OUT OF DISTRIBUTION ROBUSTNESS FOR PARTICLE ACCELERATOR APPLICATIONS

Aashwin Ananda Mishra, Auralee Edelen & Christopher Mayes
SLAC National Accelerator Laboratory
Menlo Park
California, CA 94025, USA
aashwin@slac.stanford.edu

ABSTRACT

Particle accelerators have found applications in diverse tasks such as cancer therapy, nuclear non-proliferation treaty verification, scientific discovery, etc. Such accelerators represent one of the most complex scientific devices, and are also extremely challenging to maintain and operate. Surrogate models for accelerators can aid in planning, design and operation, for instance in model based tuning. To this end, deep learning models are being used for surrogate modeling as they can process large corpora of high-dimensional data, ranging from scalars, to beam images and spectra. However, for the deployment of these models in such a high-regret and safety-critical system, verifiable out of distribution (OOD) robustness and reliable estimates of the predictive uncertainty are essential. We evaluate Bayesian Neural Networks as a means to provide OOD robustness, in conjunction with calibrated estimates of predictive uncertainty for particle accelerator systems. We outline application case studies across different accelerator designs (storage rings, beam lines for free electron lasers, injector systems). The cases have diverse data volumes and formats, e.g. particle beam phase space images and scalar parameters. We find that Bayesian Neural Networks provide accurate predictions with reliable uncertainty estimates, while ensuring robustness to dataset shifts.

1 INTRODUCTION & MOTIVATION

Particle accelerators are devices providing customized high energy beams by using electromagnetic fields to accelerate elementary particles, such as electrons or protons. Accelerators are used in diverse tasks, for instance medical therapy, where patients receive accelerator-based therapy each year for cancer (Amaldi, 2000). Accelerators play an important role in security, including cargo inspection, nuclear non-proliferation treaty verification, etc (Fazio et al., 2019). Accelerators are essential instruments to broad swaths of the scientific community; for example, accelerator based light sources are in high demand to provide scientific users with custom beams to image chemical, material, and biological samples. Such experiments have led to fundamental advances in scientific understanding of phenomena like photosynthesis (Young et al., 2016), electron-phonon interactions (Jiang et al., 2016; Singer et al., 2016), molecular interactions in drug delivery (Colletier et al., 2016), etc.

At present, the tuning, control and operation of particle accelerators is executed by trained operators, relying on intuition and experience to find appropriate accelerator settings to customize the beam for different experiments. However, this approach is time consuming and often leads to sub-optimal final settings. Machine learning based surrogate models are opportune for accelerator problems. The existence of large data sets and the need to predict complicated outputs, such as beam images, make neural networks (NNs) an appealing approach for ML-based modeling. However, to be used reliably in particle accelerator applications for prediction and control, uncertainty estimates are needed along with point predictions. Additionally, due to temporal drift in accelerator systems, model robustness to dataset shifts is critically important. Deep learning models often make overly confident predictions even for out of distribution inputs, and do not inherently include prediction uncertainties. For instance, deterministic NNs are unable to recognize out-of-distribution instances and make erroneous predictions for such cases with high confidence (Amodei et al., 2016; Nguyen et al., 2015). Such uncertainty in predictions has had grave consequences while applying deep learning to high-regret applications. For instance, the first fatality in automated driving systems occurred due to the inability of a trained AI-agent to differentiate the hue of a trailer from the color of the sky (Administration, 2017). Similarly, deep neural networks applied for facial recognition in law enforcement have exhibited critical errors (Dodds, 2018). Such epistemic uncertainty inherent to neural networks is exacerbated by the aleatoric and systemic uncertainties inherent to the modeling of particle accelerator applications. Interrelations between accelerator subsystems are complicated, involve large parameter spaces, and accelerator systems can be difficult to model *a priori*. Changes in system responses over time are also common due to drift, the existence of hidden variables, and transients. Online tuning of accelerators to meet new custom beam requests also often involves entering previously unexplored setting combinations. Beyond this, the instrumentation to characterize the beam response is often limited due to cost constraints, resulting in limited measurement data. In addition, the instrumentation and controllable components also have different levels of sensitivity and inherent noise for different beam parameters, leading to heteroscedastic effects. This uncertainty is aggravated by the presence of compounding errors in the many individual beamline components.

In this light, obtaining robustness to dataset shifts and quantified uncertainties for machine learning based models for particle accelerators is paramount if such models are to be applied in such high-regret and safety critical tasks. We evaluate Bayesian Neural Networks (BNNs) for particle accelerator applications with problems across different designs, with diverse data volumes and formats. We evaluate the ability of BNNs to provide accurate predictions of the mean and reliable estimates for the predictive uncertainty, along with OOD robustness. For inference, approximate Variational inference with the Bayes By Backprop algorithm (Blundell et al., 2015) is used. The network architectures are selected using Bayesian Optimization (O'Malley et al., 2019). To optimize the variational parameters, we utilize the Adam algorithm (Kingma & Ba, 2014) wherein the rates are set using cross-validation. The activations across all neurons are Rectified Linear Units (ReLU), and all weights and biases are initialized with standard normal priors.

2 Results & Discussion

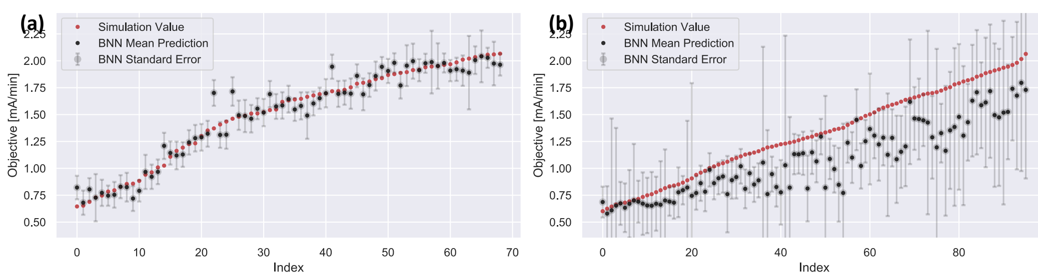

Figure 1: Comparison of Bayesian neural network predictions on (a) in-sample test dataset samples, (b) OOD dataset samples for the SPEAR3 dataset.

OOD robustness for Emittance Prediction in a Storage Ring: SPEAR3 is a 3-GeV, high-brightness electron storage ring Hettel (2005), operating with a beam current of 500 mA. The electron beam area in phase space (i.e. the *emittance*, ϵ) is an essential parameter that needs to minimized to produce a high-brightness photon beams for scientific users. In SPEAR3, 13 skew quadrupoles are adjusted to minimize the beam loss rate, which is a proxy measure for ϵ. In particle accelerator applications, dataset shifts are manifested often. For instance, the trained surrogate models can be used for exploration of the parameter space that is beyond the range of the training data. Additionally, there can be a change in the nature of the data, for instance using a surrogate model that is trained on simulation data and directly deployed to make predictions in real life applications. Herein, there is a difference between the inferred behavior of the model from simulation data and the actual machine system, as simulations rely on simplifications and often represent idealized beam dynamics. Another key concern is the temporal 'drift' in the machine system's response to given inputs. This can be manifested due to equipment aging, replacement of parts, environmental changes, etc. In such cases, the conditional distribution of the target changes due to drift. In all the aforementioned cases, deterministic surrogate models provide overly confident predictions that may be misleading. While different dataset shifts and perturbations can be introduced in image inputs, we engender an instance using the SPEAR3 simulation data. The simulation data is split into different clusters in feature space, and learning and OOD-test clusters are selected such that these have no overlap and also maximal difference between cluster centers. We train a Bayesian Neural Network on the training dataset. The performance is evaluated on the test dataset that is in-sample, and the OOD dataset. As can be seen in Figure 1, the performance of the mean predictions of the BNN are substantially depreciated when the predictions are made on the OOD samples, as compared to the in-distribution test dataset. However, the uncertainty outputted by the BNN increases significantly on the OOD dataset contra the in-sample test dataset. The Mean Prediction Interval Width on the in-sample test dataset is 0.16 and on the OOD dataset is 0.35. Furthermore, as the OOD dataset points move away from the learning dataset cluster center, the predictive uncertainty trends higher. In such a scenario, BNNs show OOD robustness, where cut-offs can be set on their prediction intervals to denote samples where the surrogate model may be relied upon and samples for which the surrogate model predictions may be unreliable.

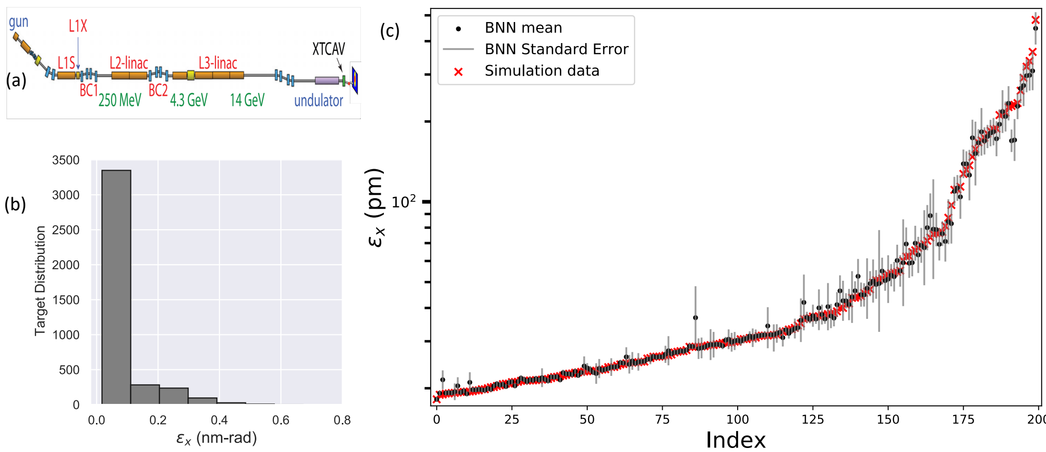

Figure 2: BNN predictions and standard error estimates for (a) In-sample test data, (b) OOD test data, for emittance prediction for the SPEAR3 storage ring.

Emittance Prediction in the LCLS Linac: In our second case, we examine modeling the transverse emittance (ϵ_x) of the Linac Coherent Light Source (LCLS) electron beamline. The LCLS is a free electron laser (FEL) based light source user facility providing customized photon beams for scientific experiments. The FEL process is extremely sensitive to variations in the electron beam phase space, which in turn is sensitive to a variety of accelerator settings and the impact of collective effects such as coherent synchrotron radiation. For this case, 4k data points are obtained using Bmad Sagan (2008) simulations, which includes nonlinear collective beam effects. The data is uniformly sampled from a large operating range of the accelerating cavity phases and voltages (6 features total), which are commonly adjusted to optimize the beam's shape. However, uniform sampling in feature space does not translate to uniform sampling in target space, as is outlined in Figure 2 (b). Due to the paucity of samples at high values of emittance, data driven model predictions have significant discrepancy in these ranges, and reliable uncertainty estimates are essential. The validation and testing sets consist of 1k and 800 samples, respectively. The BNN architecture has 8 hidden layers. The input layer has 6 features and the output ϵ_x. It can be observed in Figure 2 (c), for low values of the emittance the BNN mean predictions are accurate and the uncertainty bounds are largely negligible. In the few cases where the BNN prediction is erroneous, the predicted standard error reflects this. At high values of the emittance, where data is sparse, the BNN mean predictions have discrepancy comparable to a deterministic neural network. However, the predicted standard error is accordingly higher and reflects the discrepancy in predictions. The BNN provides prediction accuracy comparable to a deterministic neural network (MAE=0.02 for both), but the uncertainty bounds make this a reliable model for such applications.

Prediction of Phase Space Image Projections in the LCLS-II Injector: In addition to scalar quantities, prediction of beam phase space projections are often used to provide additional information about the beam. The beam itself is a collection of six-dimensional information (3 positions and

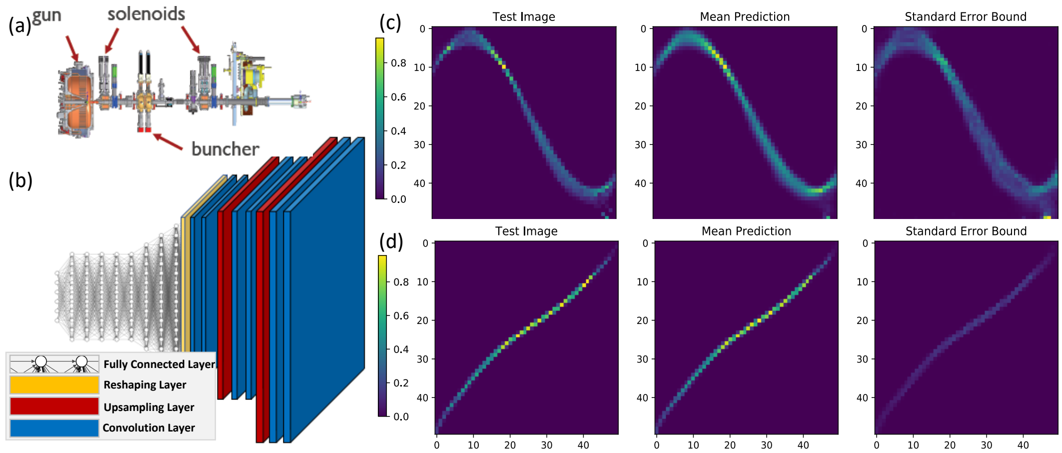

Figure 3: (a) LCLS-II Injector Schematic, (b) BNN architecture, and BNN mean predictions and predicted standard error for the LCLS-II injector longitudinal phase space projections, which show the time vs. energy histogram of the bunch (here shown as 50×50 binned images). Here we are showing two randomly-selected representative examples (c) and (d)

3 momenta for each particle), and 2D projections of the phase space can also be directly measured. Thus, for NN models of accelerator systems, it is highly desirable to predict both these 2D projections and the associated uncertainty. In addition to being useful for general accelerator modeling, ML-based image prediction is useful for providing non-invasive estimates of the beam phase space in cases where it cannot be continuously measured. Here, we focus on prediction of the longitudinal phase space images of the LCLS-II injector. To generate the dataset, simulations of the injector using the IMPACT code Ryne & Habib (1997) were carried out. The scalar inputs were generated by randomly sampling 5 settings of interest (the injector cavity phase, two solenoid strengths, and a buncher cavity amplitude and phase). The network architecture consisted of encoder and decoder sections, outlined in Figure 3 (b). The encoder section consists of 9 densely connected layers consisting of 20 ($\times$ 5), 100, 200, 600 and 10k neurons respectively. The vector output of the encoder section is reshaped into a higher-order tensor before being fed into the decoder section. The decoder section consists of sets of convolutional layers, followed by upsampling layers. Here, the upsampling factor for the rows and columns was 2. The convolutional layers had 16 filters, except for the last layer having 1 filter. The kernel sizes over the six convolutional layers were $4, 4, 4, 3, 2, 1$ respectively. The training dataset consisted of 15k pairs of scalar inputs and image outputs. The validation and testing dataset consisted of 2k pairs of scalar inputs and image outputs each. Representative predictions on the test set are shown in Figure 3. For each instance, we report the test image from the simulation, the mean prediction from the BNN and the standard error predicted by the BNN. The standard error highlights regions where there is significant discrepancy between the mean prediction and the simulation output, for instance in Figure 3 (c) . In cases where the mean prediction is in close agreement with the simulation, for instance in Figure 3 (d), the standard error is correspondingly lower.

3 Summary & Future Outlook

In this investigation, we show that BNNs can provide accurate predictions and uncertainty estimates for several different kinds of accelerator systems. Additionally, BNNs provide the OOD robustness that is critical for particle accelerator applications. We include predictions of scalar data image data that describe relevant aspects of the particle beam with respect to changing accelerator settings. Evaluating methods for incorporating uncertainty estimates into neural network-based models of particle accelerator systems is an essential step in ensuring that they can be reliably put to use in accelerator operations. Examples of use cases includes experiment planning, model-based control, and online prediction of beam parameters that cannot be continuously measured (i.e. virtual diagnostics) for use in control and data analysis.

This work was supported under the U.S. Department of Energy/ Stanford University Contract for Management and Operation of SLAC National Accelerator Laboratory under Contract No. `DE-AC02-76SF00515`. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. `DE-AC02-05CH11231`.

REFERENCES

National Highway Traffic Safety Administration. Tesla crash preliminary evaluation report ODI resume. investigation: PE 16-007, 2017.

Ugo Amaldi. The importance of particle accelerators. *Europhysics News*, 31(6):5–9, 2000.

Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*, 2016.

Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural networks. In *International Conference on Machine Learning*, pp. 1613–1622, 2015.

Jacques-Philippe Colletier, Michael R. Sawaya, Mari Gingery, Jose A. Rodriguez, Duilio Cascio, Aaron S. Brewster, Tara Michels-Clark, Robert H. Hice, Nicolas Coquelle, Sébastien Boutet, Garth J. Williams, Marc Messerschmidt, Daniel P. DePonte, Raymond G. Sierra, Hartawan Laksmono, Jason E. Koglin, Mark S. Hunter, Hyun-Woo Park, Monarin Uervirojnangkoorn, Dennis K. Bideshi, Axel T. Brunger, Brian A. Federici, Nicholas K. Sauter, and David S. Eisenberg. De novo phasing with x-ray laser reveals mosquito larvicide binab structure, 2016. URL `https://doi.org/10.1038/nature19825`.

Laurence Dodds. Chinese businesswoman accused of jaywalking after ai camera spots her face on an advert. *The Telegraph*, 2018.

Michael Fazio, George Laramore, Suresh Pillai, Ahmed Badruzzaman, Harry Martz, Jeff Buchsbaum, David Jaffray, Mary-Keara Boss, Andrea Schmidt, and Jeff Calame. Basic research needs workshop on compact accelerators for security and medicine: Tools for the 21st century, may 6-8, 2019. 2019.

R Hettel. The completion of SPEAR 3. Technical report, Stanford Linear Accelerator Center (SLAC), Menlo Park, CA, 2005.

M. P. Jiang, M. Trigo, I. Savić, S. Fahy, É. D. Murray, C. Bray, J. Clark, T. Henighan, M. Kozina, M. Chollet, J. M. Glownia, M. C. Hoffmann, D. Zhu, O. Delaire, A. F. May, B. C. Sales, A. M. Lindenberg, P. Zalden, T. Sato, R. Merlin, and D. A. Reis. The origin of incipient ferroelectricity in lead telluride, 2016. URL `https://doi.org/10.1038/ncomms12291`.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. URL `http://arxiv.org/abs/1412.6980`. cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 427–436, 2015.

Tom O'Malley, Elie Bursztein, James Long, François Chollet, Haifeng Jin, Luca Invernizzi, et al. Keras Tuner. `https://github.com/keras-team/keras-tuner`, 2019.

Robert Ryne and Salman Habib. Impact: An integrated-map & particle accelerator tracking code for linac simulation. *APS*, pp. 9P–105, 1997.

David Sagan. The bmad reference manual. *Cornell University*, 2008. URL `https://www.classe.cornell.edu/bmad/`.

A. Singer, S. K. K. Patel, R. Kukreja, V. Uhlíř, J. Wingert, S. Festersen, D. Zhu, J. M. Glownia, H. T. Lemke, S. Nelson, M. Kozina, K. Rossnagel, M. Bauer, B. M. Murphy, O. M. Magnussen, E. E. Fullerton, and O. G. Shpyrko. Photoinduced enhancement of the charge density wave amplitude. *Phys. Rev. Lett.*, 117:056401, Jul 2016. doi: 10.1103/PhysRevLett.117.056401. URL https://link.aps.org/doi/10.1103/PhysRevLett.117.056401.

Iris D. Young, Mohamed Ibrahim, Ruchira Chatterjee, Sheraz Gul, Franklin D. Fuller, Sergey Koroidov, Aaron S. Brewster, Rosalie Tran, Roberto Alonso-Mori, Thomas Kroll, Tara Michels-Clark, Hartawan Laksmono, Raymond G. Sierra, Claudiu A. Stan, Rana Hussein, Miao Zhang, Lacey Douthit, Markus Kubin, Casper de Lichtenberg, Long Vo Pham, Håkan Nilsson, Mun Hon Cheah, Dmitriy Shevela, Claudio Saracini, Mackenzie A. Bean, Ina Seuffert, Dimosthenis Sokaras, Tsu-Chien Weng, Ernest Pastor, Clemens Weninger, Thomas Fransson, Louise Lassalle, Philipp Bräuer, Pierre Aller, Peter T. Docker, Babak Andi, Allen M. Orville, James M. Glownia, Silke Nelson, Marcin Sikorski, Diling Zhu, Mark S. Hunter, Thomas J. Lane, Andy Aquila, Jason E. Koglin, Joseph Robinson, Mengning Liang, Sébastien Boutet, Artem Y. Lyubimov, Monarin Uervirojnangkoorn, Nigel W. Moriarty, Dorothee Liebschner, Pavel V. Afonine, David G. Waterman, Gwyndaf Evans, Philippe Wernet, Holger Dobbek, William I. Weis, Axel T. Brunger, Petrus H. Zwart, Paul D. Adams, Athina Zouni, Johannes Messinger, Uwe Bergmann, Nicholas K. Sauter, Jan Kern, Vittal K. Yachandra, and Junko Yano. Structure of photosystem ii and substrate binding at room temperature, 2016. URL https://doi.org/10.1038/nature20161.