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

Jinyu Wan,1, 2 Yi Jiao,1, 2 and Juhao Wu3

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: March 12, 2021)

To control the temporal profile of a charged beam to meet requirements of various accelerator applications, a widely-used technique is bunch compression via 4-dipole chicanes that may sometimes have a one-to-many map. Current approaches based on stochastic optimization or supervised learning can be limited because of the one-to-many properties. Here we demonstrate how to construct a direct and real-time solver with the aid of a semi-supervised machine learning method, the conditional generative adversarial network (CGAN), to solve one-to-many problems of temporal shaping. Unlike supervised learning that can only learn one-to-one maps, the CGAN solver can learn the one-to-many dynamics and accurately predict required longitudinal dispersion terms for a chicane to realize desired custom temporal profiles without any priori knowledge. Besides, the CGAN solver can simultaneously give multiple different solutions for a one-to-many problem, which breaks the limitation of stochastic optimization methods of finding one solution instead of many.

I. INTRODUCTION

Much of the interest in advanced applications of particle accelerators such as free-electron lasers (FELs)1, 2, terahertz radiation3, 4 and plasma wakefield accelerators (PWFAs)5 has grown in recent decades. FELs and terahertz radiation are emerging as powerful imaging tools in various fields5–8 like physics, chemistry, biology and material science. PWFAs are potential to provide accelerating gradient of multi-GV/m level for future accelerators that promises beams of higher energy than using conventional radio-frequency-based accelerators for high energy physics and photon science9. To realize these advanced accelerator applications, a critical issue is to provide electron beams of particular temporal shapes to improve performance or quality of the electron or photon beam, namely the temporal shaping of electron beams11–14. In particular, beams of flat-top and cusp-shaped temporal profiles can be used in FELs to obtain high radiation power12–15, and a linearly ramped bunch is required by PWFAs to supply a high transformer ratio of wakefield16.

One typical temporal shaping method is to let an electron bunch with inhomogeneous energy distribution pass through a bunch compressor consisting of several bending magnets to realize various temporal profiles17. For beams with small transverse emittance and large energy spread, the contribution of geometrical terms in the transfer map can be neglected and the dispersion terms will tend to dominate the temporal profile15, 19. Hence, for the initial beam having an inhomogeneous energy chirp, a desired custom temporal profile can be achieved by using an appropriate combination of longitudinal dispersion terms (see details in Methods).

It is noted that in this scenario different combinations of longitudinal dispersion terms may realize the same target profile, i.e. the temporal shaping can sometimes turn to a one-to-many problem. Grid scan20 was first used to solve such a temporal shaping problem to realize a ramped profile with second-order approximation. However, when contributions of the higher-order longitudinal dispersion terms are not negligible, the time cost of grid scan will exponentially grow due to the consideration of higher-order terms. Afterwards, it has been shown21–24 that such one-to-many problems can be solved more efficiently with stochastic optimization methods like genetic algorithm (GA), particle swarm optimization (PSO) and extremum seeking (ES). Recently, an ES method combined with a supervised machine learning (ML) surrogate model was proposed25 that can reduce the tuning time for precise bunch control from hours of manual tuning to the minute level. Although many efforts have been made to explore further improvements in the performance of stochastic optimization methods, the 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 optima25–28. In addition, for one-to-many problems that have multiple available solutions, the optimization usually stops when the first solution is found with others missed. The finally obtained results can be highly dependent on the choice of initial solution. To break these limitations, here we propose to construct a direct and real-time solver for one-to-many problems of temporal shaping with the aid of ML.

Over the past decade, great progress has been made in the ML field, leading to a revolutionary change in the way of processing complex and massive data in many areas29–32. Particle accelerators, as a collection of multiple complex physical subsystems, has profited from ML since 1980s33. In recent years, ML has attracted increasing interests of accelerator experts as a powerful tool to re-
veal the complicated correlations between various accelerator parameters [34–41]. Most ML applications in the accelerator field are based on using fast supervised ML surrogate models to predict complex accelerator parameters by learning from the previously existing simulating or experimental data without further simulation or experiment. It is noticed that, however, most supervised ML surrogate models 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 [30]. While for the inverse problem, namely, predicting the accelerator settings for a desired custom temporal profile, the supervised ML methods like multilayer perceptron could fail to give the right answer because multiple solutions possibly exist. To break this limitation, here we introduce another ML method, the generative adversarial networks (GANs) [42], that has potential to handle one-to-many problems [43, 44].

The GANs are emerging techniques of semisupervised ML that have become one of the state-of-the-art techniques to solve difficulties in image synthesis [45], style transfer [46], and image superresolution [47]. Instead of the approach of minimizing mean error in most supervised ML methods, GANs take a different approach to capture the true distribution of 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. [48] to a conditional framework by making both $G$ and $D$ class-conditional. This extended approach, called conditional GAN (CGAN), allows one to directly see the data generation process by adding an additional label to the input of the generator. The CGAN has proven to be effective to create images with a target class [49, 50].

In this paper, taking problems of temporal profile shaping as examples, we demonstrate for the first time how to construct a CGAN solver to solve one-to-many problems in the particle accelerator field. It is shown that the trained CGAN solver can accurately and quickly predict the required longitudinal dispersion terms to realize desired custom profiles. Three typical profiles, namely the flat-top, cusp-shaped and ramped profiles as well as two non-typical profiles that are too complex to realize with the conventional stochastic optimization methods, are tested and realized with the predictions of the CGAN solver. When multiple solutions exist, the CGAN solver remains effective to give different solutions that can result in the same target profile.

II. RESULTS

Temporal shaping scheme based on CGAN solver. In our temporal shaping scheme (see Fig. 1 and Methods), we consider the third-order approximation of the longitudinal dispersion. The first- to third-order longitudinal dispersion terms, i.e. $R_{56}$, $T_{566}$ and $U_{5666}$, labeled with the corresponding final temporal profile are taken as training data. The generator $G$ is trained as a solver that can produce fake samples of longitudinal dispersion terms to realize new temporal profiles. Meanwhile, the 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. Thanks to the noise component $p_z$, there is a prospect that the solver can produce different solutions of a one-to-many problem.

In this study, we use an initial electron beam (see Fig. 1) of Gaussian charge-density distribution having an energy chirp up to 5th order, with the same beam parameters as in Ref. 15. Besides, we choose a chicane-type magnetic compressor layout as shown in Fig. 1. Such a layout of chicane is chosen so 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 [51]. The initial beam is sent to the chicanes that have specific $R_{56}$, $T_{566}$ and $U_{5666}$ values, and the desired custom temporal profiles are expected to be obtained at the exit of the chicanes.

Note that here we only consider a simple chicane and neglect collective effects and beam self-interaction effects such as longitudinal space charge (LSC) and coherent synchrotron radiation (CSR). This, however, does not impact at all our primary goal that is to show proof-of-principle demonstrations of the CGAN solver to solve one-to-many problems of temporal shaping. Nevertheless, this method would be easily transferable to other examples or other settings with these detrimental effects like LSC and CSR taken into account.

| Target Profile | $R_{56}$(mm) | $T_{566}$(mm) | $U_{5666}$(m) | $R^2$ |
|----------------|--------------|--------------|---------------|-------|
| flat-top 1     | -49.9        | -18.2        | -5.00         | 0.8959|
| flat-top 2     | -49.6        | -22.9        | -4.45         | 0.8813|
| cusp-shaped 1  | -61.5        | -32.4        | 11.8          | 0.9494|
| cusp-shaped 2  | -65.6        | 1.63         | -16.0         | 0.9610|
| ramped 1       | -66.5        | 31.8         | 1.33          | 0.9672|
| ramped 2       | -66.8        | 36.9         | 3.16          | 0.9473|

Realization of three typical temporal profiles. To demonstrate the performance of the CGAN solver, three typical temporal profiles that are widely used in advanced accelerator applications are selected as test profiles (Fig. 2(a)), which are flat-top, cusp-shaped and ramped profiles. The three profiles combined with random noises are fed to the trained solver which finally results in multiple...
FIG. 1. Schematic diagram of the CGAN framework in this study. The CGAN is trained with the stochastically generated longitudinal dispersion term data labeled with the final temporal profile. The final temporal profile is obtained by letting the initial beam pass through a chicane having stochastic $R_{56}$, $T_{566}$ and $U_{5666}$. The color of the initial beam from blue to yellow represents the charge-density from low to high. $L$ is the loss function of $D$ that defined as the cross entropy loss between the output of $D$ and the ground truth for a input sample.

FIG. 2. Using the CGAN solver to realize three typical temporal profiles. (a) The three test target temporal profiles. (b) Charge-density and energy distribution of the initial beam same as in Fig. 1. (c) The temporal profiles resulted from the CGAN solver (red solid lines) and the target temporal profiles (black solid lines).

sets of fake $R_{56}$, $T_{566}$ and $U_{5666}$ samples. Two randomly selected fake samples for each test profile are listed in Table 1 and the final temporal profiles resulted from the fake samples are illustrated in Fig. 2(c). 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 flatten with an additional bunch compressor [52] 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.

The results in Fig. 2(c) indicate that the CGAN solver can predict the longitudinal dispersion terms for the custom desire profiles with high accuracy. Compared to current approaches based on grid scan or 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 sec-
FIG. 3. Longitudinal phase space distribution and temporal profiles of two different solutions predicted by the CGAN solver. (a) and (b) represent cusp-shaped 1 and cusp-shaped 2 in Table I, respectively.

FIG. 4. The temporal profiles obtained with different methods for two additional complex cases. (a) and (b) represent a double-horn structure and a single-horn structure, respectively. The red and blue lines are temporal profiles resulted from the CGAN solver. The cyan, pink and green lines are the best optimization results obtained with the PSO, GA and ES, respectively. $R^2$ is the determination coefficient.

We have demonstrated for the first time how to construct a CGAN solver for one-to-many problems of temporal shaping. By learning from the stochastically generated data, a trained CGAN solver can quickly and accurately predict multiple available combinations of the dispersion terms up to 3rd order to realize desired custom temporal profiles, without any physical priori knowledge of beam dynamics. The CGAN solver remains effective to realize three widely-used temporal profiles, namely the flat-top, cusp-shaped and ramped profiles, and two other more challenging temporal profiles.

This method can be easily transferable to other similar 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 more one-to-many problems to reduce the simulating or operational time by orders of magnitude.

IV. METHODS

Bunch compression mechanism. In a system of magnetic elements, a transfer map $M$ describes the relation of initial conditions $\zeta_i$ and final conditions $\zeta_f$, which can be symbolically written in the form $\zeta_f = M\zeta_i$. A Taylor map [18, 19] that represents the final conditions as a...
Taylor series of the initial conditions can be described as
\[
\zeta_j^f = \sum_k R_{jk} \zeta_k^0 + \sum_{kl} T_{jk} \zeta_k^l \zeta_l^0 + \sum_{klm} U_{jklm} \zeta_k^l \zeta_l^m + \ldots \tag{1}
\]
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.

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 energy spread, the contribution of geometrical terms can be described as
\[
q_{z,f} = q_{z,0} + R_{566} \delta(q_{z,0}) + T_{566} \delta(q_{z,0})^2 + U_{566} \delta(q_{z,0})^3 + \ldots \tag{2}
\]
\[
\delta(q_{z,0}) = h_1 q_{z,0} + h_2 q_{z,0}^2 + h_3 q_{z,0}^3 + \ldots \tag{3}
\]
where \(q_{z,0}\) is the initial longitudinal coordinate with respect to the bunch center, \(R_{566}, T_{566}\) and \(U_{566}\) are the first- and higher-order longitudinal dispersion terms respectively, \(\delta = \Delta E/E_0\) represents the energy deviation relative to the nominal beam energy, and \(h_1, h_2, h_3, \ldots\) are the first- and higher-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, i.e., the generator \(G\) and the discriminator \(D\). The input of \(G\) is the combination of a noise \(p_z\) 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 \(y\), and 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 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
\[
\min_G \max_D V(G, D) = \mathbb{E}_{x \sim P_{data}(x)} [\log D(x|y)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z|y)))]. \tag{4}
\]

**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 \(R_{566}, T_{566}\) and \(U_{566}\) data samples are stochastically generated. The initial beam is sent to the chicanes that have different stochastic \(R_{566}, T_{566}\) and \(U_{566}\) values, and the temporal profiles at the exit of the chicanes are calculated with an accelerator simulation code Accelerator Toolbox [53]. The final temporal profiles labeled with the corresponding \(R_{566}, T_{566}\) and \(U_{566}\) are treated as training data of the CGAN.

The temporal profile data is converted to a vector of length 200 and then fed to \(D\) combined with the longitudinal dispersion terms. A uniform distribution with dimensionality 100 is defined, from which a noise prior is selected and fed to \(G\). The training of a CGAN is difficult because one has to ensure the balance between \(G\) and \(D\). When the loss of any one of the \(G\) and \(D\) converges quickly to zero, it will cause the training to fail [54]. 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 a tenth of the training of \(G\). To further demonstrate the performance of the CGAN solver, three stochastic optimization algorithms, namely the PSO, GA and ES, are tested to realize the two non-typical temporal profiles in Fig. [1] for comparison with the results obtained with the CGAN solver. The optimization function for the three tested stochastic optimization algorithms is minimizing the mean square error with respect to the target profile and the optimized variables are the \(R_{566}, T_{566}\) and \(U_{566}\) of the bunch compressor. All the optimized variables are normalized to a range of [0, 1]. The initial population of the PSO and GA are 100 solutions randomly generated from a uniform distribution within the variable range. The initial setting of ES is 0.5 for each optimized variable. For a fair comparison, 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.
[7] S. P. Hau-Riege, R. A. London, R. M. Bionta, M. A. McKernan, S. L. Baker, Damage threshold of inorganic solids under free-electron-laser irradiation at 32.5nm wavelength, Appl. Phys. Lett. 90, 173128 (2007).

[8] G. J. Wilmink, J. E. Grundt, Invited review article: current state of research on biological effects of terahertz radiation, Journal of Infrared, Millimeter, and Terahertz Waves 32, 1074-1112 (2011).

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

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

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

[12] Y. Ding, K. L. F. Bane, W. Colocho, F. J. Decker, P. Emma, J. Frisch, et al., Beam shaping to improve the free-electron laser performance at the Linac Coherent Light Source, Phys. Rev. Accel. and Beams 19, 100703 (2016).

[13] N. Sudar, P. Musumeci, I. Gadjiev, Y. Sakai, S. Fabbi, Demonstration of cascaded modulator-chicane microbunching of a relativistic electron beam, Phys. Rev. Lett. 120, 114802 (2018).

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

[15] T. K. Charles, D. M. Paganin, M. J. Boland, R. T. Dowd, Beam by design: current pulse shaping through longitudinal dispersion control, in proceedings of IPAC2017, Copenhagen, Denmark (2017), http://accelconf.web.cern.ch/ipac2017/papers/mopik055.pdf

[16] K. L. F. Bane, P. Chen P. B. Wilson, On collinear wake field acceleration, No. SLAC-PUB—3662, https://accelconf.web.cern.ch/ipac2017/papers/mopik055.pdf

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

[18] K. L. Brown, A first- and second-order matrix theory for the design of beam transport systems and charged particle spectrometers, SLAC Report 75 (1982), https://cds.cern.ch/record/283218/files/SLAC-75.pdf

[19] D. C. Carey, K. L. Brown, F. Rothacker, Third-order TRANSPORT: A computer program for designing charged particle beam transport systems, SLAC-R-95-462, https://www.osti.gov/servlets/purl/97260

[20] P. Piot, C. Behrens, C. Gerth, M. Dohlus, F. Lemery, D. Mihalcea, P. Stoltz, and M. Vogt, Generation and characterization of electron bunches with ramped current profiles in a dual-frequency superconducting linear accelerator, Phys. Rev. Lett. 108, 034801 (2012).

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

[22] D. J. Dunning, J. K. Jones, H. M. Castañeda Cortés, N. R. Thompson, Multi-objective FEL design optimisation using genetic algorithms, in Proceedings of 39th Free Electron Laser Conf., Hamburg, Germany (2019), http://accelconf.web.cern/ch/fel2019/papers/thp005.pdf

[23] Y. Ding, K. L. F. Bane, Y. M. Nosochkov, Beam shaping for high-repetition-rate X-ray FELs, in Proceedings of 39th Free Electron Laser Conf., Hamburg, Germany (2019), http://accelconf.web.cern/ch/fel2019/papers/thp035.pdf

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

[25] A. Scheinker, A. Edelen, D. Bohler, C. Emma, A. Lutman, 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).

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

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

[28] D. C. Dang, T. Friedrich, T. Kötzing, P. K. Lehre, P. S. Oliveto, D. Sudholt, A. M. Sutton, Escaping local optima using crossover with emergent diversity, IEEE Transactions on Evolutionary Computation 22, 484-497 (2017).

[29] L. Zdeborová, Machine learning: New tool in the box, Nature Physics 13, 420-421 (2017).

[30] R. Ramakrishnan, O. A. von Lilienfeld, Machine learning, quantum chemistry, and chemical space, Reviews in computational chemistry 30, 225-256 (2017).

[31] A. L. Tarca, V. J. Carey, X. Chen, et al., Machine learning and its applications to biology, PLoS Comput Biol 3, e116 (2007).

[32] I. Kononenko, Machine learning for medical diagnosis: history, state of the art and perspective, Artificial Intelligence in medicine 23, 89-109 (2001).

[33] C. Peterson, Track finding with neural networks, Nucl. Instrum. Methods Phys. Res., Sect. A 279, 537-545 (1989).

[34] M. Wielgosz, A. Skoceń, M. Mertik, Using LSTM recurrent neural networks for monitoring the LHC superconducting magnets, Nuclear Inst. and Methods in Physics Research, A 867, 40–50 (2017).

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

[36] C. Emma, A. Edelen, M. J. Hogan, B. O’Shea, G. White, V. Yakimenko, Machine learning-based longitudinal phase space prediction of particle accelerators, Phys. Rev. Accel. and Beams 21, 112802 (2018).

[37] J. Wan, P. Chu, Y. Jiao, Yongjun Li, Improvement of machine learning enhanced genetic algorithm for nonlinear beam dynamics optimization, Nuclear Inst. and Methods in Physics Research, A 946, 162683 (2019).

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

[39] X. Xu, Y. Zhou, Y. Leng, Machine learning based image processing technology application in bunch longitudinal phase information extraction, Phys. Rev. Accel. and Beams 23, 032805 (2020).
A. Edelen, N. Neveu, M. Frey, Y. Huber, C. Mayes, A. Adelmann, Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems, Phys. Rev. Accel. and Beams 23, 044601 (2020).

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

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Advances in neural information processing systems 27, 2672-2680 (2014).

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

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

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

J. Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation with conditional adversarial networks, in proceedings of IEEE Conf. CVPR, Honolulu, Hawaii, USA (2017), https://openaccess.thecvf.com/content_cvpr_2017/papers/Isola_Unpaired Image-to-Image_Translation_CVPR_2017_paper.pdf.

C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, Photo-realistic single image super-resolution using a generative adversarial network, in proceedings of IEEE Conf. CVPR, Honolulu, Hawaii, USA (2017), https://openaccess.thecvf.com/content_cvpr_2017/papers/Ledig_Photo-Realistic_Single_Image_CVPR_2017_paper.pdf.

M. Mirza, S. Osindero, Conditional generative adversarial nets, arXiv:1411.1784 (2014).

S. Gurumurthy, R. K. Sarvadevabhatla, and V. B. Radhakrishnan, Deligan: Generative adversarial networks for diverse and limited data, in proceedings of IEEE Conf. CVPR, Honolulu, Hawaii, USA (2017), https://openaccess.thecvf.com/content_cvpr_2017/papers/Gurumurthy_DeliGAN_Generative_CVPR_2017_paper.pdf.

P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, Image-to-image translation with conditional adversarial networks, in proceedings of IEEE Conf. CVPR, Honolulu, Hawaii, USA (2017), https://openaccess.thecvf.com/content_cvpr_2017/papers/Isola_Image-To-Image_Translation_CVPR_2017_paper.pdf.

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

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

A. Terebilo, Accelerator modeling with MATLAB accelerator toolbox, in Proceedings of the 19th Particle Accelerator Conference, Chicago, IL, USA (2001), https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=988056.

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

D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980 (2014).

M. Abadi, A. Agarwal, P. Barham P, et al., Tensorflow: Large-scale machine learning on heterogeneous distributed systems, arXiv: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).