

Figure 4

The average solving time of different methods for five repeated tests.

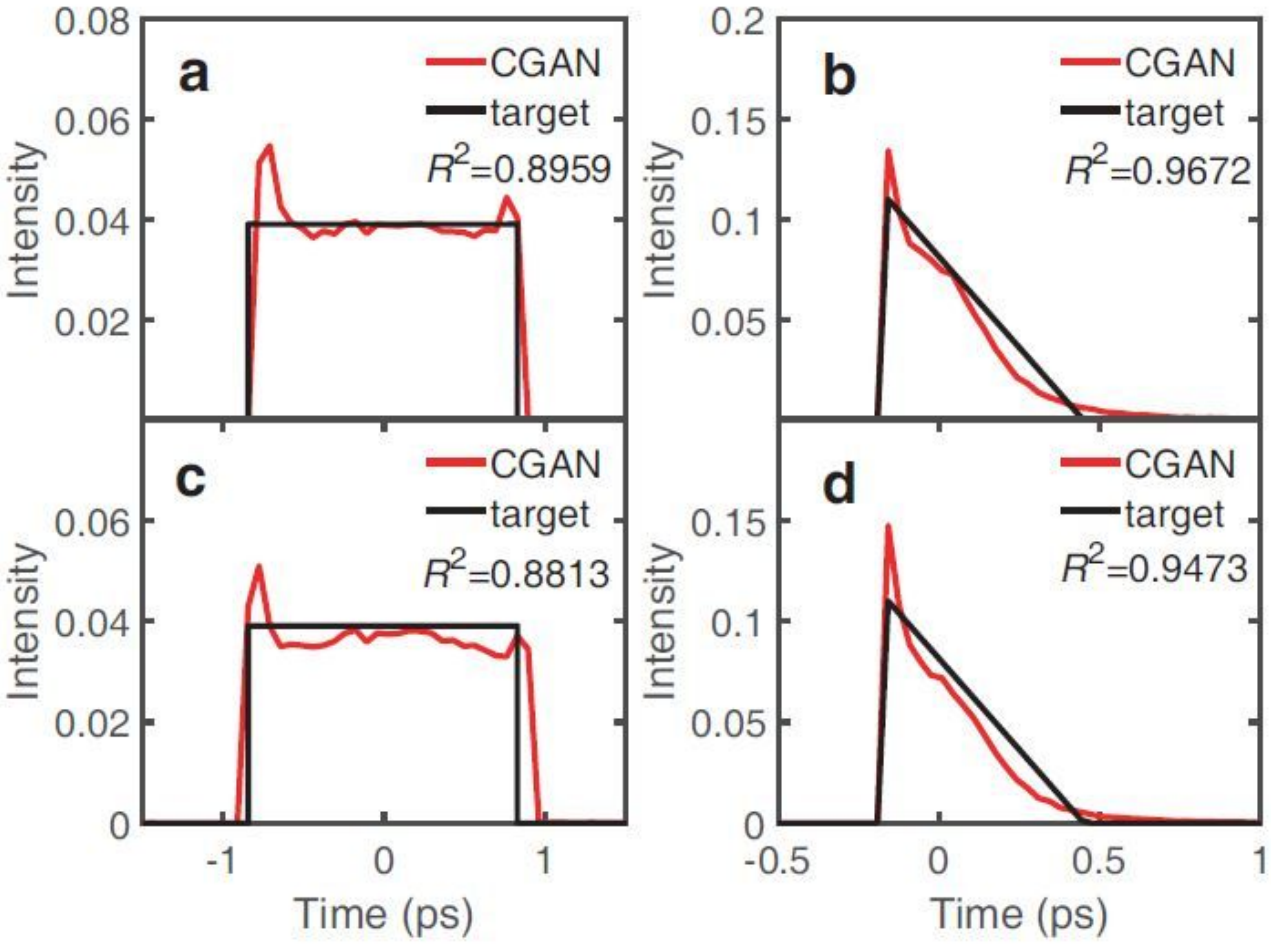

Figure 5

"Please see the Manuscript PDF file for the complete figure caption".