Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

Check for updates

Investigating particle track topology for range telescopes in particle radiography using convolutional neural networks

Helge Egil Seime Pettersen[a] , Max Aehle[b], Johan Alme[c] , Gergely Gábor Barnaföldi[d] , Vyacheslav Borshchov[e] , Anthony van den Brink[f] , Mamdouh Chaar[c], Viljar Eikeland[c], Grigory Feofilov[g], Christoph Garth[h] , Nicolas R. Gauger[b] , Georgi Genov[c] , Ola Grøttvik[c] , Håvard Helstrup[i] , Sergey Igolkin[g], Ralf Keidel[j] , Chinorat Kobdaj[k] , Tobias Kortus[j] , Viktor Leonhardt[h] , Shruti Mehendale[c], Raju Ningappa Mulawade[j] , Odd Harald Odland[a,c] , Gábor Papp[l] , Thomas Peitzmann[f] , Pierluigi Piersimoni[c] , Maksym Protsenko[e] , Attiq Ur Rehman[c], Matthias Richter[m], Joshua Santana[j], Alexander Schilling[j] , Joao Seco[n,o] , Arnon Songmoolnak[c,k], Jarle Rambo Sølie[p] , Ganesh Tambave[c] , Ihor Tymchuk[e] , Kjetil Ullaland[c] , Monika Varga-Kofarago[d] , Lennart Volz[q,r] , Boris Wagner[c], Steffen Wendzel[j] , Alexander Wiebel[j] , RenZheng Xiao[c,s], Shiming Yang[c], Hiroki Yokoyama[f], Sebastian Zillien[j] and Dieter Röhrich[c]

aDepartment of Oncology and Medical Physics, Haukeland University Hospital, Bergen, Norway; bChair for Scientific Computing, Technische Universität Kaiserslautern, Kaiserslautern, Germany; cDepartment of Physics and Technology, University of Bergen, Bergen, Norway; dWigner Research Centre for Physics, Budapest, Hungary; eResearch and Production Enterprise "LTU" (RPE LTU), Kharkiv, Ukraine; fInstitute for Subatomic Physics, Utrecht University/Nikhef, Utrecht, Netherlands; gDepartment of High Energy and Elementary Particles Physics, St. Petersburg University, St. Petersburg, Russia; hScientific Visualization Lab, Technische Universität Kaiserslautern, Kaiserslautern, Germany; iDepartment of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, Norway; jCenter for Technology and Transfer (ZTT), University of Applied Sciences Worms, Worms, Germany; kInstitute of Science, Suranaree University of Technology, Nakhon Ratchasima, Thailand; lInstitute for Physics, Eötvös Loránd University, Budapest, Hungary; mDepartment of Physics, University of Oslo, Oslo, Norway; nDepartment of Biomedical Physics in Radiation Oncology, DKFZ-German Cancer Research Center, Heidelberg, Germany; oDepartment of Physics and Astronomy, Heidelberg University, Heidelberg, Germany; pDepartment of Diagnostic Physics, Division of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway; qDepartment of Biophysics, GSI Helmholtz Center for Heavy Ion Research GmbH, Darmstadt, Germany; rDepartment of Medical Physics and Biomedical Engineering, University College London, London, UK; sCollege of Mechanical & Power Engineering, China Three Gorges University, Yichang, People's Republic of China

ABSTRACT
Background: Proton computed tomography (pCT) and radiography (pRad) are proposed modalities for improved treatment plan accuracy and *in situ* treatment validation in proton therapy. The pCT system of the Bergen pCT collaboration is able to handle very high particle intensities by means of track reconstruction. However, incorrectly reconstructed and secondary tracks degrade the image quality. We have investigated whether a convolutional neural network (CNN)-based filter is able to improve the image quality.
Material and methods: The CNN was trained by simulation and reconstruction of tens of millions of proton and helium tracks. The CNN filter was then compared to simple energy loss threshold methods using the Area Under the Receiver Operating Characteristics curve (AUROC), and by comparing the image quality and Water Equivalent Path Length (WEPL) error of proton and helium radiographs filtered with the same methods.
Results: The CNN method led to a considerable improvement of the AUROC, from 74.3% to 97.5% with protons and from 94.2% to 99.5% with helium. The CNN filtering reduced the WEPL error in the helium radiograph from 1.03 mm to 0.93 mm while no improvement was seen in the CNN filtered pRads.
Conclusion: The CNN improved the filtering of proton and helium tracks. Only in the helium radiograph did this lead to improved image quality.

Background

Proton computed tomography

The current worldwide expansion of proton therapy (PT) is being met by demands of increased diagnostic accuracy in treatment planning and treatment validation [1,2].

One of the proposed modalities that could increase the PT treatment accuracy is proton CT (pCT)/proton radiography (pRad). pCT has been shown to reduce the systematic errors of the relative stopping power (RSP) needed for treatment planning [3,4]. Furthermore, a sufficiently quick pRad system could be used as beam's eye view imaging prior to

---

treatment for accurate range verification through the Water Equivalent Path Length (WEPL) map [5].

A pCT setup needs to be able to measure the position and direction of particles going into and out from the patient, as well as the residual energy/range of the same particles. With these measurements, the energy loss along the particle's trajectory can be estimated using so-called most likely path methods [6,7], before using image reconstruction to calculate the patient's RSP map.

Several design approaches for pCT and pRad have been proposed [8–12], with different technologies and techniques for the residual energy and positional measurement.

Machine learning

Deep learning has lately achieved a large amount of success in tasks such as computer vision and speech recognition, with convolutional neural networks (CNN) introduced to computer vision by LeCun et al. [13] being of particular significance.

Common setups for CNNs for classification are a series of convolutional layers followed by a series of fully connected layers or a single global average pooling operation across the filter dimensions, which then returns a desired output.

There have been machine learning approaches to pCT-related problems in the past, such as proton path reconstruction [14], detector calibration [15] and prototype development [16]. One of the benefits of machine learning approaches is the time efficiency, when compared to previous analytical methods.

The aim for this study was to evaluate a CNN network trained to classify the track quality of both proton and helium beams, the goal being increased image quality.

Material and methods

Digital tracking calorimeter

The Bergen pCT collaboration is currently developing a Digital Tracking Calorimeter (DTC), which is a particle-tracking range telescope. The DTC consists of 43 layers (sensitive area $27 \times 16.6\,cm^2$) of high-granularity ($29\,\mu m$ pitch) pixel sensors interleaved by 3.5 mm aluminum slabs for slowing down the particles. Thus a 230 MeV/u proton (or helium ion) beam is fully stopped in the DTC with its residual range determined by the energy loss inside the imaged object. The technology behind the DTC is described in Alme et al. [17]. The pixelated design allows for handling multiple simultaneous tracks, enabling intensities of 5–20 million particles/second and thus sub-second pRad: implemented in a clinical workflow, the DTC could prove an accurate and fast RSP map and range verification device.

A track reconstruction procedure is used to disentangle 50–200 simultaneous particle trajectories. These tracks contribute to information about the average energy loss along the particles' trajectories inside the patient, so an accurate representation of the tracks entering the DTC is vital. A track reconstruction algorithm has been demonstrated for both protons and helium ions [9,18–20]. However, a fraction of the tracks are incorrectly or incompletely reconstructed due to high particle density, large-angle scattering and secondary particles. These tracks contribute to a systematic range error (incompletely reconstructed tracks) and noise (tracks not following the same particle along its path and forked paths from secondary production). Thus, a robust filtering method is needed to reduce the potential image quality degradation.

Monte Carlo simulations

The simulations have been performed using GATE 8.2 [21,22,23] and Geant4 10.5.1 [24,25]. The geometry and beam properties were described in Pettersen et al. [20]. A calibrated detector response model is applied to smear out the hits into multi-pixel clusters and to subsequently estimate the energy loss of the tracks at each sensor chip layer [26].

To generate training tracks with a sufficiently wide range of particle energies, ensuring that the CNN was trained without bias with respect to track length, a wedge-shaped water phantom with thickness 0–300 mm was used.

Additionally, to consider the effects of the filtering on proton and helium radiography (HeRad) image quality, a pediatric head phantom was implemented: see Pettersen et al. [20] for more details.

As an upper limit to the image quality, a set of simulations were performed without nuclear interactions.

Track classification

The classification procedure for determining the track quality was based on their fitted range: tracks within 2 cm of the ground truth residual range were labeled as 'good', and vice versa for 'bad' tracks. This ensured that 'good' tracks generally contributed toward increased image quality. In total, 82% (62%) of the proton (helium) tracks were labeled as 'good' after reconstruction. See Figure A-1 (Supplementary Materials) for examples of 'good' and 'bad' tracks.

Generally, the 'good' tracks contained better defined Bragg peaks compared to the 'bad' tracks. This can be explained by higher-energy tracks that were incompletely reconstructed. Both of the 'good' datasets have a common energy deposition plateau value, higher for helium than for protons. However, in the helium 'bad' dataset, a number of tracks exhibit a sudden fall to a lower plateau value due to secondary production, e.g., the track following a ^{4}He suddenly follows a proton instead. There was also a certain amount of fluctuation in the energy loss due to its statistical nature, and in some cases two close tracks were registered together (the smeared pixel shapes merged into a larger cluster).

Filter training and validation

The convolutional neural network received a vector containing 43 values, representing the deposited energy in each detector layer. This input was passed through four

convolutional layers with 32, 64, 64, and 128 filters respectively. All filters were of length 3. Then, two fully connected layers containing 256 units received the flattened output of the convolutional layers. Finally, the network gave the probability for a 'good' track. A rectified linear unit was chosen for all activations. The network was implemented in PyTorch [27].

The training datasets contained 23.4 million helium tracks and 31.3 million proton tracks. Both underwent an 80/20 training/validation split. The network was trained for 10 epochs with a batch size of 128 using the Adam optimizer [28]. To deal with the inherent class imbalance in the datasets, a weighted binary cross entropy function was used.

Two Nvidia® V100 GPUs were used, to a total training time of 2.5 h for the helium data and 3.2 h for the proton data.

In addition to the CNN filtering, a selection of energy loss threshold filters were included: the energy loss in the last layer, in the next-to-last, 3rd and 4th last layer as well as in the plateau region (the five first layers).

The area under the receiver operating characteristics curve (AUROC) was used to evaluate the CNN and energy loss filters on the validation dataset.

Particle radiographs

The head phantom was imaged using proton and helium ion beams. Track reconstruction and preliminary filtering was performed by removing tracks with incoming angle >45 mrad (3σ) and WEPL >280 mm. Tracks with >75% CNN score were used for image reconstruction. For the energy loss methods, protons were required to have >1 keV/μm in their 4th last layer, and helium ions to have >3.5 keV/μm in the last layer.

The applied image reconstruction algorithm is described in Sølie et al. [29]. In total 9.4 million protons were simulated through the head phantom, corresponding to a dose of 15 μGy [20]. For helium ions, 4.3 million primaries were simulated to a dose of 23.7 μGy. Due to the reduced scattering

and range straggling, fewer helium ions were needed compared to protons for a similar image quality [30].

Subtraction radiographs were generated between the digitized phantom and each reconstructed radiograph. They were evaluated based on the distribution (standard deviation) of the WEPL errors, both overall and in selected regions of interest (ROIs): see Figure A-2 of The Supplementary Materials.

Results

Track filtering

In general the CNN performed considerably better than the energy loss methods, as shown in Figure 1.

Mis-classifications were typically tracks undergoing nuclear interactions in the imaged object: these tracks emitted secondary particles with Bragg peaks, but with shorter ranges. This was especially true for the helium setup. Some 'good' tracks without apparent Bragg peaks were also removed.

In terms of sensitivity (specificity), using the same filter combination and thresholds as for the radiographs, the proton tracks yielded 95.5% (48.3%) using the energy loss method and 93.3% (92.2%) using the CNN method. The helium tracks yielded 97.4% (47.7%) using the energy loss method and 94.7% (98.3%) using the CNN method.

Reconstructed radiographs

The subtraction radiographs are shown in Figure A-3 (Supplementary Materials). There was a reduction of the sampling artifact in the CNN filtered HeRads. This depth-dependent oscillation of the WEPL error, especially prominent for HeRads, has been discussed previously [19,20] and is due to the 8 mm WEPL spacing between the sensor layers.

Generally, the CNN filtering leads to an image reconstruction speedup of 10%–20%, depending on the particle configuration (see the Supplementary Materials for details).
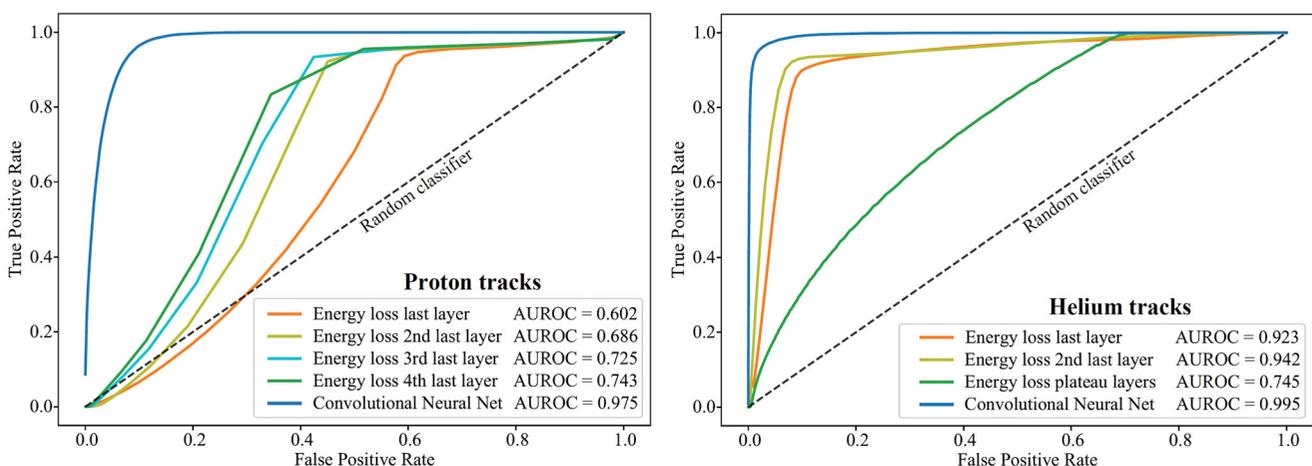
Figure 1. The receiver operating characteristics (ROC) curves for proton and helium ion tracks, using the proposed track filtering methods. The true positive rate is the proportion of 'good' tracks that were correctly identified, while the false positive rate is the proportion of 'bad' tracks that were incorrectly identified. The area under the ROC (AUROC) is also given for each scenario as a metric for model quality.
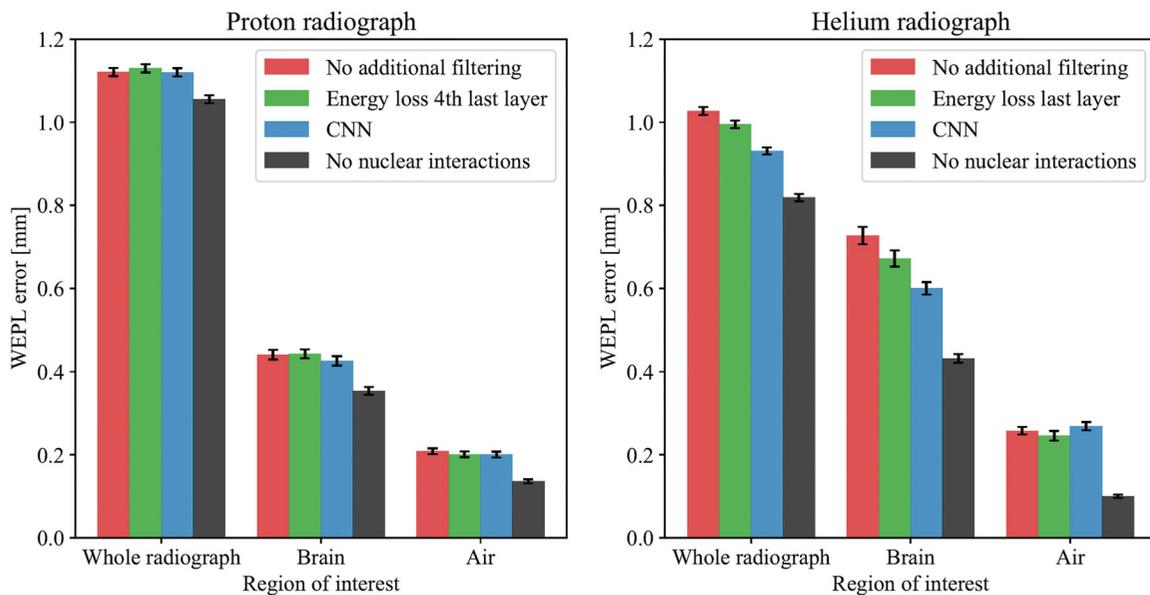
Figure 2. The water equivalent path length (WEPL) error of the three regions of interest (ROIs), calculated as the width of the distribution of the pixel-wise WEPL errors in the subtraction radiograph. The error bars showing the standard deviation were estimated using non-parametric bootstrapping with 10,000 iterations. Note that in the helium radiograph/brain ROI, the WEPL errors were larger due to the increased sampling artifact in that area.

WEPL errors

The standard deviations of the WEPL error distributions are summarized in Figure 2. No differences were seen in the proton case. For helium, the width in the overall radiograph was reduced from 1.03 mm WEPL to 0.93 mm WEPL using the CNN filter. Similar trends were also seen in the brain ROI, while no change was observed in the air ROI.

Discussion

The trained CNN exhibited a substantial improvement over simple energy loss filters when applied on a track-by-track basis, as evident in the AUROC improvements of 74.3%–97.5% for protons and 94.2% to 99.5% for helium, as well as in the specificity improvements. However, this improvement of track classification did not directly translate into improved image quality. One reason might be that the filtering commonly applied during image reconstruction already works well, where a WEPL filter removes a large fraction of 'bad' tracks with WEPL deviating from that of similar trajectories through the patient. Thus the negative effect of 'bad' tracks are already mitigated during the image reconstruction procedure, at least for pRad. The oscillating sampling artifact that degrades the WEPL accuracy, especially prominent in the HeRads, has been previously discussed [19,20]. The CNN filter was able to reduce this artifact, especially with HeRad, leading to improved WEPL accuracy.

The high AUROC values suggest that the classification properties of the network cannot be improved much. On the other hand, the design of the ground truth labeling might be a source of improvement. The applied label in this study was a rather simple one. However, during the exploratory phase of this study more complex classifications were designed: missing layers in the track (when compared to the MC truth), any secondary particles in the tracks and

confusion between track pairs. These attempts did not lead to improved AUROC scores or WEPL accuracy.

Future work would include filter implementation into the routine work flow, and coupling to tomographic reconstruction where the sampling artifact is an issue.

The presented classification can be compared to other helium imaging filtering studies. In the current work, the sensitivity (specificity) of helium classification was 94.7% (98.3%). In Pettersen et al. [20] a sensitivity (specificity) of 84.9% (97.5%) was found, and in Volz et al. [31] the sensitivity (specificity) was 84.5% (93.8%). Keeping in mind that these results reflect different imaging setups, methodologies and classification properties, this comparison is in favor of using deep learning methods for track classification.

Conclusion

In this study we trained and evaluated CNNs to classify 'good' and 'bad' tracks for proton and helium imaging, resulting from incorrect track reconstruction and secondary tracks. The goal was to improve the image quality of *in situ* particle therapy treatment verification, and ultimately, a step toward a safer and more effective particle therapy.

The presented results indicate that the CNN has very high discriminatory power for both proton and helium tracks when compared to previous methods. However, the full improvement was not carried over to the radiographic image quality due to the effective filtering inherent in the existing image reconstruction software. A modest WEPL error reduction from 1.03 mm to 0.93 mm was observed when applying the CNN filter on the HeRad: this corresponded to a visual improvement of a sampling artifact previously described. However, no improvement was observed for the pRad.

Additionally, using CNN led to a speed-up of the image reconstruction by a factor of 10%–20%.

Acknowledgements

Disclosure statement

Funding

ORCID

Helge Egil Seime Pettersen http://orcid.org/0000-0003-4879-771X
Johan Alme http://orcid.org/0000-0003-0177-0536
Gergely Gábor Barnaföldi http://orcid.org/0000-0001-9223-6480
Vyacheslav Borshchov http://orcid.org/0000-0002-5579-8932
Anthony van den Brink http://orcid.org/0000-0003-2366-7257
Christoph Garth http://orcid.org/0000-0003-1669-8549
Nicolas R. Gauger http://orcid.org/0000-0002-5863-7384
Georgi Genov http://orcid.org/0000-0002-6663-1433
Ola Grøttvik http://orcid.org/0000-0003-0761-7401
Håvard Helstrup http://orcid.org/0000-0002-9335-9076
Ralf Keidel http://orcid.org/0000-0002-1474-6191
Chinorat Kobdaj http://orcid.org/0000-0001-7296-5248
Tobias Kortus http://orcid.org/0000-0002-0987-8544
Raju Ningappa Mulawade http://orcid.org/0000-0002-0180-8517
Odd Harald Odland http://orcid.org/0000-0001-8024-8556
Gábor Papp http://orcid.org/0000-0001-5038-678X
Thomas Peitzmann http://orcid.org/0000-0002-7116-899X
Pierluigi Piersimoni http://orcid.org/0000-0002-8958-2179
Maksym Protsenko http://orcid.org/0000-0001-9313-1701
Alexander Schilling http://orcid.org/0000-0001-8802-3247
Joao Seco http://orcid.org/0000-0002-9458-2202
Jarle Rambo Sølie http://orcid.org/0000-0002-8327-8248
Ganesh Tambave http://orcid.org/0000-0001-7174-3379
Ihor Tymchuk http://orcid.org/0000-0002-6436-7253
Kjetil Ullaland http://orcid.org/0000-0002-0002-8834
Lennart Volz http://orcid.org/0000-0003-0441-4350
Steffen Wendzel http://orcid.org/0000-0002-1913-5912
Alexander Wiebel http://orcid.org/0000-0002-6583-3092
Sebastian Zillien http://orcid.org/0000-0003-3360-1251

References

[1] Paganetti H, Beltran C, Both S, et al. Roadmap: proton therapy physics & biology. Phys Med Biol. 2021;66(5):05RM01.

[2] Parodi K, Polf JC. In vivo range verification in particle therapy. Med Phys. 2018;45(11):e1036–e1050.

[3] Johnson RP. Review of medical radiography and tomography with proton beams. Rep Prog Phys. 2018;81(1):016701.

[4] Dedes G, Dickmann J, Niepel K, et al. Experimental comparison of proton CT and dual energy x-ray CT for relative stopping power estimation in proton therapy. Phys Med Biol. 2019;64(16):165002.

[5] Krah N, Patera V, Rit S, et al. Regularised patient-specific stopping power calibration for proton therapy planning based on proton radiographic images. Phys Med Biol. 2019;64(6):065008.

[6] Schulte RW, Penfold SN, Tafas JT, et al. A maximum likelihood proton path formalism for application in proton computed tomography. Med Phys. 2008;35(11):4849–4856.

[7] Collins-Fekete C-A, Volz L, Portillo SKN, et al. A theoretical framework to predict the most likely ion path in particle imaging. Phys Med Biol. 2017;62(5):1777–1790.

[8] Sadrozinski H-W, Johnson RP, Macafee S, et al. Development of a head scanner for proton ct. Nucl Instrum Methods Phys Res A. 2013;699:205–210.

[9] Pettersen HES, Alme J, Biegun A, et al. Proton tracking in a high-granularity Digital Tracking Calorimeter for proton CT purposes. Nucl Instrum Methods Phys Res, Sect A. 2017;860:51–61.

[10] Esposito M, Waltham C, Taylor JT, et al. Pravda: the first solid-state system for proton computed tomography. Phys Med. 2018;55:149–154.

[11] Gehrke T, Gallas R, Jäkel O, et al. Proof of principle of helium-beam radiography using silicon pixel detectors for energy deposition measurement, identification, and tracking of single ions. Med Phys. 2018;45(2):817–829.

[12] Miller C, Altoos B, DeJongh EA, et al. Reconstructed and real proton radiographs for image-guidance in proton beam therapy. J Radiat Oncol. 2019;8(1):97–101.

[13] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–444.

[14] Lazos D, Collins-Fekete C-AC, Bober M, et al. Machine learning for proton path tracking in proton computed tomography. Phys Med Biol. 2021;66(10):105013.

[15] Wesp P, Dickmann J, Hoyle B, et al. Machine learning based detector calibration to improve the accuracy of proton computed tomography. Poster Presented at PTCOG58, Manchester, UK; 2019

[16] Finneman GM, Meskell N, Caplice T, et al. Proton imaging with machine learning. Vol. 11595. In: Bosmans H, Zhao W, Yu L, editors. Medical imaging 2021: physics of medical imaging. Bellingham, WA: SPIE; 2021. p. 1338–1355.

[17] Alme J, Barnaföldi GG, Barthel R, et al. A high-granularity digital tracking calorimeter optimized for proton ct. Front Phys. 2020;8:568243.

[18] Pettersen HES, Meric I, Odland OH, et al. Proton tracking algorithm in a pixel based range telescope for proton computed tomography. Paper presented at: Connecting the Dots 2018. Seattle, WA. Available from: arXiv:2006.09751 [physics.med-ph].

[19] Pettersen HES, Alme J, Barnaföldi GG, et al. Design optimization of a pixel-based range telescope for proton computed tomography. Phys Med. 2019;63:87–97.

[20] Pettersen HES, Volz L, Sølie JR, et al. Helium radiography with a digital tracking calorimeter-a Monte Carlo study for secondary track rejection. Phys Med Biol. 2021;66(3):035004.

[21] Jan S, Santin G, Strul D, et al. GATE: a simulation toolkit for PET and SPECT. Phys Med Biol. 2004;49(19):4543–4561.

[22] Jan S, Benoit D, Becheva E, et al. GATE V6: a major enhancement of the GATE simulation platform enabling modelling of CT and radiotherapy. Phys Med Biol. 2011;56(4):881–901.

[23] Sarrut D, Bardiès M, Boussion N, et al. A review of the use and potential of the GATE Monte Carlo simulation code for radiation therapy and dosimetry applications. Med Phys. 2014;41(6):064301.

[24] Agostinelli S, Allison J, Amako K, et al. Geant4–a simulation toolkit. Nucl Instrum Methods Phys Res Sect A. 2003;506(3):250–303.

[25] Allison J, Amako K, Apostolakis J, et al. Recent developments in geant4. Nucl Instrum Methods Phys Res, Sect A. 2016;835:186–225.

[26] Tambave G, Alme J, Barnaföldi G, et al. Characterization of monolithic cmos pixel sensor chip with ion beams for application in particle computed tomography. Nucl Instrum Methods Phys Res Sect A. 2020;958:162626 (Proceedings of the Vienna Conference of Instrumentation 2019).

[27] Paszke A, Gross S, Massa F, et al. Pytorch: an imperative style, high-performance deep learning library. Vol. 32. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E, Garnett R,

editors. Advances in neural information processing systems. Red Hook, NY: Curran Associates, Inc.; 2019. p. 8024–8035.

[28] Kingma DP, Ba J. Adam: a method for stochastic optimization. 2017. arXiv:1412.6980.

[29] Sølie JR, Volz L, Pettersen HES, et al. Image quality of list-mode proton imaging without front trackers. Phys Med Biol. 2020; 65(13):135012.

[30] Hansen DC, Bassler N, Sørensen TS, et al. The image quality of ion computed tomography at clinical imaging dose levels. Med Phys. 2014;41(11):111908.

[31] Volz L, Piersimoni P, Johnson RP, et al. Improving single-event proton CT by removing nuclear interaction events within the energy/range detector. Phys Med Biol. 2019; 64(15):15NT01.