Abstract—Proton beam radiotherapy is a cancer treatment method that uses proton beams to irradiate cancerous tissue while simultaneously sparing doses to healthy tissue. In order to optimize radiational doses to the tumor and ensure that healthy tissue is spared, many researchers have suggested verifying the treatment delivery through real-time imaging. One promising method of real-time imaging is through a Compton camera, which can image prompt gamma rays emitted along the beam’s path through the patient. However, the reconstructed images are often noisy and unusable for verifying proton treatment delivery due to limitations with the camera. We present the usage of deep learning to remove and correct the various problems that exist within our data.

Index Terms—proton beam therapy, Compton camera, deep learning, image reconstruction

I. INTRODUCTION

Proton beams’ primary advantage in cancer treatment as compared to other forms of radiation therapy, such as x-rays, is their finite range. The radiation delivered by the beam reaches its maximum, known as the Bragg peak, at the very end of the beam’s range. Little to no radiation is delivered beyond this point. By exploiting the properties of the Bragg peak, it is possible to only irradiate cancerous tissues, avoiding any damage to the healthy surrounding tissues [1]. However, without some way to image proton beams in real time, limitations exist in our ability to take full advantage of the dose delivery properties of the proton Bragg peak. This is due to uncertainties in the beam’s position in the body relative to important organs that should not be irradiated.

The Compton camera is one method for real time imaging, which works by detecting prompt gamma rays emitted along the path of the beam. By analyzing how prompt gamma rays scatter through the camera, it is possible to reconstruct their origin. However, the raw data that the Compton camera outputs does not explicitly record the sequential order of the interaction data, which represents scatterings of a single prompt gamma ray. In addition, it often records false events, which mislabel scatterings of distinct prompt gamma rays as originating from a single ray. These problems make reconstructions based on Compton camera data noisy and unusable for practical purposes [1].

We approach these problems by leveraging several deep learning techniques. We use neural networks, which, in general, represent data transformations. The network is trained by passing data through it, then updating it systematically so as to reduce the loss of its output compared with some desired output. Doing this properly can create a model that exploits subtleties in the data which traditional models are unable to use [2]. We show how this can be done in the following sections. Additional discussion about the impact of the approach on the application area can be found in [3].

II. PROTON BEAM THERAPY

To a first order approximation, the radiation dosage emitted by a proton beam is inversely proportional to the kinetic energy of the particles within the beam. Because the beam’s particles lose kinetic energy as they traverse the patient, the amount of radiation delivered by the beam is low at its entry point, gradually rising until the beam nears the end of its range, at which point the delivered dosage rapidly reaches its maximum. This point of maximum dosage is called the Bragg peak. Little to no radiation is delivered beyond the Bragg peak. These characteristics of proton beam therapy give it a distinct advantage over x-rays. Exploiting its finite range, medical practitioners can confine the radiation of the beam to solely areas affected by cancerous tumors. Vital organs beyond the tumor can be spared [1].

While the characteristics of proton beam therapy explained above would in principle greatly reduce the negative effects of radiation therapy, there are still practical limitations. In current practice, the patient’s body is imaged before undergoing
treatment in order to map the position of the tumor. Because proton beam therapy consists of multiple sessions over a period of one to five weeks, the relative size and position of the tumor within the patient’s body may change as surrounding tissues swell, shrink, and shift as a response to radiation. Therefore, whenever using proton beams, a safety margin must be added to the position of the Bragg peak in order to fully irradiate the tumor. This rules out certain beam trajectories that would otherwise minimize damage to healthy tissue [1].

Figure 1 compares two possible beam trajectories through a cross-section of the chest [1]. In this case, the heart, outlined in purple is positioned at the top-center of the figure and a tumor, outlined in green is located next to it. The optimal trajectory, shown in the left image, uses a single beam, which is represented as the space between the dashed white lines, to fully irradiate the tumor while stopping before reaching the heart. However, due to uncertainty in the exact location that the Bragg peak occurs (and the beam stops), a safety margin is added to the optimal beam extent to ensure the tumor always receives the prescribed dose even in the presence of day-to-day changes in patient setup and patient internal anatomy. This safety margin is represented in the figure as an orange strip at the end of the beam. This partially overlaps with the heart, which would mean to possibly irradiate the heart. Therefore, in practice, the trajectory in the right image using two beams is used. Because this trajectory passes through the lungs, delivering a small dose of radiation to them, it is considered preferable, thought suboptimal [1]. If we were able to provide real-time information on the proton beam during live treatment, the safety margin could be smaller and, thus, the more optimal single-beam treatment.

III. COMPTON CAMERA IMAGING

A. Introduction to the Compton Camera

In order to exploit the full advantages of proton therapy, many researchers are investigating methods to image the beam in real time as it passes through the patient’s body [1]. One proposed method for real-time imaging is by detecting prompt gamma rays that are emitted along the path of the beam using a Compton camera.

As the proton beam enters the body, protons in the beam interact with atoms in the body, emitting prompt gamma rays. These prompt gamma rays exit the body, some of which enter the Compton camera. Modules within the Compton camera record interactions with energy levels above some trigger-threshold. These modules have a non-zero time-resolution during which all interactions are recorded as occurring simultaneously. For each interaction (that is, Compton scatter) an \((x, y, z)\) location and the energy deposited are recorded. The collection of all interaction data that a camera module collects during a single readout cycle is referred to as an event [4].

In principle, it is possible to use the data that the Compton camera outputs (paired with a suitable reconstruction algorithm) in order to image the proton beam, however, this has been shown to only be feasible at low energy levels. At the higher energy levels more typical of proton beam therapy, reconstructions of the beam are far too noisy to be helpful. This is a result of two main limitations in how the Compton camera records events further explained in [4].

At the higher energy levels typically used in treatment, proton beams emit a larger number of prompt gamma rays per unit time, increasing the likelihood of false events. Also, prompt gamma rays are more likely to scatter at higher energy levels, leading to more multi-scatter events, which, as explained above will be unordered. These two effects greatly diminish the accuracy of Compton camera reconstructions at high energy levels, making them unusable [4].

B. The Representation of Events

Multi-scatter events can be classified into five categories. A False Triple event consists of three interactions which all originate from separate prompt gamma rays that happened to enter the same module of the camera at the same time. These should be removed from the data before reconstruction. Similarly, False Double events contain two interactions originating from separate prompt gamma rays. These too should be removed. A Double to Triple event contains two interactions corresponding to the same prompt gamma ray, and one interaction from a different prompt gamma ray. The non-corresponding interaction should be removed before reconstruction. The two remaining categories of events are true double and true triple events, which, once properly ordered, can be used for reconstruction. The studies in this paper focus solely on true and false double events. Reference [5] studies the other categories in further detail.

Figure 2 shows a schematic of the Compton camera as it records events. The left side shows events produced at low energy levels and the right shows higher energy levels. Each row represents an independent module of the camera. The red arrows represent scatters, with those originating from the same prompt gamma ray being connected by a dotted line. A raised pulse represents a single readout cycle within a module of length \(T_A\). The value \(n\) is how many interactions occur during the readout cycle. Looking at just the left side, the first two rows show a True Double and True Triple event, respectively. The third row shows a False Double event consisting of two scatters originating from different prompt gamma rays. The fourth and fifth rows show two True Single events that consist of separate scatters by the same prompt gamma ray. The right
side representing higher energy levels shows a far greater proportion of false events.

The raw data output but the Compton camera for each interaction is of the form \((e_i, x_i, y_i, z_i)\), where for \(i = 1, 2, 3\), \(e_i\) is the energy level, and \(x_i, y_i, z_i\) are the \(x\), \(y\), and \(z\) coordinates respectively. \(i\) can equal 1, 2, or 3 as each event can have up to 3 interactions included. For events with less than 3 interactions, for the values of \(i\) that are not included in the event, the corresponding \((e_i, x_i, y_i, z_i)\) would be zeroed out.

To improve the performance of our networks, we find it useful to use the appended two values to the raw Compton camera data. In this version, we add the distances \(\delta r_{i,j}\) and the differences in energy levels \(\delta e_{i,j}\) between the \(i\)th and \(j\)th interactions, where \(i, j = 1, 2, 3\). Since these values have physical significance with regards to the ordering of interactions, explicitly including them in the data makes it easier for the networks to learn. For double events, the non-relevant \(\delta r_{i,j}\) and \(\delta e_{i,j}\) are set to zero.

IV. Deep Learning

We propose to train a neural network to process the data output by the Compton camera by removing false events and properly ordering the interactions within events.

The network contains three main components: an input layer which accepts the data, hidden layers which each perform some transformation on the data, and an output layer which returns the transformed data in some prescribed format [2]. We would like to train the neural network to transform the provided data in some useful way. In the case of the data output by the Compton camera, we would like the neural network to transform each multi-scatter event so that it contains only interactions originating from the same prompt gamma ray, and so that these interactions are in the correct order.

To improve the network’s performance, it is typical to train the network on all available data multiple times. One pass through all the training data is referred to as an epoch. Often, the network will be trained for hundreds or thousands of epochs. It is standard practice to set aside some data with which to evaluate the network after each epoch. These data are called the validation data. By evaluating the network at the end of each epoch, it is possible to plot how the network’s performance improves over the training process, giving insight into whether or not the network has been fully trained. After the network has finished training, a final data set separate from the training data and validation data is used to test the network. This data set is referred to as the test data.

References [5], [6] explore various network designs, activation functions, and preprocessing techniques for fully connected neural networks. Section VI are based on neural networks designed based on the results found in the reports.

V. Network Design Options

The studies in this work use a distributed-memory cluster of compute nodes with large memory, and connected by a high-performance InfiniBand network. In particular, the networks were trained and tested on one node with 4 NVIDIA Tesla V100 GPUs (5120 computational cores, 16 GB onboard memory) are by NVLink and two 18-core Intel Skylake CPUs. The node has 384 GB of memory (12 × 32 GB DDR4 at 2666 MT/s). These nodes are contained in the cluster taki of the UMBC High Performance Computing Facility (HPCF), whose webpage at https://hpcf.umbc.edu can provide more details.

VI. Results

For our studies, we considered 150MeV beams with three different dosage rates: 0kMU, 100kMU, and 180kMU. For these dosage rates, our data set consisted of 165,000, 50,000, and 28,000 samples, respectively.

Figure 3 displays three confusion matrices, one for the differing dosage rates. The leftmost column is the correct input class and the percent in each proceeding column represents the amount of data put into the class at the top of the column. Each cell within the matrix is shaded on a logarithmic scale. The shading is based on the accuracy of the cell, where higher accuracies result in darker shades of green. The darkest entry in each row is the dominant classification of the input class. Within the matrices, there are three different labels: 12, 21, and 44. Label 12 represents the case where both interactions are correctly ordered and no adjustments need to be made. Label 21 represents the scenario where the second interaction should be first, whereas the first interaction should be second. The third label 44, is a False event, where the two interactions should be thrown away. This is because the interactions originate from separate prompt gamma rays, thus making it extremely difficult to determine the origin of the two separate rays in reconstruction as there is not enough information.

For each of the confusion matrices, the dominant classification for each input class is the class itself. This indicates that the model was trained successfully and is able to correctly classify most samples. The matrix values between the dosage rates do not appear to change by more than 1%, where higher dosage rates have slightly higher accuracies. We see that the neural network classifies false events with around a 4% higher
Using the cleaned data, to Figure 4(c), there still is visual noise in the reconstruction noise as seen in uncleaned reconstruction. However, relative to Figure 4, where (a) shows the reconstruction image from the uncleaned data and (b) shows the reconstruction image from the cleaned data. Ultimately, we want the reconstruction images to resemble the Bragg Peak as shown in Figure 4(c). When comparing the Figure 4(a) and (b), we see that the cleaned data has removed a significant portion of the noise as seen in uncleaned reconstruction. However, relative to Figure 4(c), there still is visual noise in the reconstruction using the cleaned data.

Fig. 4: Comparison of gamma ray images reconstructed with (a) uncleaned data and (b) cleaned data to the (c) dose delivered by the proton beam.

VII. CONCLUSIONS

The results on the Doubles data were quite promising, with accuracies in the mid-80s. We see that with this level of accuracy, the neural network can produce cleaned data. The results then showed several reconstruction images using data with and without the support of the neural network. These images suggest that using neural networks can improve the quality of the reconstruction images by significantly reducing reconstruction noise.

There are several directions to take this research. First and foremost, the results in this paper are decent but can be improved. In-depth hyperparameter studies can be conducted to further improve network accuracy, which directly can reduce noise in the image reconstructions. Additionally, the work here only looked into fully connected neural networks. However, there are many deep learning models which may produce higher accuracies and better image reconstructions. We can explore other models such as convolutional or recurrent neural networks. Another approach for future research is to look into using deep learning techniques for the reconstruction algorithm. Using similar deep learning techniques like the ones shown in this work, we can study regression techniques that can accurately predict initial energy values for individual gamma rays. Improving the estimated initial energy values with the regression model will result in more accurate image reconstruction, which in practice can reduce the amount of uncertainty of the region influenced via the proton beam.

ACKNOWLEDGMENT

This work is supported by the National Institutes of Health National Cancer Institute under award number R01CA187416. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF). The facility is supported by the U.S. National Science Foundation through the MRI program (grant nos. CNS–0821258, CNS–1228778, and OAC–1726023) and the SREEMS program (grant no. DMS–0821311), with additional substantial support from the University of Maryland, Baltimore County (UMBC). See https://hpcf.umbc.edu for more information on HPCF and the projects using its resources.

REFERENCES