ISBN: 978-3-95450-219-6 ISSN: 2226-0366 doi: 10.18429/JACoW-LINAC2024-MOPB081

GENERALIZATION ABILITY OF CONVOLUTIONAL NEURAL NETWORKS TRAINED FOR COHERENT SYNCHROTRON RADIATION COMPUTATIONS *

C. Leon†, P.M. Anisimov, N. Yampolsky, A. Scheinker
Applied Electrodynamics Group, Los Alamos National Laboratory, Los Alamos, NM, USA

Abstract

Coherent synchrotron radiation (CSR) has a significant impact on electron storage rings and bunch compressors, inducing energy spread and emittance growth in a bunch. Calculating the effects of CSR is computationally expensive, severally limiting the use of simulations. Here, we explore utilizing neural networks (NNs) to model the 3D wakefields of electrons in circular orbit in the steady state condition. NN models were trained on both Gaussian and more general bunch distributions, which evaluate much faster than physics-based simulations. Here, we explore how well the models generalize, by testing their ability to: 1) extrapolate to Gaussians with smaller/larger widths 2) predict on distributions never encountered before (out of distribution generalization) using smoothed uniform cubes. We see the models are able to generalize, which makes them potentially useful in the design and optimization of accelerator apparatuses by enabling rapid searches through parameter space.

INTRODUCTION

Coherent Synchrotron Radiation (CSR) can significantly alter the distribution of electrons moving in circular orbit, such as in an electron storage ring or a bunch compressor. For emitted wavelengths, λ, much larger than the bunch size, then at λ-resolution the bunch looks point-like, and the N_e electrons look like they're all undergoing nearly identical motion. Consequentially, the radiation emitted by these electrons are approximately in phase and adds up coherently, resulting in an intensity of $\mathcal{O}(N_e^2)$. This can significantly distort the bunch phase space distribution. In the ultra-relativistic limit, this primarily causes the tail and center of the bunch to lose energy while the head gains, leading to an increase in emittance.

The electromagnetic (EM) field produced by a charged point particle in motion is given by the well-known Liénard-Wiechart (LW) fields [1–3]:

$$\mathrm{E} = \frac{q}{4\pi\epsilon_0}\left(\frac{\mathrm{n}-\beta}{\gamma^2(1-\mathrm{n}\cdot\beta)^3\rho^2} + \frac{\mathrm{n}\times(\mathrm{n}-\beta)\times\dot{\beta})}{c(1-\mathrm{n}\cdot\beta)^3\rho}\right)\Bigg|_{ret.}, \tag{1}$$

$$\mathrm{B} = \frac{1}{c}\mathrm{n}\times\mathrm{E}, \tag{2}$$

where the *ret*. signifies that the expression must be evaluated at the retarded time to ensure causality.

The total force an electron experiences due to the radiation emitted by other electrons in the past is called the wakefield, $\mathrm{W}(\mathrm{r},t)$. To find it, one must integrate the Lorentz force, $\mathrm{F}(\mathrm{r},t)$, generated by the LW fields in Eqs. (1) and (2) along the past light cone:

$$\mathrm{W}(\mathrm{r}) = \int_{\triangle} d^3\mathrm{r}'\,\lambda(\mathrm{r}',t')\mathrm{F}(\mathrm{r}',t')\Big|_{ret.}, \tag{3}$$

where $\triangle$ signifies the past light cone, time in the integrand is the retarded time and $\lambda(\mathrm{r}',t')$ is the number density. The computational storage needed to do the evaluations at the retarded times and the computational expense of $\mathcal{O}(N_e^2)$ interactions, with typical bunch sizes of $N_e \sim 10^{10}$, for each time step makes designing practical software for CSR calculations a formidable task. To deal with the complexities, many approximations are commonly used, such as the 1D approximation [4]. Even in the specific case of circular motion in the 1D approximation, though, the incorporation of CSR effects in a simulation can increase the running time of the simulation by an order of magnitude [5]. This makes many conventional simulations far too slow for some applications, such as a thorough exploration of parameter space to optimize accelerator component design. With the production of smaller bunch sizes in accelerator facilities, the need for a fast and accurate account of 3D CSR has become important.

Machine learning (ML) can speed up simulations to provide real-time virtual diagnostics which can be used for real-time adaptive beam control. For example, in [6] the first approach to adaptive ML was demonstrated combining a deep learning with adaptive feedback [7] for automatic control of the longitudinal phase space of the LCLS FEL electron beam. They can also be used to speed up accelerator simulations [8, 9]. Convolutional neural network (CNN) computations utilize matrix multiplications and several parallelized operations and with modern GPU's, which are optimized for such tasks, can be performed rapidly. Hence, a CNN trained on CSR simulations data can lead to very fast computations.

Previous work has investigated the use of ML surrogate models to speed up CSR calculations. Mayes and Edelen used a dense neural network in the 1D case to speed up CSR calculations [10]. CNN's in particular have been utilized in the 2D transient case [11].

Here, we take previous CNN models trained to predict the fully 3D wakefields generated by electrons at steady-state [12] and see their generalization ability. Specifically, we investigate at how well it performs at extrapolating to

* Work was supported by the Los Alamos National Laboratory Laboratory Directed Research and Development (LDRD) DR project 20220074DR.
† cleon@lanl.gov

ISBN: 978-3-95450-219-6 ISSN: 2226-0366 doi: 10.18429/JACoW-LINAC2024-MOPB081

Gaussians widths beyond their training data. In addition, we test how well it performs on out of distribution (OOD) generalization, using smoothed uniform cubes as the electron distribution, of which the model has never encountered before.

DATA

The data used here comes from the conventional (i.e, non-ML) software `PyCSR3D`, whose physics formulation is based on the Hamiltonian approach in Synder-Courant theory developed by Cai and Ding [13–15]. It is assumed the bunch is in a steady state condition (i.e, not transient) and that the shielding suppression of lower frequencies from the metallic cavity is negligible.

The Frenet-Serret coordinates with the reference path of circular motion in the horizontal plane is used. Here, the longitudinal component s is the arc length, the horizontal x is the radial outward distance from the circle, while y is the vertical distance from the horizontal plane.

For training, two datasets of distributions and wakefields were constructed, the first being referred to as the Gaussian Dataset, which consisted of 500 Gaussians distribution and their wakes. The second one, the General Dataset, was designed to be more varied with each electron distribution being a normalized sum of 2-25 Gaussians, with means near the center and the standard deviations with the same range as the Gaussian Dataset (see Table 1). Approximating a distribution as a sum of Gaussians is used in kernel density estimation. This General Dataset contained 500 instances. For both datasets, an 85/15 % train/test split was done. All field configurations were represented as 3D arrays of dimensions $128 \times 128 \times 256$, with the final index being the s component. It was given twice the resolution as the transverse components as it is usually the most important component of the three. For all sets, the bending radius is $\rho = 1\,\mathrm{m}$ and $\gamma = 500$.

Two more sets were used to evaluate generalization ability of the models. The first was a dataset of Gaussians, with fixed $\sigma_x = \sigma_y = 6\,\mu\mathrm{m}$ and going from $\sigma_z = 1\,\mu\mathrm{m}$ to $18\,\mu\mathrm{m}$, with 100 examples. This will be used to test how well the models extrapolate.

The second set tests the OOD generalization. Let's consider the probability density function (PDF) for a 1D uniform distribution between two points, a and b. It can be expressed as the difference of two step functions, each centered at a and b. This can be smoothed by replacing the step functions with sigmoids: $f(x) = \frac{1}{b-a}\left(\frac{1}{1+e^{x-a}} - \frac{1}{1+e^{x-b}}\right)$. We call a 'smoothed uniform cubical distribution' one that uses has independent PDFs of $f(x)$ for all 3 components of $i = x, y, s$, and all have the same widths. (Note: the use of curvilinear coordinates means even without smoothing it's only approximately a cube in physical space). The smoothing was done to improve the numerics of the simulation. In addition, a smooth drop-off in density is a more realistic description of actual particle beams. A dataset of 100 examples with side lengths of $l = 4\,\mu\mathrm{m}$ to $l = 50\,\mu\mathrm{m}$.

Table 1: Trainin/Test Dataset Characteristics

	Gaussian Dataset	**General Dataset**
Physical Range	$x_i \in (-48, 48)$ μm $i = s, x, y$	Same as Gaussian Dataset
Distribution Attributes	$\sigma_i \in (2, 12)$ μm $i = s, x, y$ Centered at $\mathbf{0}$	Sum of 2 to 25 Gaussians $\mu_i \in (-12, 12)$ μm $\sigma_i \in (2, 12)$ μm

MODELS

The models considered here will be labelled λ-NN, which:

$$\lambda\text{-NN}: \quad \lambda(\mathrm{r}) \to \mathrm{W}(\mathrm{r}) \tag{4}$$

In terms of arrays, $(128, 128, 256, 1) \to (128, 128, 256, 3)$. This followed a U-Net architecture [16].

We first trained λ-NN on the Gaussian Dataset, which we'll refer to as λ-NN-Gauss. We then looked to see how well what it learned transferred to the General Dataset. It was then taken as a pre-trained model and then trained on the General Dataset, with the final resulting model labelled λ-NN-Gen.

All the simulations, ML training and evaluations were performed using a workstation with an Intel Xenon Plantinum 8268 CPU with 24 cores and an NVIDIA RTX A6000 with 48 GB RAM for its GPU.

RESULTS

To evaluate the models, we used the relative weighted mean average error (RW-MAE) for each component $i = s, x, y$:

$$\text{RW-MAE}_i = \frac{\int d^3\mathrm{r}\,\lambda(\mathrm{r})|W^i_{pred.}(\mathrm{r}) - W^i_{true}(\mathrm{r})|}{\int d^3\mathrm{r}\,\lambda(\mathrm{r})|W^i_{true}(\mathrm{r})|}, \tag{5}$$

where $W^i_{true}(\mathrm{r})/W^i_{pred.}(\mathrm{r})$ is the true/predicted values for the wakefield. The idea behind this is that we care most about being correct where there are more particles, since $\mathrm{W}(\mathrm{r})$ is being used to update the phase space. Moreover, this metric differs quite a bit from the mean squared error loss (MSE) used in training, thus providing a complementary measure to the optimized MSE.

Gaussian Extrapolation

As can be seen in Fig. 1, the λ-NN-Gen model was good at extrapolating to larger σ_s with the RW-MAE error for each component rising only slightly rising outside training/test data. In contrast, the predictions for smaller σ_s did much worse and rose rapidly as it was decreased. The lattice length is 0.75 μm (0.375 μm) for the transverse (longitudinal) direction, thus it shouldn't be surprising that when σ_s approaches this value the model ceases to be accurate.

ISBN: 978-3-95450-219-6 ISSN: 2226-0366 doi: 10.18429/JACoW-LINAC2024-MOPB081

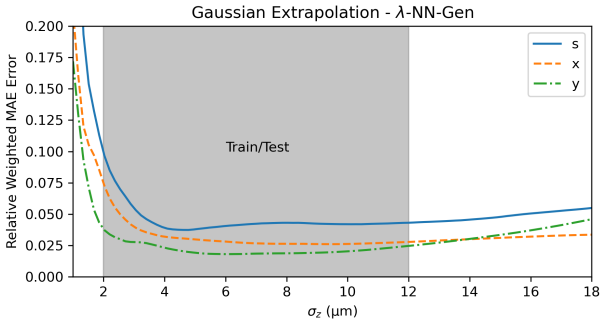

Figure 1: Gaussian Extrapolation - the performance of λ-NN-Gen with varying σ_z.

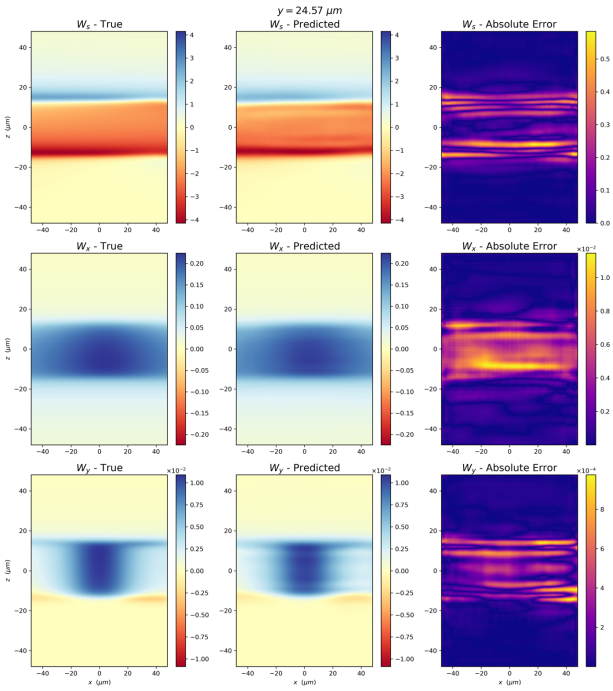

Figure 2: Smooth Uniform Cube - λ-NN-Gen predictions for wakefield at $y = 24.57\,\mu$m.

Smoothed Uniform Cubical Distribution

Fits can be seen in Figs. 2 and 3, while the RW-MAE for each component can be seen in Fig. 4. The results are similar to the Gaussian case, where the models fail at small widths, but become almost level at larger ones, rising only slowly. Based on this and the previous results, it seems the λ-NN-Gen should only be used on distributions with widths larger than $\sim 10\,\mu$m.

CONCLUSION

We have developed surrogate ML models that can accurately predict 3D wakefields of electrons in the steady state condition. We investigated the generalization ability of the models, specifically with the extrapolation ability on Gaussian widths smaller/larger than encountered before and

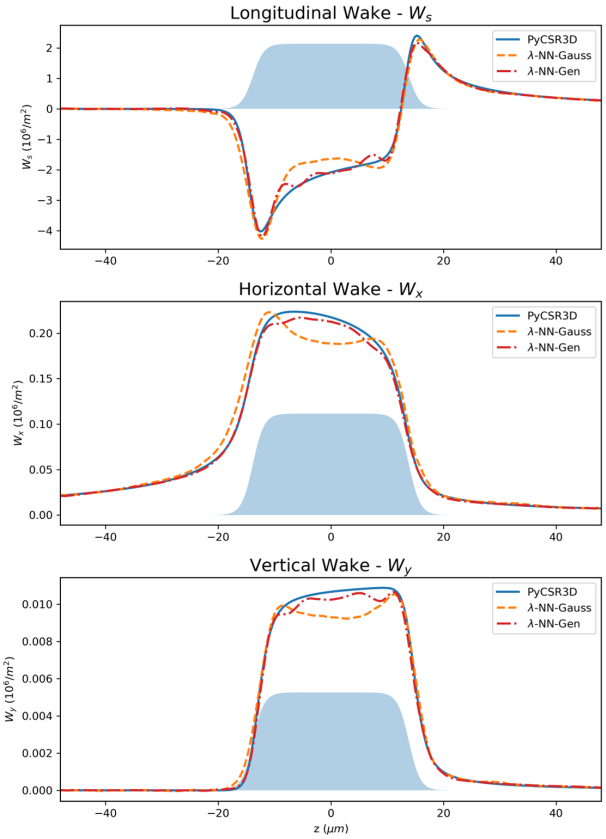

Figure 3: Smooth Uniform Cube - The 1D plots along the $x = 0\,\mu$m, $y = 24.57\,\mu$m line.

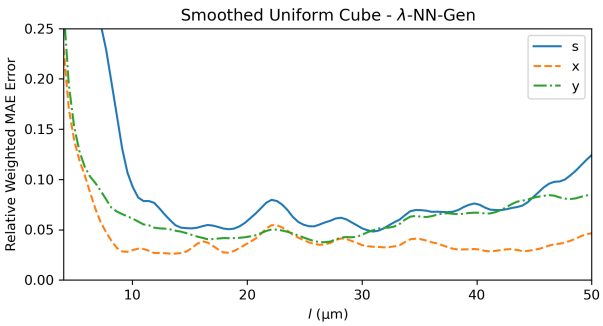

Figure 4: Smooth Uniform Cube - RW-MAE error as a function of side lengh, l.

the OOD distribution of smoothed uniform cubical distribution. These models offer the possibility of accelerator design optimization through wide parameter searches.

Future work can incorporate physics information into the networks, as previous work has shown this can lead to improved predictions and better generalization [17, 18]. Increasing the variety of distributions in the training set can increase the robustness of the λ-NN model. Finally, the energy of the beam here has been kept fixed, but in future work it can be included as an input to the networks.

ISBN: 978-3-95450-219-6 ISSN: 2226-0366 doi: 10.18429/JACoW-LINAC2024-MOPB081

REFERENCES

[1] A. Liénard, *Champ électrique et magnétique produit par une charge électrique concentrée en un point et animée d'un mouvement quelconque*. G. Carré et C. Naud, 1898.

[2] E. Wiechert, "Elektrodynamische elementargesetze", *Annalen der Physik*, vol. 309, no. 4, pp. 667–689, 1901.

[3] J. D. Jackson, *Classical electrodynamics*, 1999.

[4] M. Borland, "Simple method for particle tracking with coherent synchrotron radiation", *Physical Review Special Topics-Accelerators and Beams*, vol. 4, no. 7, p. 070 701, 2001.

[5] C. Mayes, "Computational approaches to coherent synchrotron radiation in two and three dimensions", *Journal of Instrumentation*, vol. 16, no. 10, p. P10010, 2021.

[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", *Physical review letters*, vol. 121, no. 4, p. 044 801, 2018.

[7] A. Scheinker and M. Krstić, "Minimum-seeking for clfs: Universal semiglobally stabilizing feedback under unknown control directions", *IEEE Transactions on Automatic Control*, vol. 58, no. 5, pp. 1107–1122, 2012.

[8] M. Rautela, A. Williams, and A. Scheinker, "A conditional latent autoregressive recurrent model for generation and forecasting of beam dynamics in particle accelerators", *Scientific Reports*, vol. 14, no. 1, p. 18 157, 2024.

[9] C. Leon and A. Scheinker, "Physics-constrained machine learning for electrodynamics without gauge ambiguity based on fourier transformed maxwell's equations", *Scientific Reports*, vol. 14, no. 1, p. 14 809, 2024.

[10] A. Edelen and C. Mayes, "Neural network solver for coherent synchrotron radiation wakefield calculations in accelerator-based charged particle beams", *arXiv preprint arXiv:2203.07542*, 2022.

[11] R. Robles, W. Lou, C. Mayes, R. Roussel, and A. Edelen, "Efficient computation of two-dimensional coherent synchrotron radiation with neural networks", in *Proc. 14th International Particle Accelerator Conference*, Venice, Italy, pp. 2892–2895, 2023. doi:10.18429/JACoW-IPAC2023-WEPA105

[12] C. Leon, P. M. Anisimov, N. Yampolsky, and A. Scheinker, "Utilizing neural networks to speed up coherent synchrotron radiation computations", in *Proc. 15th International Particle Accelerator Conference*, Nashville, TN, pp. 901–904, 2024. doi:10.18429/JACoW-IPAC2024-MOPS73

[13] Y. Cai, "Coherent synchrotron radiation by electrons moving on circular orbits", *Physical Review Accelerators and Beams*, vol. 20, no. 6, p. 064 402, 2017.

[14] Y. Cai and Y. Ding, "Three-dimensional effects of coherent synchrotron radiation by electrons in a bunch compressor", *Physical Review Accelerators and Beams*, vol. 23, no. 1, p. 014 402, 2020.

[15] Y. Cai, "Two-dimensional theory of coherent synchrotron radiation with transient effects", *Physical Review Accelerators and Beams*, vol. 24, no. 6, p. 064 402, 2021.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, 2015. doi:10.1007/978-3-319-24574-4_28

[17] A. Scheinker and R. Pokharel, "Physics-constrained 3D convolutional neural networks for electrodynamics", *APL Mach. Learn.*, vol. 1, no. 2, p. 026 109, 2023. doi:10.1063/5.0132433

[18] A. Bormanis, C. A. Leon, and A. Scheinker, "Solving the Orszag–Tang vortex magnetohydrodynamics problem with physics-constrained convolutional neural networks", *Physics of Plasmas*, vol. 31, no. 1, p. 012 101, 2024. doi:10.1063/5.0172075