

LEARNING TO LASE: MACHINE LEARNING PREDICTION OF FEL BEAM PROPERTIES

A.E. Pollard*, D.J. Dunning, M. Maheshwari
ASTeC, Cockcroft Institute, STFC Daresbury Laboratory, UK

Abstract

Accurate prediction of longitudinal phase space and other properties of the electron beam are computationally expensive. In addition, some diagnostics are destructive in nature and/or cannot be readily accessed. Machine learning based virtual diagnostics can allow for the real-time generation of longitudinal phase space and other graphs, allowing for rapid parameter searches, and enabling operators to predict otherwise unavailable beam properties. We present a machine learning model for predicting a range of diagnostic screens along the accelerator beamline of a free-electron laser facility, conditional on linac and other parameters. Our model is a combination of a conditional variational autoencoder and a generative adversarial network, which generates high fidelity images that accurately match simulation data. Work to date is based on start-to-end simulation data, as a prototype for experimental applications.

INTRODUCTION

Free-electron lasers (FELs) are sources of ultra-short and ultra-bright pulses of light, which are used to reveal the dynamics of important processes across various scientific disciplines [1]. For effective and efficient facility operation, it is necessary to rapidly reconfigure and optimise different setups to meet user needs. Machine diagnostics are essential for this task, and while some measurements can be taken non-invasively and on every shot, others are destructive to the beam, relatively slow, or otherwise constrained. For example, the longitudinal phase space (LPS) of the beam is critically important for FEL performance but this is an invasive measurement that can only be made at locations with dedicated hardware. Simulations are commonly used to supplement the experimental data; however, they can be computationally expensive and often require significant iteration to improve their match to experimental conditions. Machine learning offers the potential to leverage the large volumes of accessible diagnostic and simulation data to build surrogate models that enable accurate modelling with real-time execution, thereby delivering virtual diagnostics and rapid optimisation of beam conditions on demand.

Artificial Neural Networks have shown great utility in the generation of complex data. From synthesised speech to faces of non-existent people, several architectures have been developed to effectively generate artificial data that is indistinguishable from the training data. Autoencoders, autoregressive models, and generative adversarial networks comprise the three main classes of these generative networks and the focus of modern research.

* amelia.pollard@stfc.ac.uk

Autoencoders are an excellent choice for generating complex and variable data in an unsupervised manner. Their ability to condense inputs into a significantly lower dimensional representation before reconstructing the input from this representation provides a powerful method of learning the key components of target images, such as input parameters. Variational autoencoders [2] build on this to codify the latent space representation as a series of probability distributions which can be further extended by making those distributions conditional on input parameters. This Conditional Variational Autoencoder (CVAE) model enables effective learning of the parameterised generation of images.

Generative Adversarial Networks [3] (GAN) represent the current state-of-the-art approach to image generation and function on a principle of adversarial learning. Adversarial learning functions by one network attempting to generate realistic data, while another network attempts to discern real data from generated data. This technique enables image generation with fidelity orders of magnitude higher than that of generative networks alone, but it is prone to stability issues and is highly sensitive to hyperparameter choices.

Our work combines a CVAE and a GAN into the CVAE-GAN architecture [4] and uses a combination of these powerful techniques to produce high-fidelity longitudinal phase space graphs for arbitrary parameter configurations as a virtual diagnostic for end users. We also demonstrate an ability to search the space of LPS graphs for particular graphs as drawn by an end user, allowing for highly customisable longitudinal beam profiles.

RELATED WORK

The first steps to incorporating image recognition into particle accelerator control using convolutional neural networks were taken in Edelen et al. [5] and developed further in Scheinker et al. [6]. Conversely, machine learning *prediction* of longitudinal phase space from machine parameters has been explored, with good results, as in Emma et al. (2018 and 2019) [7, 8]. While these simple networks seem to lead to artifacts and sub-optimal reconstruction of fine structure details, more complex networks featuring convolutional and upsampling layers do appear to provide some improvement [9]. The use of additional information besides machine settings, in the form of non-invasive shot-to-shot spectral measurements, has also been shown to improve the accuracy of LPS predictions [10]. Additionally, recent work based solely on experimental data has demonstrated LPS prediction at significantly higher resolution to previous studies [11].

METHODOLOGY

We used a dataset of longitudinal phase space images and associated parameters, as described in [12] and briefly outlined here. This was generated from the results of previous work, using a Multi-Objective Genetic Algorithm [13] to find optimal beam characteristics on the XARA lattice [14]. The XARA lattice is a possible upgrade to the existing CLARA facility [15], using high-gradient X-band RF technology. A diagram of the XARA lattice can be found in Fig. 1. 10,000 start-to-end simulations of the accelerator were generated using the simulation code ASTRA [16] (up to the linac 1 exit) and Elegant [17] for the remainder. Using the 6D bunch distributions produced, LPS images were created by calculating the maximum and minimum values of the z-positions and beam energies for each individual distribution and binning the particles in a 2D histogram defined by this region of interest (ROI) to create 100×100 pixel images. Though not used in this study, another set of images was generated for which the maxima and minima were taken over all distributions, for a fixed ‘screen size’ (referred to as non-ROI or fixed extent images).

Each example in the dataset consists of an 100×100 pixel LPS image at a location immediately upstream of the undulator line, 4 real values describing the bounds of the ROI in the fixed extent, and 17 accelerator parameters describing the phase and amplitude of the linacs, the laser heater factor, the dechirper factor, and the bunch compression. In this work, all fixed extent images shown were generated by placing the ROI image in accordance with the bounds generated by the network.

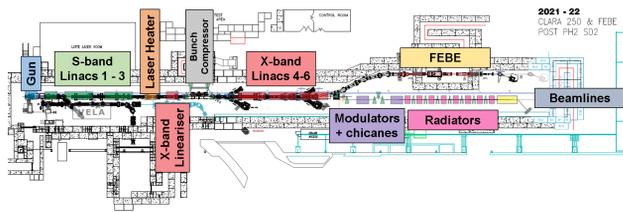


Figure 1: Schematic layout of CLARA with the potential XARA upgrade overlaid. The total length is approximately 90 m. The LPS images used in this study were generated at a location following the X-band linacs and before the modulator undulators.

Following the approach taken in previous studies [7], we began by developing a fully connected parameter to image generative model as part of Maheshwari et al. [12]. A parameter to CNN generative model was also tried, which performed poorly, generating approximations which were of the correct scale but lacking a well-defined structure. This work led directly to the development of this more complex architecture.

We then developed a conditional variational autoencoder model and experimented with CNN and DNN based generators to produce high quality images. However, as seen in previous works, the generator showed a tendency to produce a great many artifacts and fine details were lost. We therefore created a deep neural network which utilised a

convolutional conditional variational autoencoder, with the addition of a discriminator, resulting in a model known as a CVAE-GAN (as shown in Fig. 2) [4]. A discriminator endeavours to identify real and generated images. When this is affixed to the generative model (the CVAE), the weights of the discriminator are frozen and the generated images are intentionally mislabelled as being real images. The resulting error is then backpropagated through the generative network to train the generator to produce images which would ‘fool’ the discriminator. This error is combined with an absolute difference error metric between the input and output images, as well as the Kullback-Liebler divergence constraint on the latent space of the CVAE.

Hyperparameters were found by Bayesian optimisation to minimise absolute difference over a validation set composed of 10% of the total dataset. The network was trained utilising 90% of the dataset, with 10% held back for testing. The Adam optimiser was used with a learning rate of 1.0×10^{-3} , with the error function defined in Eq. (1). Note that D_{kl} represents the Kullback-Liebler divergence [18].

$$E(x) = D_{kl}(N(0, 1) || P(x|\mu, \sigma, y)) + abs(x - \bar{x}) - (x_d \log(\hat{x}_d) + (1 - x_d) \log(1 - \hat{x}_d)) \quad (1)$$



Figure 2: The model of the CVAE-GAN, with the encoder highlighted in blue, the latent space layer highlighted in orange, the decoder highlighted in green, and the discriminator highlighted in purple. Note the additional 1×4 output layer which generates the position of the region of interest in the fixed extent image.

RESULTS

We found that the CVAE-GAN model allowed for the real-time generation of longitudinal phase space graphs with high verisimilitude to simulation. This therefore achieves the stated goal of providing a virtual diagnostic of longitudinal phase space that would allow operators to carefully choose parameters to achieve specific bunch energy profiles.

Most notably, our technique delivers high quality representation of fine-grain details as can be seen in Fig. 3. This is a consequence of the application of the discriminator, which constrains the generator to produce more realistic images where more traditional error functions, such as absolute difference, generally result in blurring of fine details.

Given that the forward pass of neural networks can be run in linear time, this allows for the rapid sampling of LPS graphs from the space. In combination with Bayesian

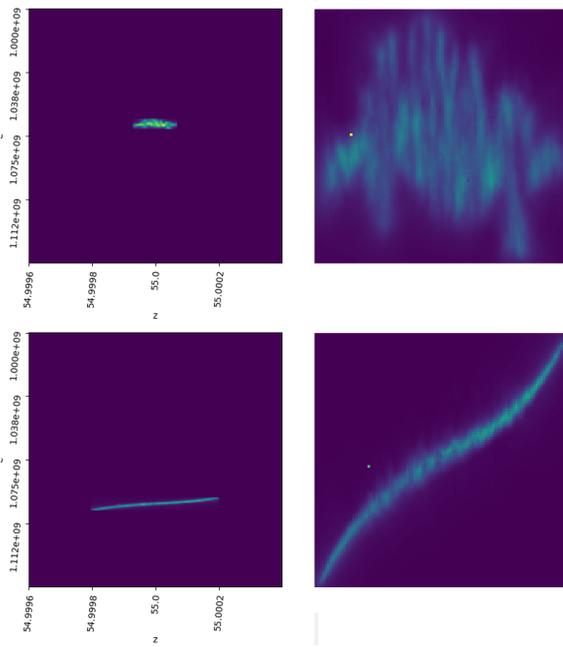


Figure 3: A pair of examples from the LPS GUI showing both the fixed extent and region of interest. Note the fine-grain details of the region of interest image, which are generally lost in simpler models.

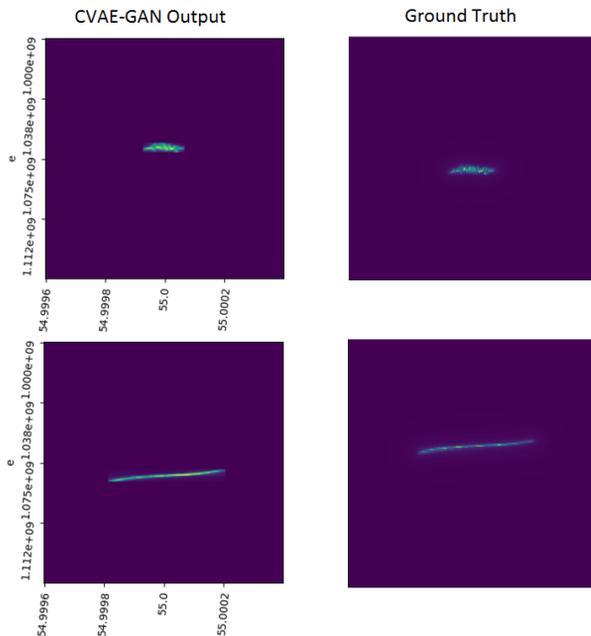


Figure 4: Fixed extent images and ground truth from Elegant simulations with the same parameters. Variations in the position of ROI are due to scale differences in the output. Significantly, the finer detail structure of the LPS graph is strongly matched to the output of Elegant.

Optimisation, we can therefore perform a guided search for image similarity with an LPS graph drawn by an operator, as shown in the example in Fig. 5, to yield the required machine settings. This feature is one part of a graphical user interface (GUI) for the model, shown in Fig. 6.

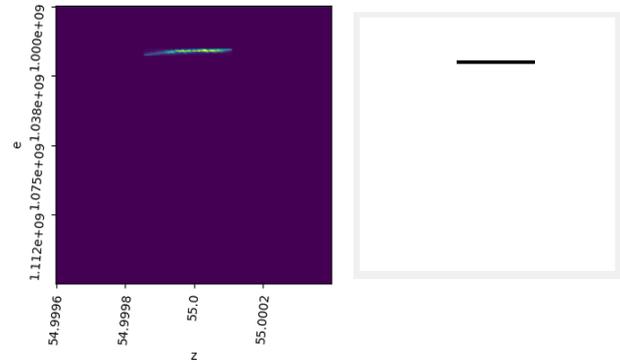


Figure 5: An example of the drawing interface with target drawn by operator (right) and the resulting LPS graph (left) which most closely matched the operator’s drawing. On a mid-range laptop, this search took approximately 20 seconds.

CONCLUSION

Utilising the CVAE-GAN architecture allows for high fidelity linear time virtual diagnostics of the longitudinal phase space for arbitrary accelerator parameters. However, expanding of the dataset to encompass a wider range of parameters significantly increases the quantity of data required to provide an accurate surrogate model. In continuing this research to expand this work to multiple screen locations along the beamline and to a wider parameter space, we have found that this increase in the generative range of the model has led to common Generative Adversarial Network pitfalls such as mode collapse and instability. We therefore conclude that this technique, while extremely powerful, is also quite difficult to implement reliably and care should be taken to avoid such issues.

Future work will include the simultaneous generation of LPS graphs for multiple points along the beamline, as well as comparisons to real-world data using a transverse deflecting cavity and potentially applying transfer learning using a small quantity of TDC data to augment a larger quantity of simulation data.

REFERENCES

- [1] E. A. Seddon *et al.*, “Short-wavelength free-electron laser sources and science: A review,” *Reports on Progress in Physics*, vol. 80, no. 11, p. 115 901, 2017, doi:10.1088/1361-6633/aa7cca
- [2] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.
- [3] I. Goodfellow *et al.*, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014.
- [4] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua, “CVAE-GAN: Fine-grained image generation through asymmetric training,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2745–2754.
- [5] A. L. Edelen, S. Biedron, S. V. Milton, and J. P. Edelen, “First steps toward incorporating image based diagnostics into particle accelerator control systems using convolutional neural networks,” in *Proc. North American Particle Accelerator*

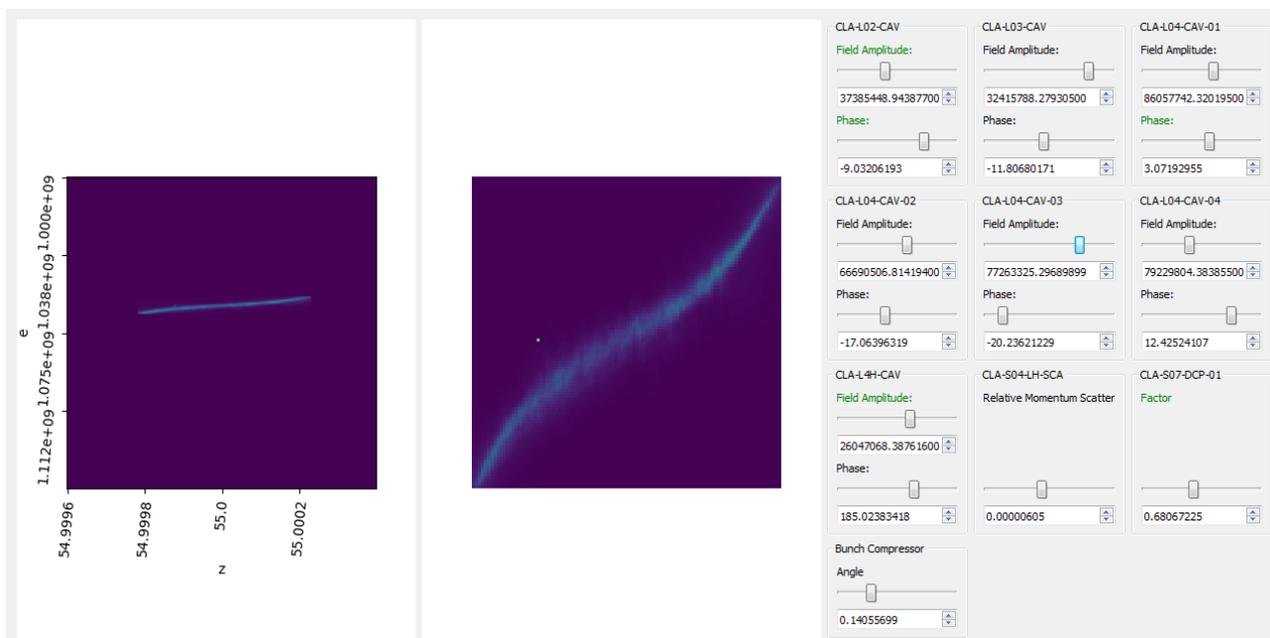


Figure 6: LPS CVAE-GAN GUI showing an example configuration, the fixed extent of the longitudinal phase space graph, and the region of interest which shows finer detail of the LPS distribution. Note the colouring of the labels above the sliders, which indicates the sensitivity of the LPS graph to that parameter in this region, with green indicating stability and red indicating significant impact from changes to this parameter.

- Conf. (NAPAC'16)*, Chicago, IL, USA, 2016, pp. 390–393, doi:10.18429/JACoW-NAPAC2016-TUPOA51
- [6] A. Scheinker, A. Edelen, D. Bohler, C. Emma, and 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.*, vol. 121, p. 044 801, 4 2018, doi:10.1103/PhysRevLett.121.044801
- [7] C. Emma, A. Edelen, M. Hogan, B. O’Shea, G. White, and V. Yakimenko, “Machine learning-based longitudinal phase space prediction of particle accelerators,” *Physical Review Accelerators and Beams*, vol. 21, no. 11, p. 112 802, 2018.
- [8] C. Emma *et al.*, “Machine learning-based longitudinal phase space prediction of two-bunch operation at FACET-II,” SLAC National Accelerator Lab., Menlo Park, CA, USA, Tech. Rep., 2019.
- [9] A. Edelen, N. Neveu, C. Emma, D. Ratner, and C. Mayes, “Machine learning models for optimization and control of x-ray free electron lasers,” in *Proc. NeurIPS Machine Learning for the Physical Sciences Workshop (NeurIPS2019)*, Vancouver, Canada, 2019.
- [10] A. Hanuka *et al.*, “Accurate and confident prediction of electron beam longitudinal properties using spectral virtual diagnostics,” *Scientific Reports*, vol. 11, no. 1, pp. 1–10, 2021.
- [11] J. Zhu, Y. Chen, F. Brinker, W. Decking, S. Tomin, and H. Schlarb, “High-fidelity prediction of megapixel longitudinal phase-space images of electron beams using encoder-decoder neural networks,” *Phys. Rev. Applied*, vol. 16, p. 024 005, 2 2021, doi:10.1103/PhysRevApplied.16.024005
- [12] M. Maheshwari, D. Dunning, J. Jones, M. King, H. Kockelbergh, and A. Pollard, “Prediction and Clustering of Longitudinal Phase Space Images and Machine Parameters Using Neural Networks and K-Means Algorithm,” in *Proc. IPAC'21*, Campinas, SP, Brazil, 2021, paper WEPAB318, pp. 3417–3420, doi:10.18429/JACoW-IPAC2021-WEPAB318
- [13] D. Dunning, H. C. Cortés, J. Jones, and N. Thompson, “Multi-Objective FEL Design Optimisation Using Genetic Algorithms,” in *Proc. FEL'19*, Hamburg, Germany, 2019, pp. 711–714, doi:10.18429/JACoW-FEL2019-THP065
- [14] D. Dunning, L. Cowie, and J. Jones, “XARA: X-Band Accelerator for Research and Applications,” in *Proc. FEL'19*, Hamburg, Germany, 2019, pp. 715–718, doi:10.18429/JACoW-FEL2019-THP066
- [15] D. Angal-Kalinin *et al.*, “Design, specifications, and first beam measurements of the compact linear accelerator for research and applications front end,” *Phys. Rev. Accel. Beams*, vol. 23, p. 044 801, 4 2020, doi:10.1103/PhysRevAccelBeams.23.044801
- [16] K. Floettmann, “Astra - a space charge tracking algorithm,” Tech. Rep., <http://www.desy.de/~mpyflo>
- [17] M. Borland, “Elegant: A flexible SDDS-compliant code for accelerator simulation,” Argonne National Lab., IL (US), Tech. Rep., 2000.
- [18] S. Kullback and R. A. Leibler, “On information and sufficiency,” *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.