

Transfer learning as a method to reproduce high-fidelity non-local thermodynamic equilibrium opacities in simulations

Michael D. Vander Wal ^{1,†}, Ryan G. McClarren ¹ and Kelli D. Humbird ²

¹University of Notre Dame, Notre Dame, IN 46556, USA

²Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

(Received 28 May 2022; revised 18 November 2022; accepted 25 November 2022)

Simulations of high-energy density physics often need non-local thermodynamic equilibrium opacity data. These data, however, are expensive to produce at relatively low fidelity. It is even more so at high fidelity such that the opacity calculations can contribute 95 % of the total computation time. This proportion can even reach large proportions. Neural networks can be used to replace the standard calculations of low-fidelity data, and the neural networks can be trained to reproduce artificial, high-fidelity opacity spectra. In this work, it is demonstrated that a novel neural network architecture trained to reproduce high-fidelity krypton spectra through transfer learning can be used in simulations. Further, it is demonstrated that this can be done while achieving a relative per cent error of the peak radiative temperature of the hohlraum of approximately 1 % to 4 % while achieving a $19.4\times$ speed up.

Key words: plasma properties, fusion plasma

1. Introduction

Inertial confinement fusion (ICF) is one of the prospective methods for achieving commercial fusion power generation. While ICF research is on the cusp of achieving breakeven net energy output for the amount of energy provided by the laser drive, there is much work to still be done (Callahan 2021). One challenge associated with ICF energy production is that the experiments, or ‘shots’, are expensive and limited in number. Simulations thus play an important role in selecting experimental parameters to maximize information gained from each shot.

The simulations of ICF experiments have their own high costs, often taking several days on several nodes of the most powerful supercomputers in the world. For many integrated hohlraum simulations, the calculation of just the spectral absorptivity and emissivity, or opacity, can easily consume as much as 90 % of the total computation time (Kluth *et al.* 2020). This is, of course, dependent on the fidelity of the spectra being produced where the high-fidelity spectra compose 96 % of the entire computation time in this work. The

[†]Email address for correspondence: mvander5@nd.edu

higher the level of fidelity of the spectra being computed, the greater the proportion of the simulation time that is dedicated to the calculating the opacities.

While the higher cost of increased fidelity computations is to be expected, it does mean that there could be physics missing when running models at low fidelity. There can be significant changes in the spectra from low fidelity to high fidelity such that a band of frequencies, absorbed or emitted, may change in width, height or depth and position. This is made rather apparent in previous work on the topic (Vander Wal, McClarren & Humbird 2022*b*). A simple example of an effect changes in model fidelity could have is the change in penetration depth, or the likelihood of escape or absorption of a photon. If that particular energy photon carries the majority of the energy emitted and the high-fidelity calculation results in an increase in that energy band, it is likely that less energy leakage will occur.

Neural networks have thus been suggested as an alternative to the normal usage of an atomic physics code such as Cretin. Previous work has demonstrated that, for a one-dimensional simulation with a krypton hohlraum, using neural networks to perform the non-local thermodynamic equilibrium (NLTE) opacity calculations for krypton can net a 10x speedup in computation time with sub-one per cent error for the radiation temperature during simulations (Kluth *et al.* 2020). Furthermore, neural networks trained on high fidelity atomic physics data could enable not only faster, but also more accurate calculations to be used in integrated ICF simulations. In this work, a technique called ‘transfer learning’ is used to create high-fidelity neural networks to predict NLTE opacities for use in integrated ICF simulations. Our results indicate the adoption of transfer learning allows high-fidelity NLTE physics to be included in routine ICF simulations at a computational cost comparable to low-fidelity calculations.

1.1. *Prior work and contributions*

This work is an extension of previous work to replace low-fidelity krypton spectra in simulations run in the radiation-hydrodynamics code Hydra (Kluth *et al.* 2020). There has also been work on transfer learning to make neural networks that reproduce high-fidelity spectra from stand-in radiative fields and a large range of input values that include realistic and non-realistic combinations of inputs (Vander Wal *et al.* 2022*b*). Other work related to the reproduction of spectra was the demonstration of a multi-element model and how the choice of data scaling is important and necessary to achieve accurate results (Vander Wal, McClarren & Humbird 2022*a*).

Elsewhere in the fields of nuclear, atomic and plasma physics have been a number of applications of machine learning. One of the more recent and substantial uses of machine learning in these fields has been the demonstration, at least by way of simulation, of the ability of neural networks to control magnetically confined plasmas in tokamak reactors (Degraeve 2022). Neural networks are heavily used to improve the understanding of simulation discrepancies with experimental data and to create more accurate models to guide experimental design (Kustowski *et al.* 2019; Humbird & Peterson 2020; Humbird *et al.* 2021). In plasma physics, physics-informed neural networks have been used to evaluate plasma conditions around space craft as well as more general plasma behaviour in space (Bard & Dorelli 2021; Zhang *et al.* 2021).

Outside of the field of NLTE opacities or anything related to nuclear, atomic or plasma physics there has been other work for the reproduction of spectra or the prediction of spectra for molecular compounds (Ghosh *et al.* 2019). A more common application has been in the use of making calibration spectra for optical measurements (Cui & Fearn 2018; Chatzidakis & Botton 2019; Chen & Wang 2019; Lan, McClarren & Vishwanath 2021). Not focusing on spectra prediction itself but rather the use of spectra, there has been work

on the utilization of spectra to predict the properties of oil spills, the composition of plants and properties of biological tissue (Liu *et al.* 2019; Ni, Wang & Tao 2019; Lan *et al.* 2021).

Finally, there is a plethora of papers addressing transfer learning. Transfer learning at the most general level is the process of taking a neural network trained on one set of data and then retraining it on similar, yet different data, of different levels of fidelity (Aydin, Braeu & Cyron 2019; Vander Wal *et al.* 2022b). The idea behind transfer learning is that neural networks possess prior knowledge about a similar task, so the network can likely learn the new information with fewer training samples (Deng *et al.* 2013, 2014; Zhuang *et al.* 2015; Eusebio 2018; Luo *et al.* 2018; Aydin *et al.* 2019; Zhu *et al.* 2019; Imai, Kawai & Nobuhara 2020; Meng & Karniadakis 2020; Zhuang *et al.* 2021). Similarly, one can take part of an already trained network and attach a new network, or set of layers, to the pre-trained partial network. This particular process is quite common for networks working with images where a convolutional autoencoder is trained and then only convolutional filters in the encoder are used. The final product of these networks are used for classification (Deng *et al.* 2013, 2014; Zhuang *et al.* 2015; Eusebio 2018; Luo *et al.* 2018; Zhu *et al.* 2019). Transfer learning is used for regression as well, and indeed, incorporating experimental data into an already trained network is a form of transfer learning (Aydin *et al.* 2019; Kustowski *et al.* 2019; Goswami *et al.* 2020; Humbird & Peterson 2020; Meng & Karniadakis 2020; Chakraborty 2021; Humbird *et al.* 2021). Some of the methods used can become rather elaborate such as adding additional networks for linear and nonlinear correlations for multi-fidelity transfer learning and requiring both levels of fidelity be utilized concurrently during training (Meng & Karniadakis 2020).

The work provided here is an immediate follow up on the demonstration that high-fidelity spectra can be transfer learned in an effective fashion. In this work, we demonstrate the viability of using transfer learned models inside integrated ICF hohlraum simulations by reproducing the maximum radiation temperature observed within the hohlraum with mean relative error in the realms of 1% with 94.8% reduction in computation time for high-fidelity spectra. Previous work used analytic radiative fields (Vander Wal *et al.* 2022b); this work instead uses realistic radiative fields from Hydra simulations of ICF experiments that must be compressed with an autoencoder first before predictions can be made. This is similar to the work previously performed by Kluth *et al.*, except we increase the resolution of the models and leverage transfer learning to improve the fidelity of the opacity calculations (Kluth *et al.* 2020). Further differentiating it from the previous work, while not fully explored here, is the implementation of a novel neural network architecture and improved loss function for effective training.

2. Methods

2.1. Data

This work utilizes two separate datasets: one high-fidelity and one low-fidelity dataset. The low-fidelity model utilizes 1849 levels, and the high-fidelity model utilizes 25 903 levels in atomic state. The models have sample rates of 35.3 and 3.2 samples min^{-1} respectively. Ultimately, this work uses 26 000 low-fidelity samples and 1000 high-fidelity samples. In the following sections we outline the process of generating the data, and then discuss the training of the neural networks.

2.1.1. Step one – production of radiative field data

An initial set of low-fidelity Hydra simulations are generated where Cretin provides the NLTE opacities, and Hydra produces radiative fields that are extracted for training. In addition to the input radiative fields, the input temperature and density and the output

absorptivity and emissivity spectra are also extracted for neural network training. In this work, 16 000 samples are obtained from the Hydra simulations.

2.1.2. Step two – production of low-fidelity absorptivity and emissivity data

Second, Cretin produces an additional 10 000 samples of low-fidelity absorptivity and emissivity spectra of krypton using the radiative fields and randomly sampled density and electron temperatures as inputs. The spectra produced consist of 400 energy bins ranging from 0 to 40 000 eV. The density is sampled using a quasi-latin-hypercube sampling (McClarren 2018) weighted by the distribution of inputs obtained from the simulation data. The radiative fields are randomly sampled from a list of radiative fields that are binned together with the input values based on the integrated intensity of the radiative fields.

To provide some more clarity, imagine a three-dimensional histogram that is binned with the electron temperature, the \log_{10} transformation of the density and the integrated intensity of the radiative fields in the three sampled dimensions and the counts of those bins in the fourth dimension. The bins with temperatures >300 eV are more heavily weighted to increase the number of samples above 300 eV. The resultant weighting provides a roughly 50–50 split of temperature values above and below 300 eV as opposed to the 20–80 split that is actually present.

The histogram is normalized and ordered in ascending order of the normalized counts with all bins containing a value of zero being excluded. A bin is then randomly sampled using latin-hypercube sampling. The density and temperature are then randomly sampled from random uniform distribution bounded by the edges of the bin. The radiative field is then sampled from a list of radiative fields that possess an integrated intensity that falls between the edges of the sampled bin. The binning is rather coarse to allow for the sampling of input values beyond the edge cases found in the simulation data. This is done to help improve the generalizability of the neural network beyond the simulations used to produce the radiative fields and other input values. Finally, the reason for not sampling from a random uniform distribution between two extrema is because that results in many input combinations that would not reasonably occur during an ICF simulation.

2.1.3. Step three – production of high-fidelity absorptivity and emissivity data

Third, Cretin further produces higher-fidelity NLTE absorptivity and emissivity spectra utilizing the same process and distribution as in step two. The production of the higher-fidelity spectra utilizes the same radiative fields from the low-fidelity Hydra simulations, since the premise of this study is that the high-fidelity simulations are too expensive to run hundreds of times. It should be noted that none of the absorptivity and emissivity samples directly from the simulations are used in the high-fidelity dataset because they are from low-fidelity models. Instead, the entire high-fidelity dataset is sampled using the prescribed sampling method.

2.2. Process

2.2.1. Training

While this work is a direct continuation of the work in Kluth *et al.* (2020) and Vander Wal *et al.* (2022b) it does not use the same method of building a prediction model. Instead of creating separate autoencoders for the radiative fields and absorptivity and emissivity spectra that are purely sequential, fully connected or convolutional neural networks, it utilizes a single, non-sequential, full-connected autoencoder for all radiative fields, absorptivity spectra and emissivity spectra. The autoencoder is constructed as shown in figure 1 such that each layer of the encoder passes information to both the next layer in

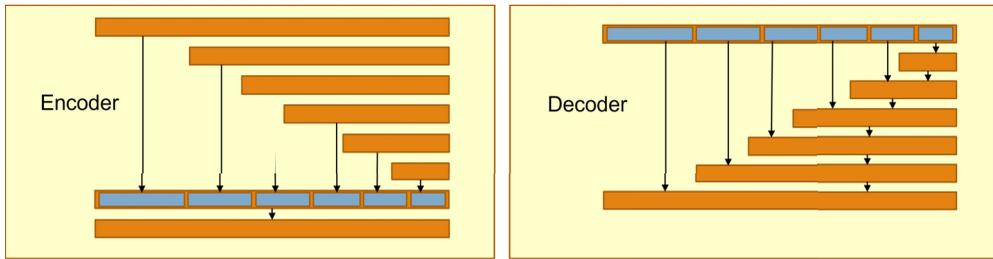


FIGURE 1. This is the general architecture of the encoder and decoder of the autoencoder. The latent space is a fully connected layer that mixes the prior layer which contains sets of nodes that are directly connected to their own specific layers including the input and output.

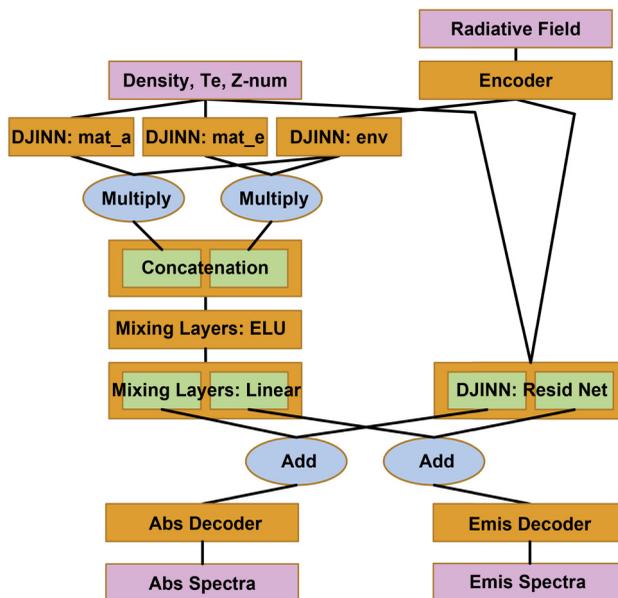


FIGURE 2. This is the full architecture as described. Excluded from this diagram are the locations and types of transformation operations.

the encoder as well the layer just prior to the latent space. The decoder does the exact opposite. Further, this work does not strictly use separate Deep, Jointly-Informed Neural Network (DJINN) (Humbird, Peterson & McClarren 2019) models for the absorptivity and emissivity neural network models as is done with the models in Kluth *et al.* (2020) and Vander Wal *et al.* (2022a,2022b) where a single DJINN model takes the radiative field latent space, density, temperature, atomic number as inputs to predict the latent space absorptivity or emissivity latent space. Instead, as seen in Figure 2, two separate DJINN models are trained to predict the latent space of the absorptivity and emissivity spectra as encoded by the single, multiscale autoencoder by using only the material properties: density, temperature and atomic number but not the radiative field which is treated as an environmental property. The density, temperature and atomic number are minmax scaled. The two DJINN models are trained for 100 epochs and then combined in parallel into a single model.

Next, another DJINN model is made using only the radiative field as encoded by the multiscale autoencoder. It is initialized by using the encoded emissivity spectra but not actually trained initially; instead this DJINN model is placed in parallel with the other two DJINN models. The concatenated output of the three parallel models is then passed through three fully connected layers of dimensionality equal to the concatenated output of the DJINN modules. These layers use the exponential–linear unit activation function. The output of these layers is then passed to a fully connected layer of dimensionality equal to two times the dimensionality of the latent space of the multiscale autoencoder. This layer uses a linear activation function. All of these sub-models are combined into a single model, and the output of the model is expected to be the encoded absorptivity and encoded emissivity concatenated to each other.

The combined model is then evaluated with the training data, and the signed error is used to initialize yet another DJINN model that uses the entire input space: encoded radiative field, density, temperature and atomic number, which are all minmax scaled, with the error in the latent space predictions being the output for initialization. No initial training is done here. The output of this DJINN model is then added to the output of the first combined model much like a residual block, and this newly combined model is then trained for 100 epochs to once again predict both the encoded absorptivity and encoded emissivity. Finally, the encoder of the multiscale autoencoder is prepended to the combined latent space prediction model, and the encoded absorptivity is passed to one instance of decoder of the multiscale autoencoder and the encoded emissivity is passed to a duplicate instance of the same decoder such that they can be trained independently. These are all combined into one model and then trained for 10 000 epochs.

The training datasets consist of 80 % of all of the available data with the other 20 % used for validation. All training, including for the autoencoder, is done with a batch size that is 1 % of the training dataset. The learning rate for low-fidelity training is set to 0.0001. The autoencoder is trained for 2500 epochs. The high-fidelity transfer learning only consists of performing additional training the completed model with the high-fidelity data. The high-fidelity transfer learning is performed with a learning rate of 0.00001, and the training is performed for 1000 epochs. The loss function in all cases is the fraction of unexplained variance is given by

$$\text{Loss} = \frac{\sum_{i=1}^N (y' - y)^2}{\sum_{i=1}^N (y' - \bar{y}')^2}. \quad (2.1)$$

Here, y is the expected value, y' is the predicted value and \bar{y}' is the mean of the predicted values. We found the use of this loss function is particularly important because it enables better shape matching of the predicted spectra.

Some additional points that must be known is that the DJINN software was modified to enable the usage of unity input and output scaling (Humbird). The DJINN models, as previously described, use various combinations of scalings, but all outputs use unity scaling. Also, the autoencoder input and output, and thus the input and output of the full model, use 30th-root scaling as explained in Vander Wal *et al.* (2022b).

Finally, while a single neural network can be used for each model, the authors opted to make two separate networks to constitute one neural network model. One model is for inputs with an electron temperature above 500 eV, and the other model is for inputs with an electron temperature below 500 eV. They are trained with inputs that

have electron temperatures 100 eV below and 100 eV above the 500 eV split point for the ‘high-temperature’ and the ‘low-temperature’ models, respectively. This is done so that the split point is not the edge of the model’s trained input space. The 500 eV split point is chosen because there is a non-hard transition where the prediction accuracy quickly worsens with decreasing temperature that starts somewhere in the realms 300 to 500 eV. Emissivity prediction is particularly poor for low temperatures. Absorptivity prediction is worse at increased temperatures; however, it is not necessarily the worst at the highest of temperatures. Rather, absorptivity performance is worst in a region located roughly around 900 to 1400 eV.

2.2.2. Hydra

The simulations of ICF experiments are performed using the radiation-hydrodynamics code Hydra. The simulations use the same set-up as those performed in Kluth *et al.* (2020) with the exception of the laser pulse profiles used and the binning structure used for the spectra.

The simulations use a one-dimensional spherical hohlraum made of krypton with 64 cells. The krypton hohlraum is set to have a density of 7.902 g cc^{-1} . Inside the hohlraum there is a fuel capsule consisting of an ablator made of polyethylene filled with deuterium–tritium ice fuel. There are various additional trace elements present in the fuel and ablator such as oxygen, germanium and a few others. The absorptivity and emissivity of these materials were calculated using standard, tabulated values. The use of neural networks will be applied to the approximation of the krypton hohlraum.

The laser pulse is applied as an internal source. This means that the pulse does not traverse space in the simulation and is applied to the interior cell, ‘cell 0’, directly. The ten laser pulse profiles that are used for both data generation and testing are shown in figure 3. During the generation of the radiative fields and a portion of the training data, the simulation of the hohlraum is handled as an NLTE problem with the radiation-hydrodynamics software Cretin computing the NLTE opacities and handled as an inline operation in Hydra. Cretin was set to use steady-state approximations and use Steward–Pyatt continuum lowering. When using a neural network to compute the NLTE opacities, Hydra passes the input data to a Python function which calculates the opacities with the neural networks.

These same ten laser profiles are used for the evaluation of the performance of the neural networks. First, all neural network models will be used in conjunction with one laser profile. Then, based on the median error in the radiative temperature, the fourth-best neural network model will be used to simulate the outcomes from all ten of the laser profiles used to produce the training data.

3. Results

Three simulation outputs of interest that are used to evaluate performance in the simulations are the radiative temperature, the electron or material temperature and the density. The radiation temperature is of particular interest because it can be measured in ICF experiments, whereas the material temperature and density have to be inferred. Before getting into the simulation results, the neural networks performance is demonstrated in table 1, which shows the median values for each of the means, medians, 90th percentiles and maximums across all ten models. The most notable observation is that the median of the maximum $1 - R^2$ value, where R^2 is the coefficient of determination, is below one for both absorptivity and emissivity for both models above and below 500 eV. This is not possible with the original method of neural network construction utilized in Kluth *et al.*

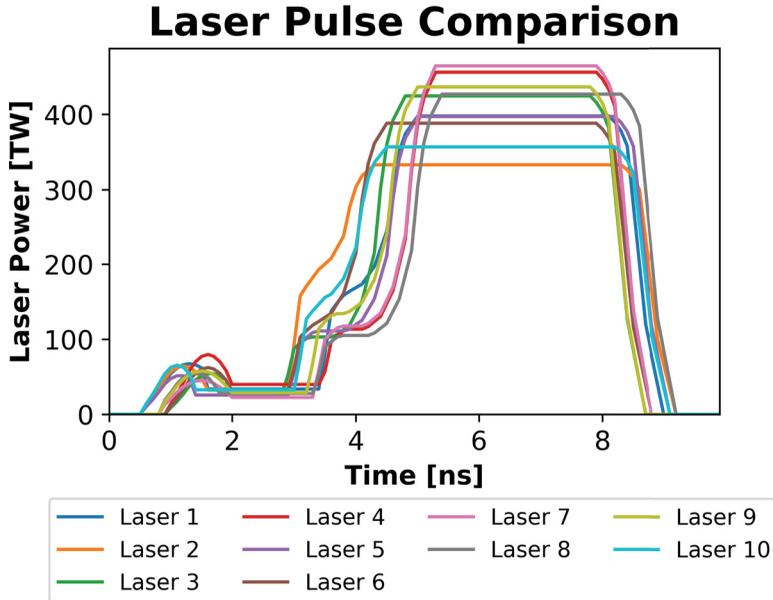


FIGURE 3. These are the ten laser profiles used in the simulations. They are randomly generated perturbations to the profile of the N210808 shot at the National Ignition Facility (Callahan 2021).

(2020) and Vander Wal *et al.* (2022a,2022b), thus the method described in section II was chosen. The median of the median $1 - R^2$ value is exceptionally low as well.

In our 64 cell discretization in space, cell 0 is the surface that faces the fuel capsule and is, therefore, considered the most important cell of the simulation. The maximum radiation temperature could also be considered to be a good point of comparison for the level accuracy achieved. The left-hand plot in figure 4 shows the the interquartile range as well as the ranges of the lower and upper quartiles. They are plotted against the results of simulations ran with low-fidelity spectra and the high-fidelity spectra as well as the mean and median predicted values of the radiation temperature between the ten different neural network models. The interquartile range of the radiation temperature for each cell at various times are displayed in the right-hand plot of figure 4. Surprisingly, there is significantly higher uncertainty at earlier times and in cells between times and cells with low uncertainty. This is particularly apparent in the cells near cell 48, which have a significantly higher level of uncertainty around 4.0 ns into the simulation, yet by the end of the simulation, there is very little uncertainty in the simulation as seen in the 8.0 and 9.0 ns lines.

The right-hand plot also displays an apparent shift in the cell number as time advances where the median prediction begins to significantly deviate from the expected temperature profile. This deviation is expected to arise from a rapid and significant change in the radiative field intensity with the rapid change arising from the energy imparted by the laser pulse, shock heating and the radiative field from the surrounding material. Further, the contributions from the laser and the other cells are filtered by the plasma in between the sources and a given cell. The radiative fields in this ‘front’ should not prevent the implementation in future use because the rapidly transitive and evolving nature of the front makes it difficult to capture without prior knowledge, so the radiative fields in this front likely make up a small portion of the training data. Thus, the authors expect that

Absorptivity				
Binwise % Error				
	Mean	Median	90th-Perc.	Max.
>500 eV	2.784	1.682	6.477	71.54
<500 eV	20.134	11.26	42.32	1151
Unexplained Variance				
	Mean	Median	90th-Perc.	Max.
>500 eV	8.780×10^{-4}	5.342×10^{-4}	1.621×10^{-3}	7.699×10^{-3}
<500 eV	1.821×10^{-2}	5.589×10^{-3}	4.177×10^{-2}	0.3729
Emissivity				
Binwise % Error				
	Mean	Median	90th-Perc.	Max.
>500 eV	7.082	3.826	14.78	3479
<500 eV	1.691×10^{17}	51.71	771.7	7.175×10^{21}
Unexplained Variance				
	Mean	Median	90th-Perc.	Max.
>500 eV	2.603×10^{-4}	1.581×10^{-4}	4.413×10^{-4}	2.272×10^{-3}
<500 eV	1.333×10^{-2}	5.298×10^{-3}	3.126×10^{-2}	0.1659

TABLE 1. The neural networks’ performance on per cent relative error and unexplained variance for absorptivity and emissivity as compared against high-fidelity data.

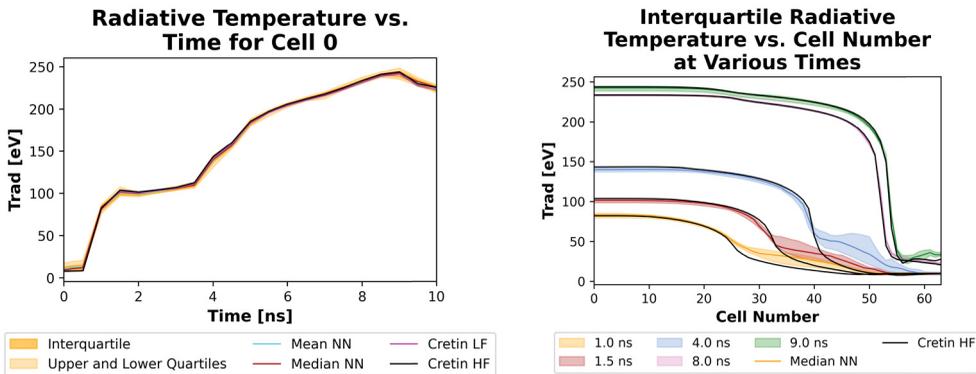


FIGURE 4. These plots track the radiative temperature across time for cell 0 (left) and across the cells for various times (right). The temperature tracking at cell 0 is quite effective; however, for other cells, the cells where the shock is present for a given time step have significant deviation from the expected values. In the bottom plot, the shaded regions represent the interquartile range with median represented by the solid line sharing the same colour as the shaded region. The black lines are the expected values from the simulations.

increasing the number of samples from this phase of the simulation by either increasing the sampling rate globally for cells and time or in a more targeted fashion such as around the periods of increase in laser output power or potentially by even making a moving frame of increased sampling that follows the front.

The relative error near the peak of the radiation temperature is quite small. Interestingly, this happens to occur after having significantly higher relative error at the earlier times of

% Error	Time Agnostic	Time Corrected
Mean	1.042	1.232
Median	1.049	1.122
Maximum	1.585	2.871

TABLE 2. The per cent relative error of the maximum radiation temperature in cell 0 for both time agnostic and time corrected considerations.

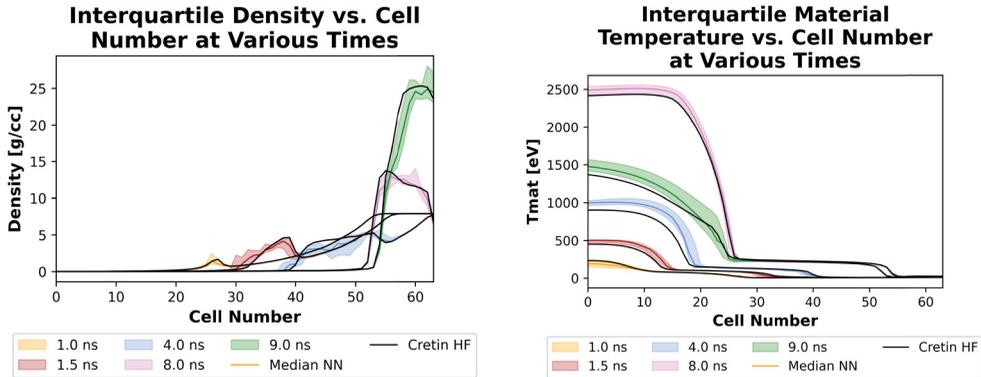


FIGURE 5. These plots of the density and material (electron) temperature across the cells for various times demonstrate that the density and temperature are generally predicted well when using neural networks. The density has the highest deviations near the shock, and the material temperature has the highest deviations near the laser source and fuel capsule. The shaded regions represent the interquartile range with median represented by the solid line sharing the same colour as the shaded region. The black lines are the expected values from the simulations.

the simulation. The mean, median and maximum per cent relatives errors can be seen in table 2. The ‘time agnostic’ column refers to the fact this is the error between the two maximum temperatures found anywhere in cell 0 for a given simulation. The ‘time corrected’ column refers to fact that this column displays the error found at the same time as the maximum radiation temperature in the high-fidelity simulation. The time-corrected median relative error comes out at 1.05 %, and the maximum time-corrected relative error comes out at 2.87 %.

The density and material temperature perform better in mean per cent relative error than the radiation temperature. The mean error across all cells and sample times for density and material temperature are 18.7 % and 23.8 %, respectively, as opposed to 26.9 % for radiation temperature. However, as can be seen in the two plots in figure 5, large portions of the cells for both variables see any deviation in areas that might be considered of low interest based on the baseline computation. In areas of increased activity or deviation from a seemingly constant state, the predictions become significantly worse. Additionally, just as in the radiation temperature plot in figure 4, the cell profile at 4.0 seconds shows some of the largest deviations from the expected values.

As a note of understanding what goes on during the simulation, one can see the propagation of the shock through the material in the left-hand plot of figure 5 as the hump of increased density moves through the cells and actually grows over time. Similarly, in

% Error	Time Agnostic		Time Corrected	
	5th Model	6th Model	5th Model	6th Model
Mean	0.4063	0.5819	3.0945	1.130
Median	0.3048	0.6414	3.760	1.193
Maximum	1.280	1.043	5.985	1.729

TABLE 3. The per cent relative error of the maximum radiation temperature in cell 0 for both time agnostic and time corrected considerations for two models as used with simulations with each laser pulse.

the right-hand plot, one can see how the radiation is filtered out by the plasma closest to the laser source by the fact that only the low numbered cells see significant increases in temperature. The shock heating can also be observed by comparing the location in the step in the material temperature near cell 50 at 8.0 and 9.0 seconds with the density profiles at those same times.

The second part of testing the neural networks involves choosing the fifth and sixth best neural networks based on their mean radiation temperature error across all cells and times for laser profile 1. These two models were selected for the purpose of choosing models that may represent a more common outcome of training the neural networks. They are neither the best nor the worst models. The results are simultaneously surprising and not surprising. The mean and median error in radiation temperature across all cells, times and laser pulses is lower for the fifth model than the sixth model, with mean errors of 20.3 % and 21.4 %, respectively. The median errors are even more different at 4.40 % and 7.70 %. However, when the results of [table 3](#) are considered, the time-corrected error for the maximum radiation temperature of cell 0 is significantly better for the sixth model with a median error of 1.193 % compared with the 3.760 % error from the fifth model. The maximum time-corrected error offers an even starker difference. That aside, if the time of the maximum temperature is not as important as simply obtaining the correct maximum temperature, the fifth model performs better with a median error of 0.3048 % compared with the 0.6414 % of the sixth model.

Lastly, the speed up in computation of opacities is considerable based on the timing of the neural network evaluation compared with inline calls from Hydra to Cretin. For reference, the speedup over the low-fidelity simulations is roughly $2.36\times$ or rather the neural networks consume 42.4 % of the time that the computations performed by Cretin do. The speed up for the high-fidelity simulation is exceptional with a speedup of $19.4\times$ or rather neural networks consume 5.20 % of the regular computation time. An important note to make about this speedup in comparison with the speedup demonstrated by Kluth *et al.* (2020) is that we are reporting the speed up in the opacity calculations only. The time comparisons provided in Kluth *et al.* (2020) are based on running Cretin in a non-standard way involving Python script calls that introduces significant overhead. This leads to a significant slow down in the process of using Cretin, and more dramatic performance gains. In our results we are reporting neural network time to standard inline Cretin execution. Therefore, the speedup values should be expected to be smaller.

4. Conclusion

In this work, it has been demonstrated that neural network emulators of Cretin can recover the maximum radiation temperature in an integrated Hydra hohlraum simulation

with a mean relative error estimated to be 1.23 %. It has also been shown that a single network from an ensemble of 10 models, can achieve a mean relative error in the maximum radiation temperature in the range 1.13 %–3.10 % across multiple laser pulses if time corrected but can achieve around 0.41 %–0.58 % if not time corrected. Moreover, this model requires only 5.20 % of the computation time compared with inline Cretin calculations. The density and material temperature were also demonstrated to closely match Cretin; however, there were regions of space and time that saw fairly significant deviation. As such, those who desire to analyse the physics further from the innermost cells will need to further improve upon these results before that can be reliably done.

An improvement in results could be achieved by using more low-fidelity data. There could also be better tuning of when the simulations are sampled for the radiative fields, absorptivity spectra and emissivity spectra. The sampling times used here were regularly spaced, but the sampling times could be optimized in frequency and not necessarily be regularly spaced. The measured uncertainty with these models could also be further improved by utilizing laser profiles that were not used to generate training data unlike what was done in this work.

An alternative to a single neural network that predicts the properties of multiple chemical elements would be to use several, single element models (Vander Wal *et al.* 2022a). This could be used to replace the fuel and fuel capsule models, which are currently based on table lookups, with neural networks might be one way multi-element models can be used. While this approach may be slower, high-fidelity neural networks trained on more accurate NLTE data may produce more realistic simulations. Further, as expansion on previous work for multiple elements and transfer learning as well as this work, high-fidelity multi-element models could be produced that actually utilize mix ratios of multiple elements such as would be seen in the fuel and potentially fuel capsule (Vander Wal *et al.* 2022a,2022b).

More work could also be done on the transfer learning process, which could sequentially or concurrently use multiple levels of fidelity (Aydin *et al.* 2019; Meng & Karniadakis 2020). In particular, a model that uses linear and nonlinear correlations to compensate may be great choice to further expand the accuracy (Meng & Karniadakis 2020). For the ambitious, physics informed neural networks (PINNs) may be an interesting path to take. This route could provide high quality results if past work is a reasonable indicator of performance; however, it would likely require a custom training loop that would have to be run in concurrently with a radiation-hydrodynamics software package like Cretin (Raissi, Perdikaris & Karniadakis 2019; Yang & Perdikaris 2019; Dwivedi & Srinivasan 2020; Jagtap, Kharazmi & Karniadakis 2020; Meng & Karniadakis 2020; Rao, Sun & Liu 2020; Karniadakis *et al.* 2021; Zhang *et al.* 2021; Kim *et al.* 2022). Additionally, the PINNs could be used in special ways to perform transfer learning (Goswami *et al.* 2020; Meng & Karniadakis 2020; Chakraborty 2021).

There could also be further improvements in the architecture in ways not related to PINNs. The architecture used in this work are fully connected with additional skip-layer features included, but a convolutional network with similar skip-layer features may perform better. There could also be an introduction more closely related to artificial intelligence (AI) for the task of determining which network should take high-temperature inputs and which network should take low-temperature inputs. Rather, a network that is introduced for this reason could choose to split along a nonlinear boundary that is based on all input variables. Finally, there could exist a better way to auto-construct networks than using DJINN. One such possibility is the class of evolutionary neural networks which could potentially be smaller while also better predicting than networks made with DJINN

(Cui, Shabash & Wiese 2019; O'Neill, Xue & Zhang 2021; Xue *et al.* 2021a; Xue, Wang & Liang 2021b).

Acknowledgements

Editor William Dorland thanks the referees for their advice in evaluating this article.

Funding

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. The work of MDV and RGM was supported by Lawrence Livermore National Laboratory under contract B638638, 'Non-LTE Physics in Integrated Calculations using Machine Learning'.

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation or favouring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes. Released as LLNL-JRNL-834755-DRAFT.

Declaration of interests

The authors report no conflict of interest.

REFERENCES

- AYDIN, R.C., BRAEU, F.A. & CYRON, C.J. 2019 General multi-fidelity framework for training artificial neural networks with computational models. *Front. Mater.* **6**.
- BARD, C. & DORELLI, J. 2021 Neural network reconstruction of plasma space-time. *Front. Astron. Space Sci.* **8**.
- CALLAHAN, D.A. 2021 Achieving a burning plasma on the national ignition facility laser. In *Presented at the 63rd Annual Meeting of the APS Division of Plasma Physics*. American Physical Society.
- CHAKRABORTY, S. 2021 Transfer learning based multi-fidelity physics informed deep neural network. *J. Comput. Phys.* **426**, 109942.
- CHATZIDAKIS, M. & BOTTON, G.A. 2019 Towards calibration-invariant spectroscopy using deep learning. *Sci. Rep.* **9**.
- CHEN, Y.-Y. & WANG, Z.-B. 2019 Feature selection based convolutional neural network pruning and its application in calibration modeling for nir spectroscopy. *Chemometr. Intell. Lab. Syst.* **191**.
- CUI, C. & FEARN, T. 2018 Modern practical convolutional neural networks for multivariate regression: applications to nir calibration. *Chemometr. Intell. Lab. Syst.* **182**.
- CUI, P., SHABASH, B. & WIESE, K.C. 2019 Evodnn - an evolutionary deep neural network with heterogeneous activation functions. In *2019 IEEE Congress on Evolutionary Computation*. IEEE.
- DEGRAVE, J.E.A. 2022 Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature* **602**.
- DENG, J., XIA, R., ZHANG, Z., LIU, Y. & SCHULLER, B. 2014 Introducing shared-hidden-layer autoencoders for transfer learning and their application in acoustic emotion recognition. In

- 2014 *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, pp. 4818–4822. IEEE.
- DENG, J., ZHANG, Z., MARCHI, E. & SCHULLER, B. 2013 Sparse autoencoder-based feature transfer learning for speech emotion recognition. In *2013 Humaine Conference on Affective Computing and Intelligent Interaction*.
- DWIVEDI, V. & SRINIVASAN, B. 2020 Physics informed extreme learning machine (pielm) - a rapid method for the numerical solution of partial differential equations. *Neurocomputing* **391**, 96–118.
- EUSEBIO, J.M. 2018 Learning transferable data representations using deep generative models. Master's thesis, Arizona State University.
- GHOSH, K., STUKE, A., TODORVIĆ, M., JØRGENSEN, P.B., SCHMIDT, M.N., VEHTARI, A. & RINKE, P. 2019 Deep learning spectroscopy: neural networks for molecular excitation spectra. *Adv. Sci.* **6**.
- GOSWAMI, S., ANITESCU, C., CHAKRABORTY, S. & RABCZUK, T. 2020 Transfer learning enhanced physics informed neural network for phase-field modeling of fracture. *Theor. Appl. Fract. Mech.* **106**.
- HUMBIRD, K.D. 2022 DJINN. Github repository, <https://github.com/LLNL/DJINN>.
- HUMBIRD, K.D. & PETERSON, J.L. 2020 Transfer learning to model inertial confinement fusion experiments. *IEEE Trans. Plasma Sci.* **48**.
- HUMBIRD, K.D., PETERSON, J.L. & MCCLARREN, R.G. 2019 Deep neural network initialization with decision trees. *IEEE Trans. Neural Netw. Learn. Syst.* **30** (5).
- HUMBIRD, K.D., PETERSON, J.L., SALMONSON, J. & SPEARS, B.K. 2021 Cognitive simulation models for inertial confinement fusion: combining simulation and experimental data. [arXiv:2103.10590v1](https://arxiv.org/abs/2103.10590).
- IMAI, S., KAWAI, S. & NOBUHARA, H. 2020 Stepwise pathnet: a layer-by-layer based transfer learning algorithm. *Nature* **10**.
- JAGTAP, A.D., KHARAZMI, E. & KARNIADAKIS, G.E. 2020 Conservative physics-informed neural networks on discrete domains for conservation laws: applications to forward and inverse problems. *Comput. Meth. Appl. Mech. Engng* **365**.
- KARNIADAKIS, G.E., KEVREKIDIS, J.G., LU, L., PERDIKARIS, P., WANG, S. & YANG, L. 2021 Physics-informed machine learning. *Nat. Rev. Phys.* **3**.
- KIM Y., CHOI Y., WIDENMANN D. & ZOHDI T. 2022 A fast and accurate physics-informed neural network reduced order model with shallow masked autoencoder. *J. Comput. Phys.* **451**, 110841.
- KLUTH, G., HUMBIRD, K., SPEARS, B.K., PETERSON, J.L., SCOTT, H.A., PATEL, M.V., KONING, J., MARINAK, M., DIVOL, L. & YOUNG, C.V. 2020 Deep learning for NLTE spectral opacities. *Phys. Plasmas* **10**.
- KUSTOWSKI, B.L., GAFFNEY, J.A., SPEARS, B.K., ANDERSON, G.J., THIAGARAJAN, J.J. & ANIRUDH, R. 2019 Transfer learning as a tool for reducing simulation bias: application to inertial confinement fusion. *IEEE Tran. Plasma Sci.* **48** (1), 46–53.
- LAN, Q., MCCLARREN, R.G. & VISHWANATH, K. 2021 Neural network forward model and transfer learning calibration from monte carlo to diffuse reflectance spectroscopy. In *Optical Tomography and Spectroscopy of Tissue XIV*, vol. 11639, p. 116390E. International Society for Optics and Photonics.
- LIU, B., LI, Y., LI, G. & LIU, A. 2019 A spectral feature based convolutional neural network for classification of sea surface oil spill. *ISPRS: Intl J. Geo-Inf.* **8**.
- LUO, W., LI, J., YANG, J., XU, W. & ZHANG, J. 2018 Convolutional sparse autoencoders for image classification. *IEEE Trans. Neural Netw. Learn. Syst.* **29**.
- MCCLARREN, R.G. 2018 *Uncertainty Quantification and Predictive Computational Science*. Springer.
- MENG, X. & KARNIADAKIS, G.E. 2020 A composite neural network that learns from multi-fidelity data: application to function approximation and inverse PDE problems. *J. Comput. Phys.* **401**.
- NI, C., WANG, D. & TAO, Y. 2019 Variable weighted convolutional neural network for the nitrogen content quantization of masson pine seedling leaves with near-infrared spectroscopy. *Spectrochim. Acta A* **200**.
- O'NEILL, D., XUE, B. & ZHANG, M. 2021 Evolutionary neural architecture search for high-dimensional skip-connection structures on densenet style networks. *IEEE Trans. Evol. Comput.* **25** (6).

- RAISSI, M., PERDIKARIS, P. & KARNIADAKIS, G. 2019 Physics-informed neural networks: a deep learning framework for learning forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**.
- RAO, C., SUN, H. & LIU, Y. 2020 Physics-informed deep learning for incompressible laminar flows. *Theor. Appl. Mech. Lett.* **10**, 207–212.
- VANDER WAL, M.D., MCCLARREN, R.G. & HUMBIRD, K.D. 2022a Neural network surrogate models for absorptivity and emissivity spectra of multiple elements. *Mach. Learn. Appl.* **8**.
- VANDER WAL, M.D., MCCLARREN, R.G. & HUMBIRD, K.D. 2022b Transfer learning of high-fidelity opacity spectra in autoencoders and surrogate models. *IEEE Trans. Plasma Sci.* doi: 10.1109/TPS.2022.3227506
- XUE, Y., JIANG, P., NERI, F. & LIANG, J. 2021a A multiobjective evolutionary approach based on graph-in-graph for neural architecture search of convolutional neural networks. *Intl J. Neural Syst.* **31** (9).
- XUE, Y., WANG, Y. & LIANG, J. 2021b A self-adaptive mutation neural architecture search algorithm based on blocks. *IEEE Comput. Intell. Mag.* **16** (3), 67–68.
- YANG, Y. & PERDIKARIS, P. 2019 Adversarial uncertainty quantification in physics-informed neural networks. *J. Comput. Phys.* **394**, 136–152.
- ZHANG, Y., FU, H., QIN, Y., WANG, K. & MA, J. 2021 Physics-informed neural network for inhomogeneous magnetized plasma parameter inversion. *IEEE Antennas Wirel. Propag. Lett.* **21** (4), 828–832.
- ZHUANG, F., CHENG, X., LUO, P., PAN, S.J. & HE, Q. 2015 Supervised representation learning: transfer learning with deep autoencoders. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*.
- ZHUANG, F., QI, Z., DUAN, K., XI, D., ZHU, Y., ZHU, H., XIONG, H. & HE, Q. 2021 A comprehensive survey on transfer learning. *Proc. IEEE* **109** (1), 43–76.
- ZHU, Y., WU, X., LI, P., ZHANG, Y. & HU, X. 2019 Transfer learning with deep manifold regularized auto-encoders. *Neurocomputing* **369**.