r-UNet: Leaf Position Reconstruction in Upstream Radiotherapy Verification

C. De Sio, J. J. Velthuis, L. Beck, J. L. Pritchard, and R. P. Hugtenburg

***Abstract*—Monolithic active pixel sensor (MAPS) devices are an effective tool for upstream verification of intensity-modulated radiotherapy (IMRT) treatments. It is crucial to measure with high precision the positions of the multi-leaf collimators (MLCs) used to shape the beam in real time, in order to enhance the quality and safety of treatments. This article describes r-UNet, a deep learning-based solution for leaf position reconstruction. The model is used to analyze the high-resolution images produced by a Lassena MAPS device in order to automatically determine the leaf positions. Image segmentation and leaf position estimation are performed simultaneously in a multitask setting. r-UNet obtained an average Dice coefficient of 0.96 ± 0.03 for the reconstructed image masks in the held-out test set; whilst the mean squared error (MSE) resulting from the estimation of the MLC positions is 0.003 mm, with a resolution ranging between 45 and 53 μm for leaf extensions between 1 and 35 mm. On unseen leaf positions, r-UNet yielded a single-leaf resolution between 54 and 88 μm depending on the leaf extension, and an average MSE of 0.07 mm. These results were obtained using single frames of data collected at 34 frames/s.**

***Index Terms*—Deep learning, image segmentation, monolithic active pixel sensors (MAPSs), multileaf collimator (MLC), position reconstruction, radiotherapy.**

I. Introduction

INTENSITY-MODULATED radiotherapy (IMRT) is a technique used to treat cancerous tumors by shaping the photon beam through multi-leaf collimators (MLCs) [1], in order to deliver a concentrated radiation dose to the target region, while preserving the surrounding healthy tissue. As treatments get more complex, it is crucial to measure the delivered dose and the position of the collimator components (i.e., tungsten leaves) with high accuracy. MLC position calibration is usually performed on a monthly basis, and given the precision of standard calibration methods [2], a calibration tolerance of ± 1 mm for the MLC leaf edge position is currently reached [3]. From the evaluation of the effects of MLC positioning errors on the delivered treatments, it has been suggested that the precision limit to maintain total dose errors below 2% is around 300 μm [4], [5]. As the calibrations are generally performed pretreatment, a high-precision, real-time verification device would improve the treatment monitoring and could allow for faults to be detected during the treatment delivery and immediately addressed. This article describes an approach to MLC position reconstruction based on artificial intelligence (AI) techniques. In particular, we define r-UNet, a UNet-based model that is able to simultaneously detect the leaves in the images produced by MAPS sensors, and to estimate their position in a multitask learning setting.

Manuscript received December 23, 2019; revised March 9, 2020; accepted April 17, 2020. Date of publication May 15, 2020; date of current version March 3, 2021. This work was supported by the Science and Technology Facilities Council under Grant ST/T002646/1. *(Corresponding author: C. De Sio.)*

C. De Sio, L. Beck, and J. L. Pritchard are with the School of Physics, H. H. Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL, U.K. (e-mail: chiara.desio@bristol.ac.uk).

J. J. Velthuis is with the School of Physics, University of Bristol, BS8 1TL Bristol, U.K., also with the School of Nuclear Science and Technology, University of South China, Hengyang 421001, China, and also with Medical School, Swansea University, Swansea SA2 8PP, U.K.

R. P. Hugtenburg is with Medical School, Swansea University, Swansea SA2 8PP, U.K., also with the Department of Medical Physics and Clinical Engineering, Abertawe Bro Morgannwg University Hospital Board, Swansea SA2 4QA, U.K., and also with the School of Physics, University of Bristol, Bristol BS8 1TL, U.K.

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRPMS.2020.2994648

II. Radiotherapy Verification

Several monitoring approaches exist to verify the radiotherapy treatment and measure the position of the MLC. The verification devices that are currently used can be divided into downstream and upstream detectors. The first is positioned behind the patient, and produce images that can be analyzed to reconstruct the delivered dose and the shape of the treatment received by the patient. The most common devices are radiographic films and electronic portal imaging devices (EPIDs). The produced images, also known as portal images, are affected by beam distortion and scattering due to the presence of the patient.

Upstream detectors are placed before the beam reaches the patient. The produced image is, therefore, more clear compared to the previous approach, but the key challenge is keeping the beam attenuation to a minimum.

An alternative approach is using the log files produced by the LINAC itself as a measure of the MLC positions [6]. Use of log files is technically not verification, as it does not take into account the possibility that the MLC monitoring system is incorrectly calibrated or misaligned. Monolithic active pixel sensor (MAPS) devices have been proven effective for upstream verification of radiotherapy treatments [7]–[9], as they can be made very thin and MAPS-based systems can be produced with an attenuation below 1% [10].

A. Monolithic Active Pixel Sensors

A MAPS device consists of three layers (Fig. 1). From bottom to top: a highly p-type doped substrate, referred to as

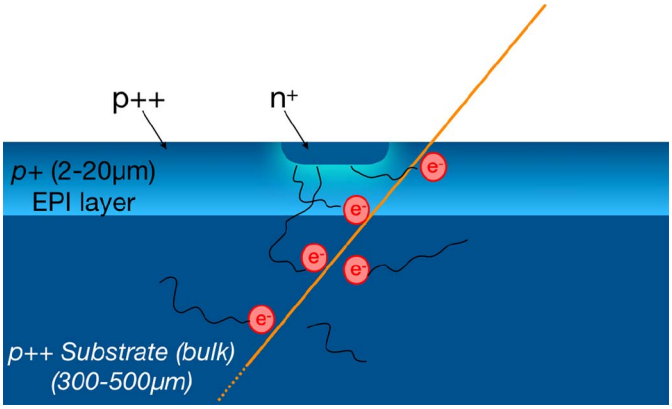


Fig. 1. Diagram of MAPS sensor. Incoming particles (photons) generate ionization and electron-hole pairs are created. Electrons propagating in the silicon are then collected in the *n*-well.

the bulk; a lower *p*-type doped epitaxial layer (EPI) of silicon; and a highly *p*-doped layer on top of the EPI layer. In the top layer, *n*-wells are created to collect the charge. The top layer also houses electronics, typically a single transistor where the first stage of amplification takes place. Sensors with more complicated circuits do exist.

When ionizing radiation propagates through the sensor, electron-hole pairs are generated. The electrons produced in the EPI layer are confined to that layer by the built-in potential due to the different p doping concentrations. These electrons diffuse through the EPI layer, and eventually reach the depleted zone underneath the diode, in which they are collected.

The top layer that houses the transistors is very thin, much less than a micron thick. The top of the substrate layer is needed to create the built-in potential difference. The rest of the bulk is only for mechanical support. The EPI layer, where signal generation occurs, is typically between 2 and 20-μm thick. The device can thus be thinned from the back, up to 30 μm or less, without significant loss of signal-to-noise. In a radiotherapy verification scenario, a sensor with this thickness would attenuate the beam by less than 0.1%.

Previously, with the Achilles sensor [11] which has an area of 6×6 cm^2 and a 15-μm pitch, a leaf edge resolution of 52 ± 4 μm at the iso-center has been achieved. This article was done using 0.1 s of data taken at 400 monitor units (M.Us.) per minute, for leaves with a width of 1 cm at isocenter [10]. The algorithm used to calculate the leaf profile was based on a Sobel filter. Leaf misplacements as small as 0.5 mm were detected and moving leaves were tracked [12].

In this article, we use the Lassena sensor [13]: a 12×14 cm^2, $3T$ sensor with 50-μm pitch. Despite having larger pixel sizes, leading to a lower resolution, this device is 3-side buttable, which allows the coverage of large areas by tiling sensors in a $2 \times N$ configuration, without significant dead space in between. For example, a 2×2 matrix of these sensors can cover a large enough area to verify a radiotherapy treatment with field as large as 30×30 cm, thus it could be clinically deployed in the future. Moreover, the sensor can be operated at a frame rate of 34 frames/s.

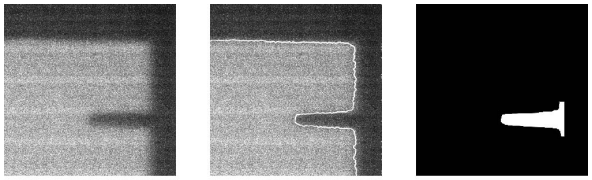
III. Artificial Intelligence for Radiotherapy Verification

Nowadays, machine learning (ML) is used for several applications in medicine, medical imaging, and radiotherapy [14], [15]. In particular, different ML-based solutions have been defined in the literature for radiotherapy verification. For example, volumetric dose prediction [16] was performed using fully convolutional neural networks (FCNNs) to predict new dose distributions from the analysis of existing treatment plans; portal images are used as input data to assess the quality of radiotherapy treatments, and classify the delivered treatment as acceptable or not acceptable, compared to the treatment plans [17]. ML models are also used to estimate the position of the MLC, using the positions reported in the LINAC log files as ground truth, as well as estimating the delivered dose [18].

In this article, images obtained using a Lassena sensor are used as input to an FCNN model that is trained to simultaneously detect the leaf in the sensor image, and extract its position (i.e., displacement from the leaf bank hosting the components of the MLC). From an ML perspective, this can be seen as the combination of an image segmentation problem and a regression problem. Once trained, the model produces a multiple output consisting of the segmented images and the estimated positions, inferred directly from corresponding input sensor images. The aim of this article is to achieve the best possible position resolution with the least possible amount of radiation, in order to be able to provide an early intervention. We are exploring ML techniques for their power in working with large datasets and providing a fast response (after the training process has been completed).

A. Data Description and Experimental Setup

The Lassena device produces 2800×2400 pixel images, showing the position of the MLC leaves in terms of attenuation of the radiotherapy beam. Different regions are associated with different pixel intensities. A series of measurements was recorded keeping the sensor fixed in one position and moving a single central leaf of the MLC forward from the leaf bank in a 10×10 cm^2 square field. The dataset used in this article is composed of approximately 9000 images. For each considered leaf position (i.e., 1, 2, 3, 4, 10, 15, 20, 25, 30, and 35 mm) a sample of 900 consecutive frames was acquired with our sensor. The 900 frames correspond to $\sim$ 30 s of data per position, acquired in 3 consecutive runs of 300 frames. During the acquisition, the MLC configuration and the detector were kept still. To avoid overfitting, as well as to test the generalization capabilities of the model, two separate evaluation protocols have been adopted in this article. 1) To test the ability of the model to correctly estimate the position of a leaf from a single image, all the images corresponding to 4 (out of 10) positions, namely, 1, 3, 15, and 30 mm, have been used only for testing. In the rest of this article, this dataset will be referred to as "external unseen dataset." None of these images, as well as their corresponding leaf positions, have been used during the training of the neural network. The aim is to evaluate the ability of the model to generate correct estimations for "new"



(a) (b) (c)

Fig. 2. (a) Input image. (b) Contour selected by the Chen-Vese algorithm, superimposed to the input image. (c) Ground truth binary mask used for training. The shown images refer to a leaf displacement of 25 mm.

leaf positions never seen before, emphasising the advantages of the regression settings, over a more simplistic classification. 2) The data for the remaining six positions (i.e., 2, 4, 10, 20, 25, and 35 mm, corresponding to 5400 images) are further split into a training set and a "held-out" test set, accounting for the 80% and the 20% of this data set, respectively (i.e., 4320 and 1080 images). It is worth noting that the selection of the frames during the split is completely randomized (i.e., refers to randomly selected frames within the 30 s interval) so to avoid any possible bias in the selection of training and evaluation data. The held-out test set is kept completely unseen, whereas an additional 20% of the training set is further used for internal validation during the training process (i.e., 864 out of 4320 samples). This internal validation set is used to analyze the evolution of the loss during the training, and to monitor possible cases of overfitting. The random selection algorithm used to sample images to account for the internal validation set is the same one used for the held-out test set generation. Stratification is used in the splitting, to keep the proportions of each leaf position in the three subsets and reduce biases in the training. A fixed seed is set in the random data splitting, for reproducibility purposes. The single leaf images used in this article have been collected at the Singleton Hospital in Swansea, using an Elekta Synergy LINAC to deliver a user-defined treatment at 400 M.U./min, in a 6-MV photon field. The MLC system is an Elekta Agility, featuring 160 tungsten leaves (2×80) with 5 mm width at iso-center. The sensor was operated at 34 fps, collecting approximately 30 s of treatment per leaf position.

B. Ground Truth Masks Generation

Binary masks are used as ground truth for the segmentation task. The mask generation steps are illustrated in Fig. 2. Input images (a) were initially processed to identify the contour of the leaf. The Chan-Vese segmentation algorithm [19] was used, as available in the scikit-image library [20]. From the obtained profile (b), the portion of the image containing the leaf was selected, and plotted on a black background, using the "fill_between" method from the matplotlib library [21]. This plot, showing a white leaf on a black background, is the binary mask (c). The mask is generated with the same size of the input image, so that it can be superimposed to filter the leaf as a segmentation mask. We considered single-leaf images due to the complexity in the generation of the ground truth masks.

IV. Model and Training

As the problem of identifying leaves in the sensor images requires the localization of objects in the input data, a neural network model inspired by UNet [22] was defined. It has been shown that UNet-based models are an effective technique for medical image segmentation (see [15], [23]). Our r-UNet model extends the classic UNet, including regression layers to estimate the MLC positions. This approach is end-to-end and multitask, as the two estimations are produced simultaneously and directly from the single input images. The main advantage of this learning strategy is that the model is able to identify the leaves directly from raw data, and calculate their position without requiring any manual intervention (e.g., manually selecting the leaves). A schematic representation of r-UNet is shown in Fig. 3.

r-UNet is composed of a contracting path and an expansive path; its structure allows it to learn a segmentation of the input image, and to produce an output image of the same size (typically a binary mask), isolating the target structure. It is composed of convolutional blocks, each composed of two 2-D Convolutions, each followed by a rectified linear unit (ReLU) [24], with increasing or decreasing feature size depending on the path. The contraction is implemented by adding MaxPooling [25] downsampling layers after each convolutional block. Each convolutional block has double the number of feature channels of the previous one. Differently from the original UNet model, r-UNet uses smaller feature map sizes (i.e., 16, 32, 64, 128, and 256). These sizes have been chosen experimentally, considering a tradeoff between performance and computation time.

Each up-sampling step is performed with 2-D transposed convolutions, and the number of feature channels is halved in the successive layers. Zero-Padding [25] is used to match the size of consecutive feature maps.

The two paths are also connected horizontally, concatenating every last layer of the convolutional blocks in the contracting path to the first layer of the corresponding block in the expansive path. This allows for the learnt features to be shared during the training of the model.

A fully connected layer with sigmoid activation is added at the end of the contracting path to enable the MLC position estimation. The leaf position regression is performed considering the same feature embedding used in the expanding path to generate the output binary masks.

r-UNet is trained using the sensor images as input data, along with the binary masks containing the corresponding leaf region as ground truth for the segmentation, and the LINAC set leaf positions as target labels for the regression. Input images are converted to 8-bit, cropped, and rescaled by a factor 0.25 (size reduced to 350×350 pixels), to ease the processing and reduce the memory occupation. For each input image, r-UNet produces the pair (segmented image and estimated position). The performance for both tasks is evaluated separately and combined.

A. Loss Functions

The r-UNet is trained on the 3456 images of the training dataset, using a combined loss function. The Dice-Sørensen

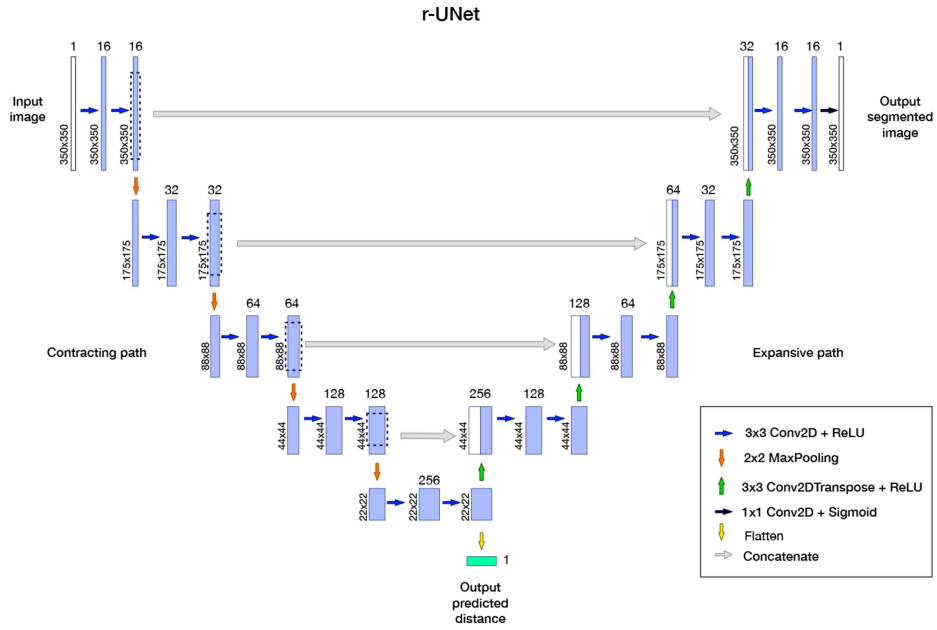


Fig. 3. r-UNet, the multitask FCNN model, composed of a contracting path (left-hand side), an expansive path (right-hand side), and additional regression layers (light green), to perform image segmentation and position estimation simultaneously. The dashed blue lines are used to picture the downsampling performed by pooling layers.

Coefficient [23], [26], [27] is used to optimize the leaf detection, whereas the leaf position regression is optimized by calculating the mean-squared error (MSE) between target and predicted distances. The Dice coefficient is a metric used to estimate the similarity (spatial overlap) between two sample sets (i.e., X and Y) [26]. It is defined as

$$\mathrm{DSC} = \frac{2|X \cap Y|}{|X| + |Y|}. \tag{1}$$

The Dice loss is expressed as $1 - \mathrm{DSC}$ to be used as loss function.

The resulting combined loss is defined as

$$L = \alpha \cdot \text{Dice Loss} + (1 - \alpha) \cdot \mathrm{MSE}. \tag{2}$$

The Adam optimizer has been used, with learning rate $lr = 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, eps $= 10^{-8}$ and no weight decay [28]. Different values for the α coefficient have been tested in the range $0.25 - 0.75$ with a 0.05 step, and performance have been evaluated for each trained model. The evaluation of the *best model* was conducted by considering the value of the combined loss on validation data, as well as the absolute difference between the average distance estimation on test data and on external unseen data. The best estimations were obtained for $\alpha = 0.4$, corresponding to $L = 0.4 \cdot \text{Dice Loss} + 0.6 \cdot \mathrm{MSE}$. This particular choice allows both problems to be optimized, whereas other values led to model overfitting, e.g., low MSE in distance prediction but no (or poor) binary mask generated, or good quality mask but bad distance estimation. For example, for $\alpha = 0.75$ the leaf segmentation (Dice Loss) reached a plateau almost immediately (see Fig. 4). In this case no masks were generated.

On the other hand, although the average MSE on the held-out set seemed promising (MSE $= 0.005$ mm), the same model, tested on unseen positions led to an MSE $= 0.809$ mm.

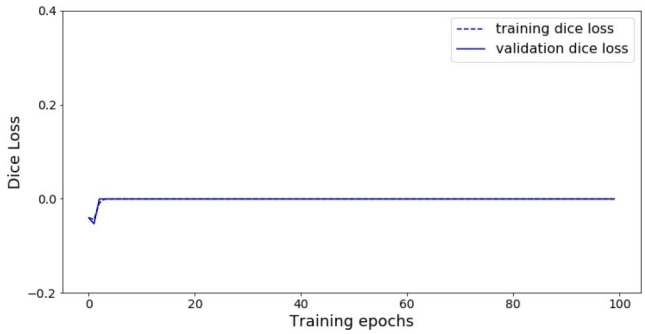


Fig. 4. Evolution of the Dice Loss for $\alpha = 0.75$. The loss function reaches a plateau around 0.0, indicating that the model is not learning.

Thus, the model was clearly overfitting the training positions. This further confirms the benefits of the multitask method, in which the two tasks are bounded to be learnt simultaneously; i.e., the leaf position estimation is constrained by its corresponding leaf segmentation.

The best model was trained for 100 epochs, using an early stopping criterion. The evolution of the combined loss function across the training and validation phases is reported in Fig. 5, along with the evolutions of the Dice Loss and MSE. These two plots are shown separately due to difference in the corresponding scales. The peak observed in the Dice Loss evolution plot around epoch 55 may be due to the fact that the optimizer yielded an update to the learning rate to skip over potential saddle points. As can be seen from the figures, the training converges very quickly to a stable solution for the two learning tasks, with this coefficient value. To validate the conclusion, several starting points have been experimentally tested, leading to the same best model.

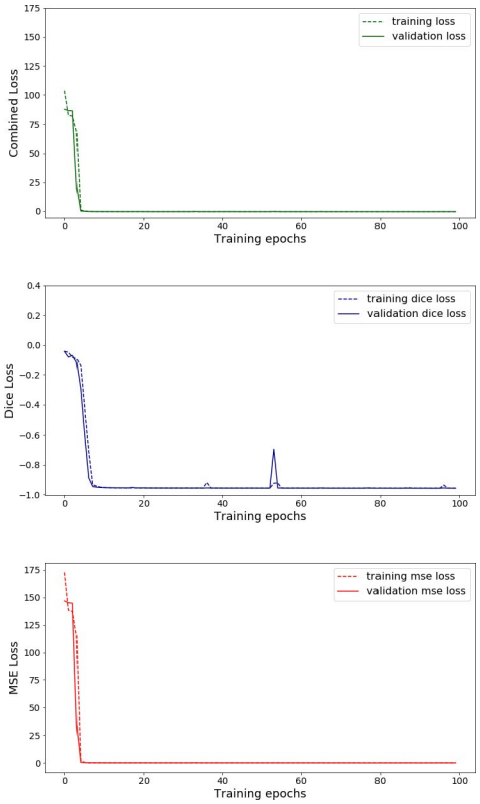


Fig. 5. Evolution of the loss functions during training (dashed lines) and validation (solid lines) of the best model ($\alpha = 0.4$). From top to bottom: combined loss, according to 2; Dice Loss; MSE loss.

B. Model Implementation

The model was implemented in Python using the PyTorch framework [29]. It was trained on a NVIDIA GeForce Titan Xp, producing a memory footprint of 2.5 GB, and performing 1 epoch in 4 min on the train dataset (composed by 3456 images in training and 864 in validation). The inference time was 0.17 s per image (1080 in the test dataset and 3532 images in the external unseen dataset).

V. Results

A. Performance on Test Dataset

The trained r-UNet model was tested on the 1080 images, corresponding to 20% of the total dataset, not used for training. The Dice coefficient resulting on test set images is 0.96 ± 0.03, whereas the average MSE on the same images is 0.003 mm. The distribution of the prediction error (evaluated as difference between estimated and true leaf extension) for the test dataset is shown in Fig. 6. Calculated as the σ of the Gaussian fit, the position resolution at iso-center for all test data is 50 ± 1 μm, ranging from 45 μm to 53 μm depending on the leaf extension, as shown in Fig. 9.

An example of the binary mask estimation from an input image as produced by the r-UNet is shown in Fig. 7.

B. Performance on External Unseen Dataset

To assess the generalization capability of the proposed approach, the model performance was tested on four samples

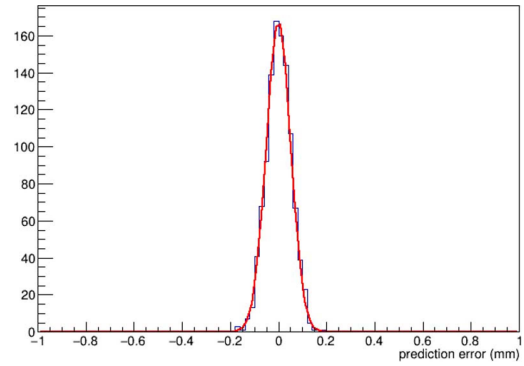
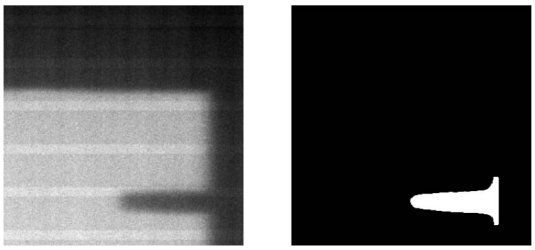


Fig. 6. Error distribution on all test dataset and gaussian fit.



(a) (b)

Fig. 7. (a) Input test image. (b) r-UNet-generated binary mask.

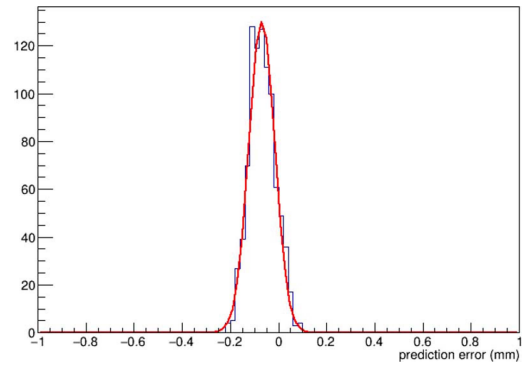


Fig. 8. Distribution of reconstructed position error for on a sample of unseen data (15 mm) and gaussian fit, yielding a reconstruction resolution of 54 μm and an MSE of 0.07 mm.

of unseen positions: 1, 3, 15, and 30 mm. Notably, r-UNet was not shown these positions in training. The obtained results are good: the average MSE ranges from 0.01 to 0.10 mm, depending on the positions. The distribution of the prediction errors for a sample of unseen images obtained for a leaf extension of 15 mm is shown in Fig. 8. A Gaussian fit yielded a reconstruction resolution of 54 μm and an MSE of 0.07 mm.

The single-leaf resolutions, calculated for each estimated position (held-out test and external unseen datasets) are shown in Fig. 9. Overall, the position resolutions range between 54 and 88 μm and are slightly worse in the case of unseen positions, but still below 100 μm even in the worst case (3 mm leaf extension). Furthermore, the leaf position resolution does not depend on the leaf extension.

These leaf position resolutions are far below the 300 μm precision required to limit dose errors to below 2%. Please

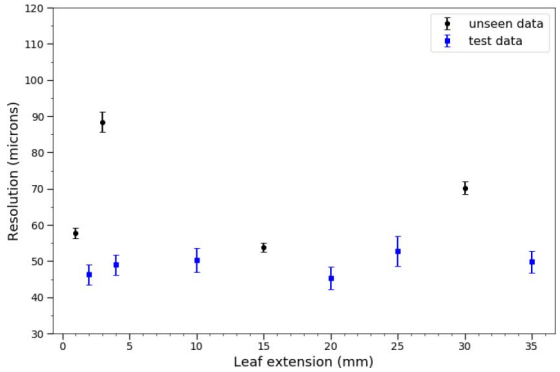


Fig. 9. Position resolution at iso-center for the test data (blue) and unseen data (black).

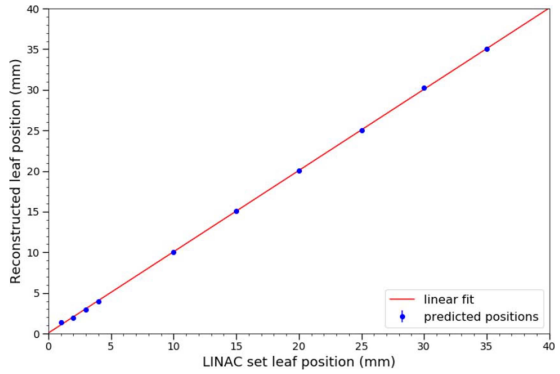


Fig. 10. Estimated versus set leaf positions, with linear fit compatible with $y = x$.

note that these results are achieved in single frames which are taken at 34 frames/s. The speed and precision allows for real-time intervention during treatments.

In Fig. 10, the estimated positions are plotted as a function of the LINAC set positions. A straight line fit resulted in a slope of 0.999 ± 0.004 and an intercept of 0.061 ± 0.032, which is compatible with the expected relation of $y = x$. Hence, the procedure yields the correct positions with excellent precision.

C. Uncertainty in the Reference Positions

The "intended" leaf positions, i.e., the MLC positions set in the Linac, were used as ground truth to train the network and evaluate the performance of the estimations. The data was collected in small consecutive sessions, to reduce the possibility of leaf repositioning by the machine. For each configuration (i.e., single leaf displacement), 3 runs of 300 frames were taken, corresponding to $\sim$10 s each, without changing the leaf position or moving the sensor between consecutive runs. The stability of the detected leaf positions was analyzed as a function of the frame number, to assess whether a (major) leaf repositioning occurred during the data taking.

The results are shown in Fig. 11 for the 15 mm case. The estimated positions appear to be flat throughout the whole $\sim$30-s data taking. This was confirmed by selecting a subset of 200 frames for each acquisition and calculating the residual distributions for each region (see Fig. 12). Finally, Fig. 13 shows that no significant repositioning took place

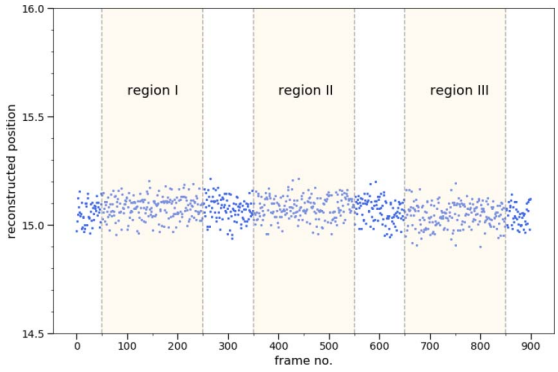


Fig. 11. Estimated position as a function of the frame number for the 15 mm case. The three highlighted regions indicate the central 200-frame subsets selected to study the stability.

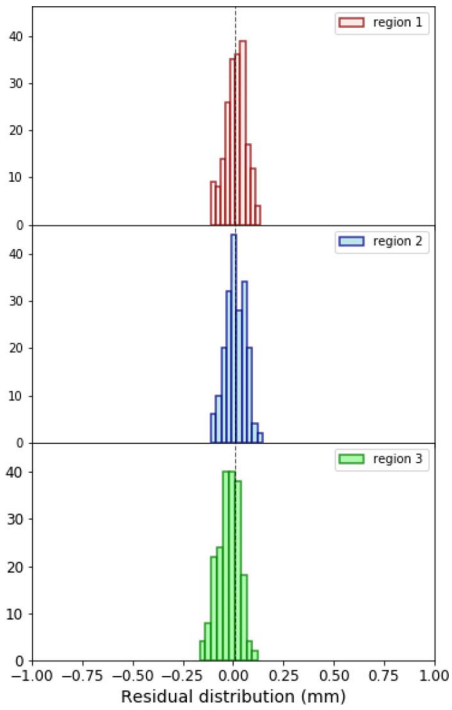


Fig. 12. Residual distributions for 15 mm, shown for the three regions separately. The corresponding resolutions are 51 μm, 53 μm, and 54 μm, respectively. The dashed black line, indicating the mean of the central distribution (blue, region 2), is shown for reference.

during our data taking: for each leaf position, the average estimated positions in the three regions are compatible with one another. Hence, the obtained resolution is a good measure of the uncertainty on the leaf position.

VI. Toward Multileaf Reconstruction for Radiotherapy Verification

The results shown in this article were obtained for a 6 MV flattened radiotherapy beam (FF) for a single central MLC leaf displaced forward in a 10×10 cm^2 photon square field. Single leaf resolutions were calculated for ten different displacements and the obtained results demonstrate the power of the novel approach, which is based on ML. However, the signal in the sensor, and thus the reconstructed leaf position, does not only

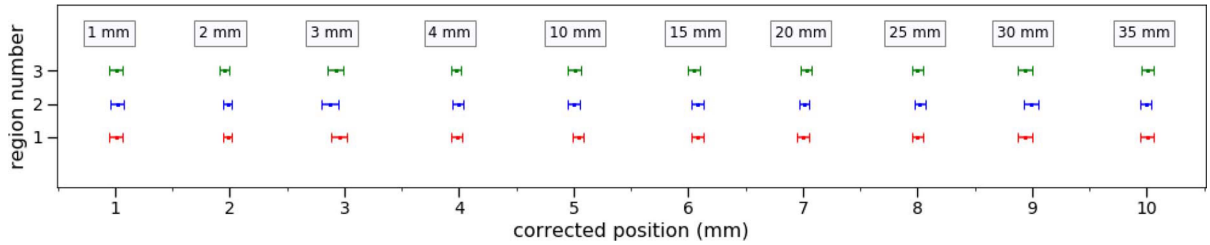


Fig. 13. Average of estimated position for single-leaf configuration, shown for each region, separately. The estimations in each consecutive data acquisition run are compatible with one another. Corrected (shifted) positions are shown to ease the visualization, with the corresponding leaf position in mm above.

depend on the position of the leaf under study but also on the position of its neighbors and the field configuration. Now that it has been proven that the approach presented here works very well for the simple configurations used here, we plan to extend this approach to more complicated scenarios and study the dependency of the leaf resolution on different field sizes, different beam energies, as well as using flattening filter free (FFF) beams. We will progress to the study of multiple leaf detection and complex radiotherapy treatments with moving leaves. In essence, leaf position reconstruction is about detecting the transition between the covered part of the sensor, which will be in the dark, and the uncovered part, where the signal will be high(er) due to the beam. The electron field smears this transition to a certain extent. Changing the field size results in a different contribution due to the electron field but as the rapid light-dark transition due to the partially blocked photon beam is still present, the algorithm will still detect the leaf position. The different electron field might lead to a different slope of this transition and thus to a different leaf position resolution, but the technique will still work and the effect should be minimal. Similarly, an FFF beam will mainly change the signal in the region where the sensor is not covered by the leaves as its amplitude will become dependent on the leaf position, but the rapid transition will still be present. Hence, we do not expect a major difference in leaf position resolution for different fields, beam energies and FFF beams. As for the detection of multiple leaf positions, we will use a region of interest approach. Once the regions of interest have been identified (image segmentation), the estimation will be performed for each identified leaf iteratively, taking into account the estimated positions for the neighboring leaves. Applying the technique to treatments with moving leaves like VMAT is more challenging, partly because it is difficult to define a reference position to compare to for a moving leaf. However, our sensor integrates images at 34 frames/s. Hence, the maximum movement between consecutive frames is about 1.47 mm for a Varian linac [30] and 1.91 mm for a Elekta Agility [31]. Since the detector integrates between readout cycles, the light-dark transition will be smeared out which will give information on the leaf position and average leaf speed. This information can be used in training to obtain better estimates. This will be the subject of further study.

VII. Conclusion

Now that IMRT is widely in use and with the novel trends toward more intense treatments, it becomes more and more important to independently and dynamically verify the positions of the MLC. Here, an ML algorithm is presented which reconstructs the leaf positions in real time with a precision well below 100 μm. All the estimations are obtained on a single frame basis, which corresponds to 0.03 s of treatment. The resolutions are well below the proposed limit of 300 μm, which in turn results in dosimetric errors below 2%. Multiple leaf position reconstruction is currently under development.

Acknowledgment

The Titan Xp used for this research was donated by the NVIDIA Corporation.

References

[1] M. Jeraj and V. Robar, "Multileaf collimator in radiotherapy," *Radiol. Oncol.*, vol. 38, no. 3, pp. 235–240, 2004.
[2] M. N. Graves, A. V. Thompson, M. K. Martel, D. L. McShan, and B. A. Fraass, "Calibration and quality assurance for rounded leaf' end MLC systems," *Med. Phys.*, vol. 28, no. 11, pp. 2227–2233, 2001.
[3] E. E. Klein *et al.*, "Task group 142 report: Quality assurance of medical accelerators A," *Med. Phys.*, vol. 36, no. 9, pp. 4197–4212, 2009.
[4] A. Rangel and P. Dunscombe, "Tolerances on MLC leaf position accuracy for IMRT delivery with a dynamic MLC," *Med. Phys.*, vol. 36, no. 7, pp. 3304–3309, 2009.
[5] A. Agarwal *et al.*, "Evaluating the dosimetric consequences of MLC leaf positioning errors in dynamic IMRT treatments," *J. Radiotherapy Pract.*, vol. 18, no. 3, pp. 1–7, 2019.
[6] A. Agnew, C. E. Agnew, M. W. D. Grattan, A. R. Hounsell, and C. K. McGarry, "Monitoring daily MLC positional errors using trajectory log files and EPID measurements for IMRT and VMAT deliveries," *Phys. Med. Biol.*, vol. 59, no. 9, pp. 49–63, 2014.
[7] J. J. Velthuis *et al.*, "The VANILLA sensor as a beam monitoring device for *X*-ray radiation therapy," *Appl. Radiat. Isotopes*, vol. 83, pp. 8–11, Jan. 2014.
[8] J. J. Velthuis, R. P. Hugtenburg, C. Hall, R. Page, and P. Stevens, "Monitoring of a therapeutic *X*-ray beam using an active pixel sensor detector," EU Patent EP2 654 895 B1, Mar. 2018.
[9] J. J. Velthuis, R. P. Hugtenburg, C. Hall, R. Page, and P. Stevens, "Upstream direct *x*-ray detection," U.S. Patent 9 517 358, 2016.
[10] R. F. Page *et al.*, "Using a monolithic active pixel sensor for monitoring multileaf collimator positions in intensity modulated radiotherapy," *IEEE Trans. Nucl. Sci.*, vol. 61, no. 1, pp. 74–78, Feb. 2014.
[11] N. Guerrini, R. Turchetta, G. Van Hoften, R. Henderson, G. McMullanc, and A. R. Faruqic, "A high frame rate, 16 million pixels, radiation hard CMOS sensor," *J. Instrum.*, vol. 6, pp. 1–10, Jun. 2010.
[12] L. Beck, J. J. Velthuis, S. Fletcher, J. A. Haynes, and R. Page, "Using a TRAPS upstream transmission detector to verify multileaf collimator positions during dynamic radiotherapy delivery," *Appl. Radiat. Isotopes*, vol. 156, Feb. 2020, Art. no. 108951.
[13] I. Sedgwick, D. Das, N. Guerrini, B. Marsh, and R. Turchetta, "LASSENA: A 6.7 MegaPixel, 3-sides buttable wafer-scale CMOS sensor using a novel grid-addressing architecture," in *Proc. Int. Image Sensor Workshop*, Snowbird, UT, USA, Jun. 2013.
[14] V. Weidlich and G. A. Weidlich, "Artificial intelligence in medicine and radiation oncology," *Cureus*, vol. 10, no. 4, 2018, Art. no. e2475.

[15] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[16] V. Kearney, J. W. Cham, G. Valdes, T. Solberg, and S. S. Yom, "The application of artificial intelligence in the IMRT planning process for head and neck cancer," *Oral Oncol.*, vol. 87, pp. 111–116, Dec. 2018.
[17] K. Leszczynski *et al.*, "Application of a fuzzy pattern classifier to decision making in portal verification of radiotherapy," *Phys. Med. Biol.*, vol. 44, no. 1, p. 253, 1999.
[18] J. N. K. Carlson, J. M. Park, S.-Y. Park, J. I. Park, Y. Choi, and S.-J. Ye, "A machine learning approach to the accurate prediction of multi-leaf collimator positional errors," *Phys. Med. Biol.*, vol. 61, p. 2514, May 2016.
[19] R. Cohen, "The Chan–Vese algorithm," Jul. 2011. [Online]. Available: arXiv:1107.2782.
[20] S. Van der Walt *et al.*, "Scikit-image: Image processing in Python," *PeerJ*, vol. 2, pp. 2–6, May 2014.
[21] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007.
[22] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.
[23] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. IEEE Int. Conf. 3D Vis.*, 2016, pp. 565–571.
[24] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, 2010, pp. 807–814.
[25] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," 2016. [Online]. Available: arxiv.abs/1603.07285
[26] K. H. Zou *et al.*, "Statistical validation of image segmentation quality based on a spatial overlap index," *Acad. Radiol.*, vol. 11, no. 2, pp. 178–189, 2004, doi: 10.1016/s1076-6332(03)00671-8.
[27] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, M. Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (Lecture Notes in Computer Science), vol. 10553, M. Cardoso *et al.*, Eds. Cham, Switzerland: Springer, 2017. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-67558-9_28
[28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014. [Online]. Available: arXiv:1412.6980.
[29] A. Paszke *et al.*, "Automatic differentiation in PyTorch," in *Proc. 31st Conf. Neural Inf. Process. Syst. (NIPS)* 2017, pp. 1–4.
[30] T. Y. Lim, I. Dragojević, D. Hoffman, E. Flores-Martinez, and G. Y. Kim, "Characterization of the HalcyonTM multileaf collimator system," *J. Appl. Clin. Med. Phys.*, vol. 20, pp. 106–114, Mar. 2019, doi: 10.1002/acm2.12568.
[31] V. P. Cosgrove *et al.*, "Physical characterization of a new concept design of an Elekta radiation head with integrated 160-leaf multi-leaf collimator," presented at the ASTRO, 2009.