Received: 31 July 2020 | Revised: 15 July 2021 | Accepted: 5 September 2021

DOI: 10.1002/mp.15229

RESEARCH ARTICLE

MEDICAL PHYSICS

Deep learning-augmented radioluminescence imaging for radiotherapy dose verification

Mengyu Jia | **Yong Yang** | **Yan Wu** | **Xiaomeng Li** | **Lei Xing** | **Lei Wang**

Department of Radiation Oncology, Stanford University, Stanford, California, USA

Correspondence
Lei Wang, Department of Radiation Oncology, Stanford University, Stanford, CA 94305, USA.
Email: leiwang@stanford.edu

Funding information
National Cancer Institute, Grant/Award Numbers: 1R01CA223667, 1R01CA227713

Abstract

Purpose: We developed a novel dose verification method using a camera-based radioluminescence imaging system (CRIS) combined with a deep learning-based signal processing technique.

Methods: The CRIS consists of a cylindrical chamber coated with scintillator material on the inner surface of the cylinder, coupled with a hemispherical mirror and a digital camera at the two ends. After training, the deep learning model is used for image-to-dose conversion to provide absolute dose prediction at multiple depths of a specific water phantom from a single CRIS image under the assumption of a good consistency between the TPS setting and actual beam energy. The model was trained using a set of captured radioluminescence images and the corresponding dose maps from the clinical treatment planning system (TPS) for the sake of acceptable data collection. To overcome the latent error and inconsistency that exists between the TPS calculation and the corresponding measurement, the model was trained in an unsupervised manner. Validation experiments were performed on five square fields (ranging from 2×2 to 10×10 cm^2) and three clinical intensity-modulated radiation therapy (IMRT) cases. The results were compared to the TPS calculations in terms of gamma index at 1.5, 5, and 10 cm depths.

Results: The mean 2%/2 mm gamma pass rates were 100% for square fields and 97.2% (range from 95.5% to 99.5%) for the IMRT fields. Further validations were performed by comparing the CRIS results with measurements on various regular fields. The results show a mean gamma pass rate of 91% (1%/1 mm) for cross-profiles and a mean percentage deviation of 1.15% for percentage depth doses (PDDs).

Conclusions: The system is capable of converting the irradiated radioluminescence image to corresponding water-based dose maps at multiple depths with a spatial resolution comparable to the TPS calculations.

KEYWORDS
deep learning, dosimetry, radioluminescence

1 | INTRODUCTION

The rapid development in precision radiation therapies such as intensity-modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT) dramatically increases the demands for accurate and efficient pretreatment dosimetric quality assurance (QA). The verification is typically implemented by comparing plan-calculated dose distributions or intensity maps, with those measured using QA devices, such as 2D array with diodes, ionization chambers, radiographic films, and so forth. Radiographic film dosimetry has the advantage of low introduction cost and high spatial resolution measurement but is a labor-intensive method. On-board electronic portal imaging device (EPID)-based dose verification has been popular in recent years benefiting from its wide availability, quick setup, and high image quality.[1] Limited by the on-board mounting manner, they

do not provide independent measurement that allows for end-to-end type verification on image-guided treatments, in which the QA device should be independent of treatment system and the QA procedure should follow the patent treatment procedure including simulation, plan creation, and plan delivery phases.[2] Detector array-based devices, such as ArcCHECK (Sun Nuclear Corporation, Melbourne, FL, USA), extend the independent measurements with coplanar detection ability to enable verifications on linac gantry angle.[3,4] However, the current pixel pitch (~0.7 cm for ArcCHECK) could be marginal to meet a stringent passing criterion, especially for small field measurements in stereotactic radiosurgery applications. Challenges still remain when good spatial resolution, digitalized acquisition, and independent measurement are jointly considered.

Camera-based radioluminescence imaging has been proposed as a low-cost yet promising alternative for QA applications.[5–8] In reality, however, radioluminescence imaging suffers from adverse influence of radioluminescence photon scattering, leading to blurred field edges and mirror-glare artifacts that make it difficult for accurate dosimetric measurements. Flat-field corrections are traditionally used to remove the artifacts, which, however, are highly dependent on beam energies and location in the detector's field of view. The relation between the radioluminescence image and the corresponding dose map is mathematically analogous to a convolution process.[1,9] In the past, various approaches were proposed in kernel-based approach. Most studies assumed that the dose kernel is spatially invariant and angularly isotropic when homogeneous medium is considered.[7,9,10] Brost and Watanabe advanced the field by developing region-based kernels calculated in sparsely partitioned subregions.[11] Alhazmi et al. proposed an EPID-based 3D dosimetry using a modified back-projection algorithm, where the volumetric dose was converted slice-by-slice from an input EPID image that was pre-proceeded with a series of rigorous calibrations.[12] Image-to-dose conversion was also implemented using neural networks. Liu et al. used a three-layer network to correct the dose profile measured by diode detector that suffered from the volume-averaging effect.[13] This method was further advanced by Cheon et al. to achieve 2D dose prediction from radioluminescence images using a shallow convolution neural network (CNN), which demonstrated the outperformance of using CNN to typical deconvolutions.[8] Nevertheless, conventional deep learning models could be problematic for certain dosimetric applications due to limited scalability of conventional approaches in handling more than two domains, which is desired (dose at different depths) in certain scenarios, for example, 3D dosimetry.[9] A deep learning model that relates image to multiple domains (depths) without going through the hassle of multiple training is highly desirable.

The purpose of this work is to develop a dose prediction strategy for a camera-based radioluminescence imaging system (CRIS) to circumvent the challenges in existing dosimetric tools as mentioned above. The CRIS innovatively involves a cylindrical sensing receptor to allow for a coplanar detection fashion similar to a volumetric dosimeter such as ArcCHECK, while maintaining a spatial resolution as high as that of an EPID (0.5 mm). The specific geometric design in turn complicates the dose response function both mathematically and physically in the following aspects: (i) diversity of incident beam angle on a curved surface; (ii) nonuniformity of optical coupling efficiency due to the perspective view and the vignetting effect; and (iii) complexity of physics behind the interreflections between the hemispheric mirror and the phosphor screen (i.e., the mirror-glare artifacts). The downstream task of traditional image-to-dose conversion would be thus challenged in terms of reliability, practicability, and robustness. Benefiting from the high nonlinearities in a CNN that enables to approximate the mapping physics with increased flexibility and accuracy, a data-driven learning model is established to effectively relate CRIS image to the desired absolute dose distribution in a specific water phantom, named as functional generative adversarial network (fGAN). The training ground-truths are those calculated in the treatment planning system (TPS), which facilitate the acquisition of high-resolution dose maps at multiple depths and therefore benefits for translational applications. In view of the latent error and inconsistency such as the field geometry, fGAN is trained in an unsupervised manner that originally aims at training with unpaired images.[14] Compared to simple pixel-wise mapping, the advantage of the unsupervised training is that it essentially enables learning on abstractive features (e.g., blurring behavior in the penumbra region) that are less sensitive to field geometry variations. The technique is applied to a series of regular fields and three clinical IMRT cases, and the results are compared with TPS counterparts.

2 | METHOD AND MATERIALS

2.1 | Design of radioluminescence imaging system

2.1.1 | Hardware design

Hardware design of the radioluminescence imaging system is shown in Figure 1. The inner surface of a 3D-printed cylindrical chamber is coated with a Gd_2O_2S-based scintillator material, which emits 545 nm light upon interaction with the megavoltage (MV) photons.[15–17] The radiation-induced light from the scintillator-coated layer is reflected by a hemispheric mirror mounted at far end of the cylinder, which provides a panoramic view of the chamber inner wall and

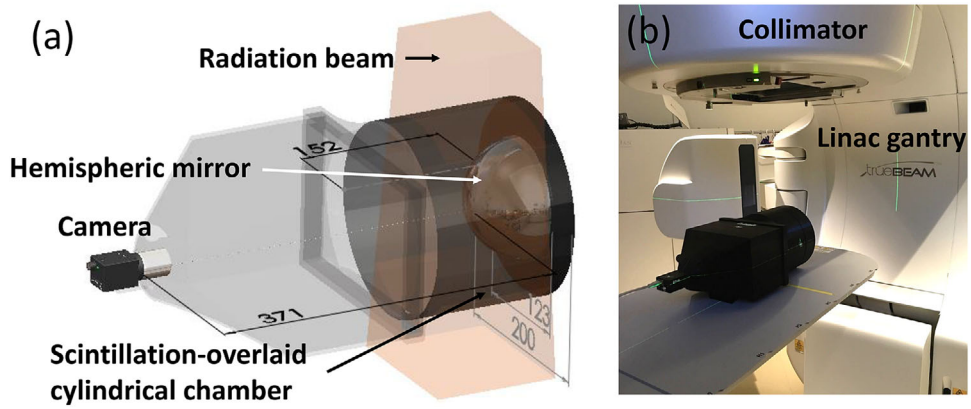

FIGURE 1 (a) Schematic diagram of the developed camera-based radioluminescence imaging system (CRIS) phantom, and (b) experimental setup. The system consists of a cylindrical receptor with its inner surface coated with scintillator, a hemispheric mirror mounted at the cylinder end, and a camera at the opposite end to capture the luminescence signals from the inner surface of the cylinder. The main dimensions are indicated in unit of millimeter

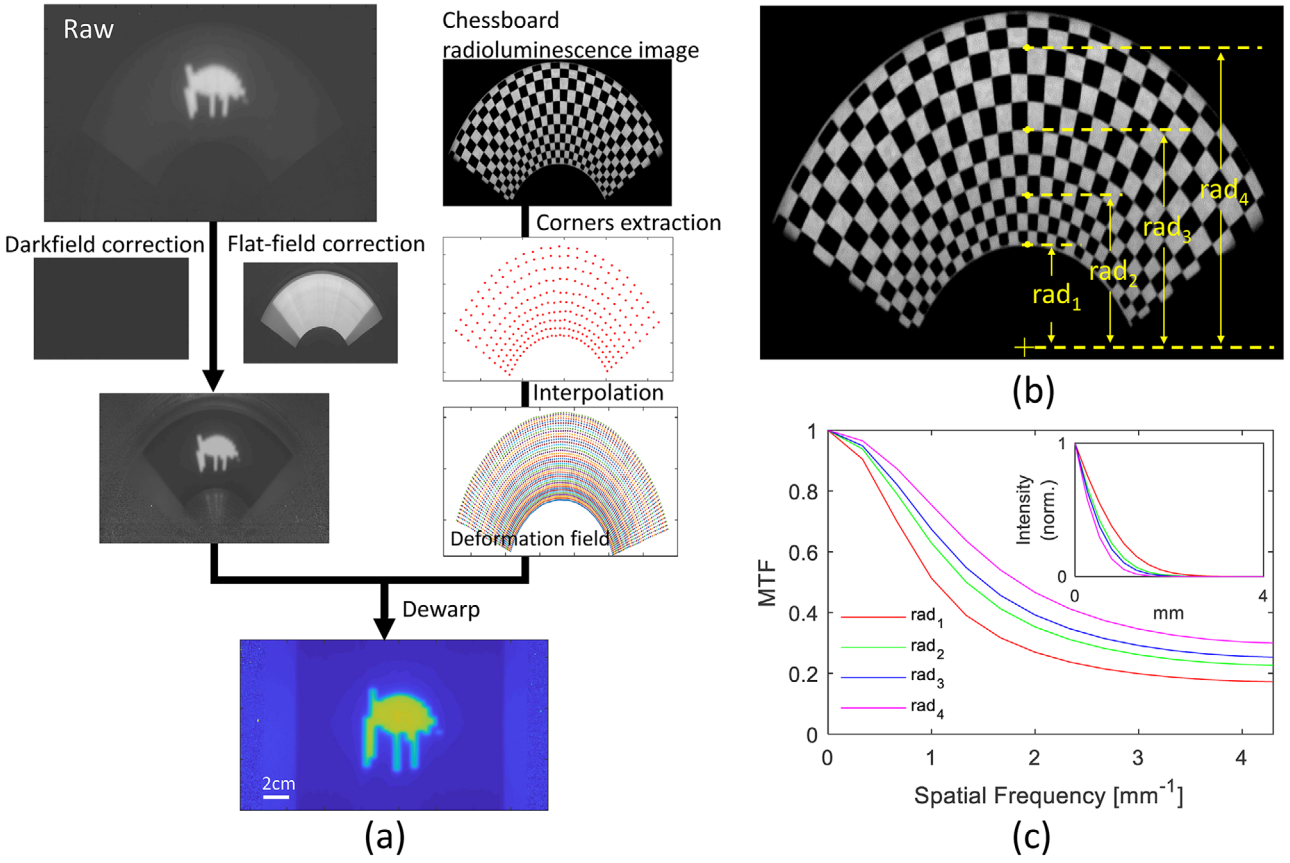

FIGURE 2 (a) Diagram showing the calibration procedure of camera-based radioluminescence imaging system (CRIS) image, (b) locations used to estimate the modulation transfer functions (MTFs), and (c) the system MTFs with the corresponding line spread functions (LSF) inserted

recorded by a CMOS camera with 1920 × 1200 pixel (GS3-PGE-23S6M-C, Point Grey Research, Inc., Richmond, Canada) mounted at the other end of the cylinder. By aligning the central axis of the hemispheric mirror to that of the cylinder, the CRIS provides a consistent beam's eye view for all the gantry angles.

2.1.2 | System calibration

A diagram showing the calibration process is presented in Figure 2a. The system was sequentially calibrated with dark-field and flat-field (a field of 15 × 15 cm^2 was delivered) corrections. To perform geometrical

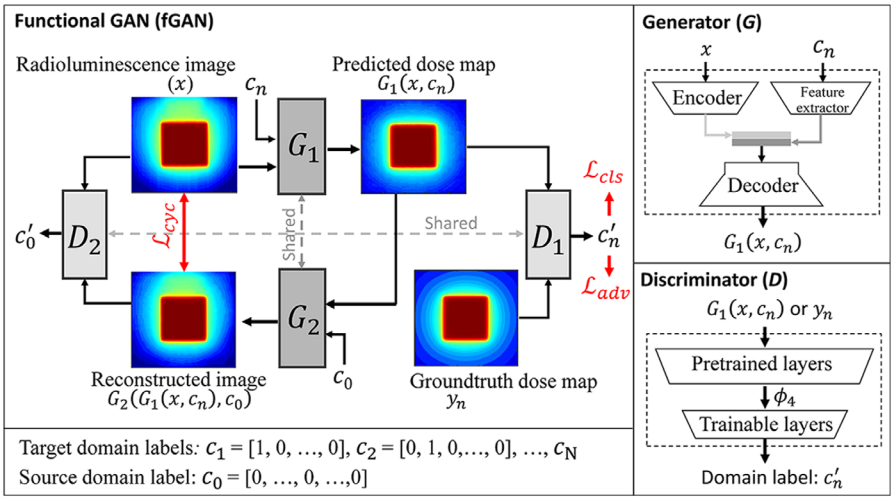

FIGURE 3 Architecture of the proposed fGAN, consisting of two identical generators (G_1 and G_2) and two identical discriminators (D_1 and D_2). G_1 takes input from both radioluminescence image x and a target domain label vector ($c_n|n > 0$) to a synthesize dose map at a desirable depth, that is, $G_1(x, c_n)$. By passing a predicted target image and a source domain label c_0 into G_2, the source image is reconstructed, that is, $G_2(G_1(x, c_n), c_0)$ and used to formulate a cycle consistency loss $\mathcal{L}_{cyc}$. The discriminators play dual roles: (i) a real/fake identifier, which contributes to $\mathcal{L}_{adv}$, and (ii) a domain classifier to evaluate the similarity between the predicted dose map and the ground truth at every depth, which contributes to $\mathcal{L}_{clc}$

restoration, a chessboard was overlaid to the scintillator sheet and the captured scintillation image was analyzed with OpenCV to automatically extract the corner points, which were then cubically interpolated to form a deformable field for nonrigid transformation. To quantify spatial frequencies that can be resolved versus locations at various radial distances (see Figure 2b), that is, $rad_1 = 30.8$ mm, $rad_2 = 44.9$ mm, $rad_3 = 64.1$ mm, and $rad_4 = 88.0$ mm, the calibrated chessboard image was analyzed to obtain modulation transfer functions (MTFs). Results are shown in Figure 2c, where the line spread functions (LSFs) are insert correspondingly in the subplot. Spatial resolutions achievable are related to the inverse of the MTF function where there is measurable amplitude, and here it was taken as 10% of the maximum. Results show that the system maintains spatial resolutions of 0.47, 0.38, 0.35, and 0.30 mm for locations at rad_1, rad_2, rad_3, and rad_4, respectively. Calibrated images still suffer from edge blurring, mirror-glare artifacts, and residual nonuniformity issues caused by light scattering, which will be mitigated using the deep learning model. To maintain reasonable signal-to-noise ratio and save the memory usage, the camera runs at five frames per second.

2.2 | Deep learning model for image–dose conversion

2.2.1 | Functional GAN (fGAN)

Figure 3 shows the pipeline of the proposed fGAN, which takes advantages of the powerful adversarial learning mechanism in the prediction of dose maps.[18]

While dose maps at different depths are deemed to belong to different image domains, our goal is to train a single network that learns a one-to-many domain translation. In brief, the architecture consists of two identical generators (G_1 and G_2) and two discriminators (D_1 and D_2). G_1 takes input from both radioluminescence image and target domain labels (c_n) to synthesize dose maps at multiple depths. By passing a predicted dose map and the corresponding source label (c_0) into G_2, a radioluminescence image is reconstructed and used to formulate a cycle consistency loss. In fact, an identical G is used for both G_1 and G_2. Meanwhile, the discriminators play dual roles: a real/fake identifier and a domain classifier. An identical discriminator (D) is employed for both D_1 and D_2.

Generator

Every generator (G_1 and G_2) is composed of three subnetworks, an encoder, a decoder, and a feature extractor. To enable dose prediction at a target depth, G_1 is trained to translate an input image x into an output dose map $G_1(x, c_n)$ conditioned on the target domain label c_n ($n > 0$), which is a one hot vector with a single bit identifying the nth depth. The domain label is integrated into the encoder–decoder framework via a feature extractor. Specifically, each element in the one hot vector is replicated to form a uniform matrix with the same size as the input x, yielding a matrix stack (with number of the stack layers equal to the vector length) imported into the feature extractor to elicit the abstractive information. Similarly, G_2 is trained to reconstruct the radioluminescence image $G_2(G_1(x, c_n), c_0)$ from the corresponding predicted dose map $G_1(x, c_n)$ and a source domain label c_0 using the same encode–decoder architecture.

Notice that a unified G is used for both G_2 and G_2, benifiting from the flexible translation ability of the network. A cycle consistency loss is achieved by comparing the reconstructed image and the original image. By importing images from various domains, the encoder learns domain-invariant features that are shared by all domains (e.g., the primary beam shape and content) as well as domain-specific features that are functions of depth. Subsequently, with the target depth fed into the decoder, the dose map at a specific depth is predicted with domain-specific and domain-invariant features incorporated as a prior information. The disentangled learning of feature representation and manipulation via respective encoding and decoding is achieved by integrating the domain label into the decoder, which has been demonstrated to benefit image synthesis.[19,20]

Discriminator

Similar to G, two identical discriminators, D_1 and D_2, are responsible for evaluating the dose map prediction and radioluminescence image reconstruction, respectively. In a conventional GAN, D serves as a real/fake identifier to distinguish the generated images from the ground truth. In the one-to-many domain translation problem, D has two responsibilities. First, in every target domain c_n, D differentiates predicted dose maps ($G_1(x,\ c_n)$) from the ground truth (y_n), which is similar to the discriminator in a conventional GAN. Moreover, D acts as a multidepth discriminator or a domain classifier that estimates the similarity between $G_1(x,\ c_n)$ and every y_n. For this purpose, we modify the last fully connected layer in D to export a one-hot vector c'_n that contributes to a classification loss $\mathcal{L}_{\text{cls}}$. Furthermore, we select an element ($c'_n(n)$) from the predicted one hot vector to make a real/fake assessment for the predicted $G_1(x,\ c_n)$ that results in an adversarial loss $\mathcal{L}_{\text{adv}}$. The above two quantities ($c'_n(n)$ and c'_n) are typically produced by a discriminator consisting of two components in cross-domain networks.[20,21]

The top layers of D are transferred from a VGG16 network to provide low-level features, which are aimed to pay attention to interdomain information (e.g., the distribution in the penumbra regions). The VGG16 is pretrained for a classification task using the ImageNet database.[22] Figure 4 shows a CRIS image and the corresponding feature map, the latter of which is the integration of output from the convolutional layer before the fourth maxpooling in the VGG16 and nominated as ϕ_4. In ϕ_4, pixels with an intensity larger than 50% maximum are observed inside the pernumbra regions, where the interdomain information can be extracted.[7,8] By adversarially learning of the interdomain information, both the image quality and rate of convergence can be improved.[23]

Loss functions

In fGAN, there are three loss functions involved: a perceptual adversarial loss $\mathcal{L}_{\text{adv}}$, a classification loss $\mathcal{L}_{\text{cls}}$,

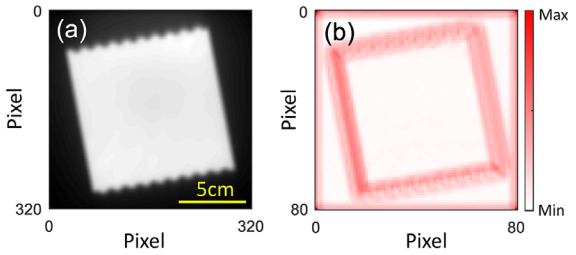

FIGURE 4 (a) The raw measurement of camera-based radioluminescence imaging system (CRIS), and (b) its representation in feature space, extracted from the convolutional layer before the fourth maxpooling layer in a pretrained VGG16. The extracted features were taken as input of the discriminator in fGAN

and a cycle consistency loss $\mathcal{L}_{\text{cyc}}$. $\mathcal{L}_{\text{adv}}$ is formulated as

$$\begin{aligned}\mathcal{L}_{\text{adv}} = &\mathbb{E}_y\left[\log D^{(n)}\left(\Phi_4\left(y_{c_n}\right)\right)\right] \\ &+\mathbb{E}_x\left[\log\left(1-D^{(n)}\left(\Phi_4\left(G_1\left(x,c_n\right)\right)\right)\right)\right],\end{aligned} \tag{1}$$

where $D^{(n)}$ outputs an element ($c'_n(n)$) of the one hot vector (i.e., c'_n output by D) that evaluates the similarity between the synthetic dose map and the corresponding groundtruth y_n at a depth specified by c_n. In the training of fGAN, the radioluminescence image and the TPS dose map that corresponds to the same irradiation field (i.e., paired data) are used as opposed to a typical unsupervised learning, which is demonstrated to effectively enhance the synthesis. $\mathcal{L}_{\text{cls}}$ is expressed as the sum of two cross-entropies calculated for the real dose maps and the synthetic ones, respectively,

$$\begin{aligned}\mathcal{L}_{\text{cls}} = &-\sum_{n=1}^{N} c_n \log D\left(\Phi_4\left(y_{c_n}\right)\right) \\ &-\sum_{n=1}^{N} c_n \log D\left(\Phi_4\left(G_1\left(x,c_n\right)\right)\right).\end{aligned} \tag{2}$$

The cycle consistent loss $\mathcal{L}_{\text{cyc}}$ is intended to minimize the difference between a reconstructed radioluminescence image $G_2(G_1(x,\ c_n),\ c_0)$ and the input x:

$$\mathcal{L}_{\text{cyc}} = \mathbb{E}_x\left[\left\|G_2\left(G_1\left(x,c_n\right),c_0\right)-x\right\|_1\right], \tag{3}$$

where $\|\cdot\|_1$ is *l1*-norm. $\mathcal{L}_{\text{cyc}}$ is used to guarantee that the predicted dose maps preserve the content of their input radioluminescence images, while changing only the interdomain part of the inputs. In this manner, an unsupervised learning that avoids traditional end-to-end fashion is applied to enforce attentions on both the high-frequecy structures and low-dose contents (demonstrated later). Finally, the total objective is

$$\begin{cases}\mathcal{L}_D = -\mathcal{L}_{\text{adv}} + \lambda_{\text{cls}}\mathcal{L}_{\text{cls}} \\ \mathcal{L}_G = \mathcal{L}_{\text{adv}} + \lambda_{\text{cls}}\mathcal{L}_{\text{cls}} + \lambda_{\text{cyc}}\mathcal{L}_{\text{cyc}}\end{cases}, \tag{4}$$

where λ_{cls} and λ_{cyc} are hyperparameters that balance the contributions from classification loss and

consistency loss, respectively. In Equation (4), the domain classification loss ($\mathcal{L}_{cls}$) is actually reduced to $\sum_n -c_n \log D(\phi_4(y_n))$ (as a function of the real CRIS images) in the calculation of $\mathcal{L}_D$, and reduced to $\sum_n -c_n \log D(\phi_4(G(x, c_n)))$ (as a function of the predicted dose maps) in the calculation of $\mathcal{L}_G$.

2.2.2 | Architecture and training of the network

The generator G has an encoder–decoder structure similar to that in,[24] using a scaling factor of two. The decoder involves six residual blocks at the top. The feature extractor is constructed using the same structure as the encoder with the number of input channels equal to that of the desirable domains. Instance normalization and LeakyReLU activation with $\alpha = 0.2$ are used.[25] The discriminator D is based on a VGG structure in which the top layers are transferred from a pertained VGG16 and the last linear layer is modified to output a vector of size N. The subnetworks (G and D) were sequentially trained at each epoch to minimize $\mathcal{L}_G$ and $\mathcal{L}_D$ in Equation (4) until the model converges. λ_{cls} and λ_{cyc} were empirically set as 0.5 and 1, respectively. During the training process, the VGG16 top layers in D were frozen, while the subsequent layers keep trainable to yield the high-level features. All the images and dose maps were resized to 320 × 320. The batch size was set to 12. The whole framework is built on PyTorch with an NVIDIA TITAN V GPU. We trained our model with the Adam optimizer by setting $\beta_1 = 0.5$ and $\beta_2 = 0.999$.[26] The learning rate was initialized as 10^{-4} and linearly decayed after half of the training epochs (150). The training time of the network was around 12 h and the inference time was 0.05 s per image. During the training process, the one hot domain label vector was generated randomly. The image pairs were augmented with random rotations within a range of 15°.

2.3 | Data collection

A LINAC (Varian Clinac 2100 CD, Varian Medical Systems, USA) equipped with Millennium MLC was used for all the experiments with photon energy of 6 MV and dose rate of 600 monitor units per minute (MU/min). Before measurement, the machine performance was checked following the AAPM TG142.[27] The radial section of CRIS phantom was centered to the LINAC isocenter. Dose calculations were conducted in Eclipse (Varian Medical Systems, Palo Alto, CA, USA) using the anisotropic analytical algorithm (version 15.6.05), and the plane dose was exported at a resolution of 0.29 mm. The calculation was performed for a cubic water phantom (45 × 45 × 45 cm^3) constructed in the TPS with a source-to-surface distance (SSD) of 100 cm. For all the experiments, the CRIS settings were fixed to maintain a dynamic range of ∼57 dB (80% of that of the CMOS camera used) to ameliorate the latent nonlinearities when CMOSs near the physical saturation. The measured dataset is divided for network training and validation.

The training dataset involves 58 shapes (see Figure 5). The circular and comb-like shapes were designed to learn the intra- and interleaf features, respectively. The measurement was taken for multiple collimator rotations ranging from 0° to 180° at a step size of 15°. That is, at 12 MLC rotations angles, 12 × 58 images were taken in all. The gantry angle was fixed to 0° during the data collection process. Dose was calculated in the TPS at the depths of 1.5, 5, and 10 cm, and the model was trained for the three depths. The test dataset consists of regular fields and three clinical IMRT cases. The regular fields include a set of MLC-shaped square fields: 2 × 2, 4 × 4, 6 × 6, 8 × 8, and 10 × 10 cm^2. The clinical IMRT cases include a brain, a lung, and a prostate case, with the number of fields (and step-and-shot segments) being six (136), six (142), and seven (77), respectively. The ranges of field sizes are 8.0 × 7.5–13.0 × 8.0 cm^2, 8.0 × 6.5–11.5 × 7.0 cm^2, and 9.0 × 6.5–11.1 × 6.5 cm^2, for the brain, lung, and prostate cases, respectively. The predicted dose was calibrated to the TPS calculation for a 10 × 10 cm^2 field at a depth of 10 cm. A linear relation was assumed between the predicted and calculated doses, with the scaling factor determined from the calibration.

As an additional independent validation to our work, measured dose profiles and percentage depth doses (PDDs) were used to compare with CRIS results. Dose profiles were acquired using an IC Profiler (Sun Nuclear Corporation, Melbourne, FL, USA) at a depth of 1.5 cm for field sizes of 2 × 2, 4 × 4, 6 × 6, 8 × 8, and 10 × 10 cm^2. PDDs for the field sizes of 4 × 4, 6 × 6, and 10 × 10 cm^2 were acquired using an ion chamber in a water phantom (IBA water scan system).

2.4 | Evaluation metrics

The predicted dose maps were evaluated by comparing to the corresponding TPS calculations in terms of global gamma index (Υ).[28] The gamma passing rates were calculated (Υ_{pass}) using gamma criteria of 1% (global intensity)/1 mm (distance-to-agreement) and 2%/2 mm with a low-dose cutoff threshold value of 10%.

3 | EXPERIMENTS AND RESULTS

3.1 | Dose prediction for regular fields

Figure 6 shows an example of 6 × 6 cm^2 field. The images are displayed in log-scale to reveal the details

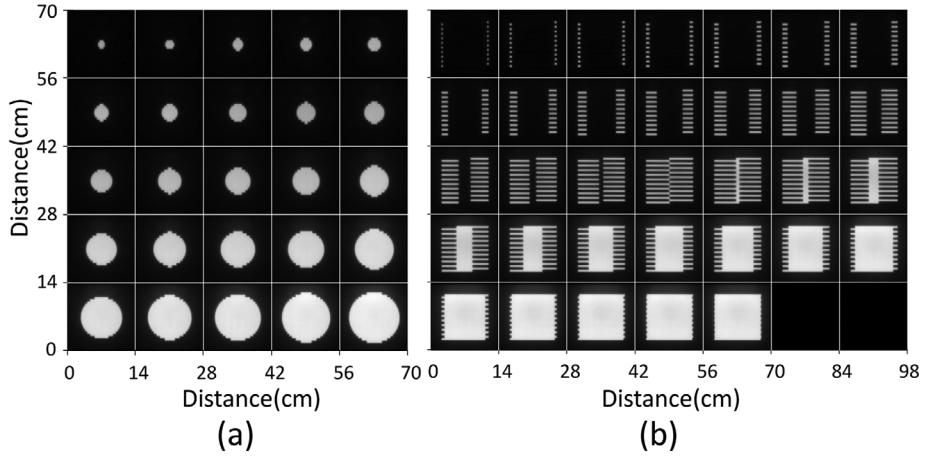

FIGURE 5 Training dataset includes MLC-defined (a) circular and (b) comb-like fields. The two designs are intended for learning disentangled features regarding content and edge. These images presented were collected by using the camera-based radioluminescence imaging system (CRIS)

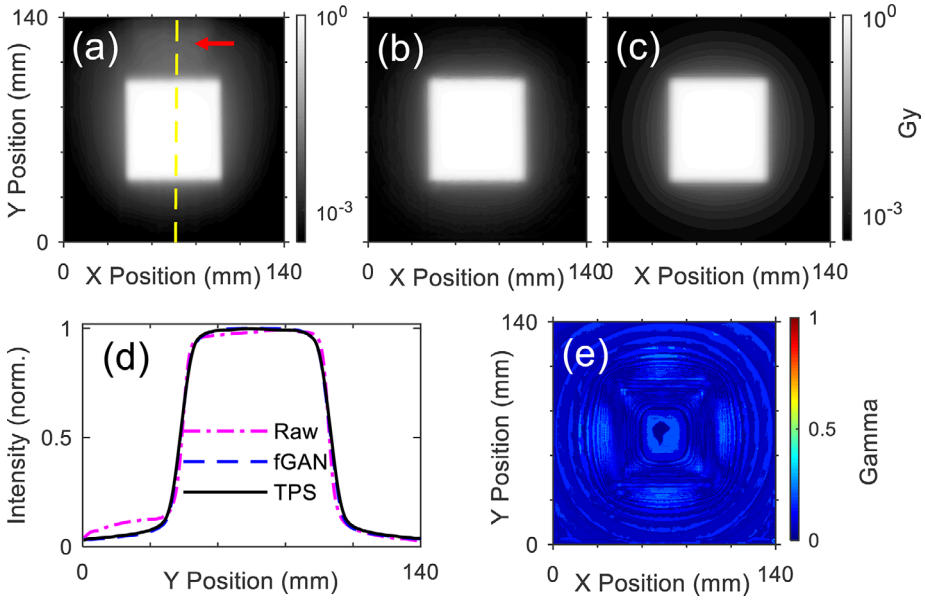

FIGURE 6 Dose prediction for a 6 × 6 cm^2 open field at a water depth of 10 cm. (a) Raw measurement collected from our camera-based radioluminescence imaging system (CRIS), (b) predicted dose map, (c) TPS calculation, (d) intensity profiles along the yellow dashed line in (a), and (e) gamma map of the prediction (2%/2 mm). Images are displayed in log-scale

in the penumbra regions. In this case, the mirror-glare artifacts can be observed on top of radioluminescence images, as indicated with a red arrow in Figure 6a. In the cross-beam profiles shown in Figure 6d, the artifacts lead to a 10% deviation relative to the maximum dose. These artifacts are eliminated in the predicted dose map (Figure 6b), which has a 100% γ_{pass} (2%/2 mm) as shown in Figure 6e ($\gamma < 1$ for all the pixels). More quantitative analysis on fields of $2 \times 2, 4 \times 4, 6 \times 6, 8 \times 8$, and 10×10 cm^2 at depths of 1.5 and 10 cm are listed in Table 1. All γ_{pass} reach 100% with 2%/2 mm criteria and exceed 99% for more stringent 1%/1 mm. In comparison, the raw images have mean γ_{pass} of 80.5% (2%/2 mm) and 61.9% (1%/1 mm) comparing with the dose maps at a depth of 1.5 cm, and 87.6% (2%/2 mm) and 60.7% (1%/1 mm) at a depth of 10 cm. γ_{pass} in the raw data are found inversely related to the field size. For example, the mean γ_{pass} (2%/2 mm) is 99.0% for 2×2 cm^2 and reduces to only 47.3% for 10×10 cm^2.

3.2 | Dose prediction for IMRT cases

We further investigated three IMRT treatment plans delivered at a gantry angle fixed to 0°. As an example, the predicted dose at a depth of 10 cm for the last

TABLE 1 Gamma analysis of predictions and raw images (in bracket) for regular open fields

	Depth (cm)	2×2 cm^2	4×4 cm^2	6×6 cm^2	8×8 cm^2	10×10 cm^2
1%/1 mm	1.5	99.5% (92.1%)	99.9% (81.0%)	100% (57.4%)	100% (48.2%)	100% (30.6%)
	10	99.7% (75.1%)	99.9% (69.9%)	100% (69.8%)	99.3% (60.1%)	99.9% (28.4%)
2%/2 mm	1.5	100% (98.7%)	100% (96.3%)	100% (78.8%)	100% (86.7%)	100% (41.8%)
	10	100% (98.2%)	100% (97.5%)	100% (98.0%)	100% (91.2%)	100% (52.8%)

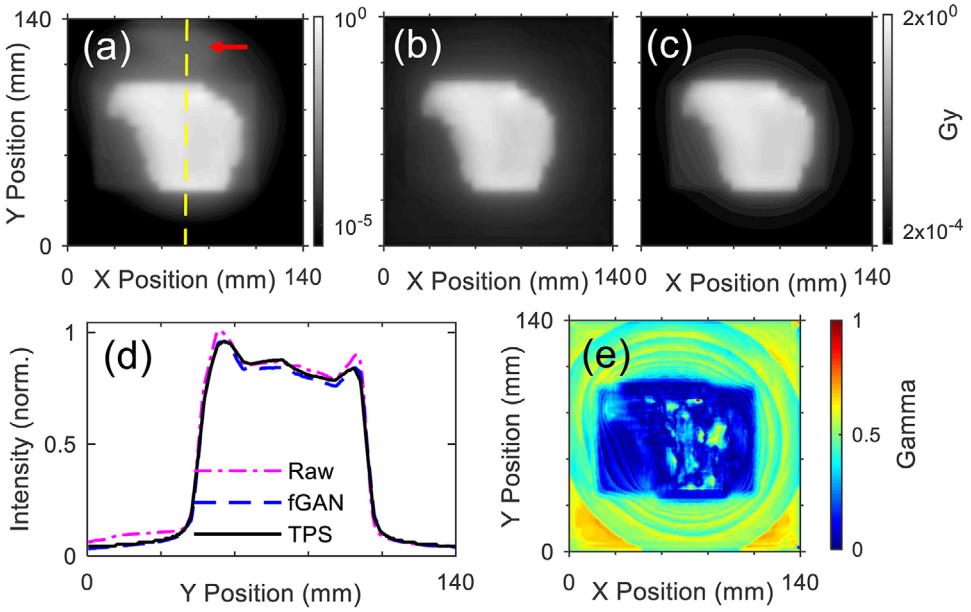

FIGURE 7 Dose prediction for a prostate intensity-modulated radiation therapy (IMRT) case (Field 7) at a water depth of 10 cm. (a) Raw measurement collected from our camera-based radioluminescence imaging system (CRIS), (b) predicted dose map, (c) TPS calculation, (d) intensity profiles along the yellow dashed line in (a), and (e) gamma map of the prediction (2%/2 mm). Images are displayed in log-scale

field of the prostate case is presented in Figure 7. The mirror-glare artifacts indicated with a red arrow are visible in a similar position to that in Figure 6a, accounting for a profile deviation of 4% as shown in Figure 7d. Figure 7e shows the gamma map using 2%/2 mm criterion, where moderate deviations on γ (~0.5) can be found in the low-dose regions surrounding the primary beam. By checking the TPS settings, these regions are mostly covered by the secondary collimator jaws. The quantitative results of the three IMRT plans with 2%/2 mm gamma criteria are summarized in Table 2 for depths of 1.5 and 10 cm. The last column shows the mean gamma pass rates over the total fields in each plan. Single-field passing rate ranges from 91.8% to 99.5%, with mean passing rate all above 95%.

3.3 | Comparison to measurement

Figure 8a shows the predicted cross-beam dose profiles along the *X* and *Y* directions for MLC-shaped fields of 2×2, 4×4, 6×6, 8×8, and 10×10 cm^2. The data obtained from TPS and the measurements via the IC profiler are also plotted for comparison. The gamma pass rates with 1%/1 mm and 2%/1 mm criterions are shown in Table 3. For objective comparison, the 1D gamma index was calculated with a low-dose cutoff threshold value down to 1%. It can be found that the predictions agree well with the corresponding TPS calculations with 100% pass rates, and maintain mean pass rates of 95.0% (*X*-direction) and 87.0% (*Y*-direction) with respect to the measurements. In the latter case, small discrepancies are observed in the shoulder and trail regions of the profiles, presumably because of the ion chamber array has more volume averaging effects and lateral scatter equilibrium problems due to the air cavities of the air-filled ion chamber array. We further compare the PDD results among predictions, TPS calculations, and the measurements for field sizes of 4×4 cm^2 (top), 6×6 cm^2 (middle), and 10×10 cm^2 (bottom), as shown in Figure 8b and quantized in Table 4 in terms of percentage deviation that was defined as the ratio between the predictions and TPS calculations or measurements at 1.5, 5, and 10 cm depths. On average,

TABLE 2 Gamma analysis (2%/2 mm) on three intensity-modulated radiation therapy (IMRT) cases

	Depth (cm)	Field 1 (%)	Field 2 (%)	Field 3 (%)	Field 4 (%)	Field 5 (%)	Field 6 (%)	Field 7 (%)	Mean (%)
Brain	1.5	97.2	99.3	95.9	97.3	94.8	99.8	N/A	97.4
	10	98.9	99.6	91.8	94.4	93.2	97.9	N/A	96.0
Lung	1.5	92.7	96.8	99.3	97.8	96.9	95.2	N/A	96.5
	10	93.1	95.8	98.3	94.4	93.2	97.9	N/A	95.5
Prostate	1.5	99.0	98.8	99.3	99.3	99.5	99.1	99.3	99.2
	10	98.8	99.4	98.8	99.5	98.5	99.2	99.3	98.7

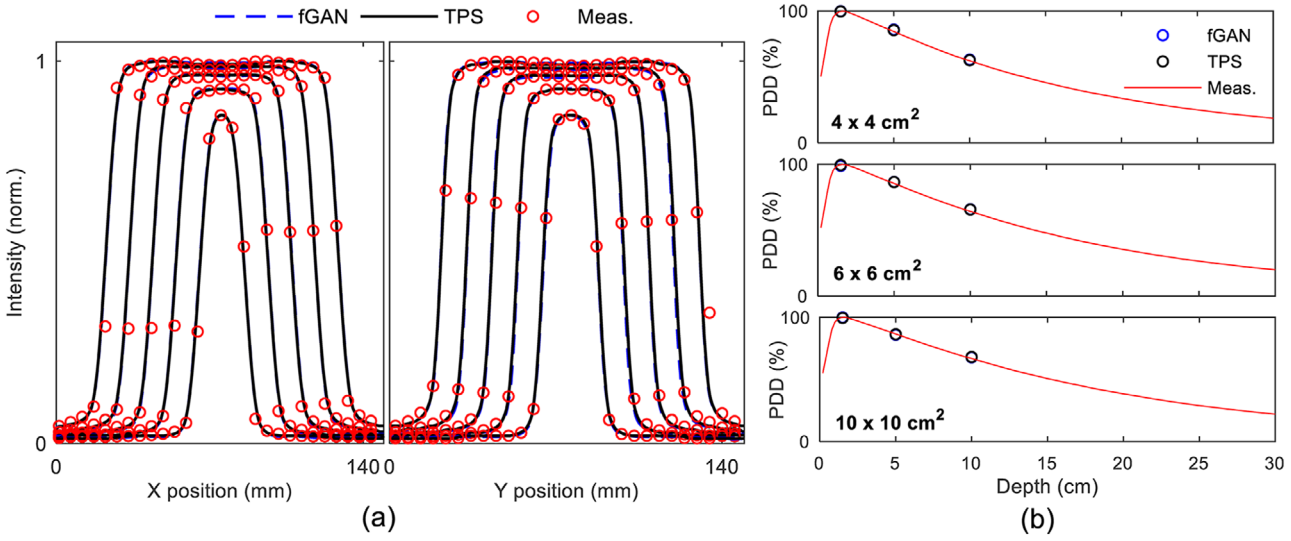

FIGURE 8 (a) Dose profiles along *X* (left) and *Y* (right) directions obtained with prediction, TPS and measurement for field sizes of 2×2, 4×4, 6×6, 8×8, and 10×10 cm^2, and (b) percentage depth dose curves for field sizes of 4×4 cm^2 (top), 6×6 cm^2 (middle), and 10×10 cm^2 (bottom). *X* and *Y* directions are defined in Figure 6(a)

the predictions were reported within 0.40% and 1.15% of the results from TPS calculations and IC profiler measurements, respectively.

4 | DISCUSSION

4.1 | Verification of failure deliveries

The capability to detect a failure during the treatment was demonstrated. To this end, 20 MLC leaf pairs with various displacement errors (d) were used to deliver a field of 100.0×100.0 mm^2, as depicted in Figure 9a. The fields were delivered four times with $d = 1, 2, 3,$ and 4 mm. The reference dose map is shown for a normal

TABLE 4 Percentage deviations of predicted PDD relative to TPS counterparts (ΔPDD_{TPS}) and measurements (ΔP_{meas}) for various fields

	4×4 cm^2	6×6 cm^2	10×10cm^2
ΔPDD_{TPS}	0.41%	0.30%	0.39%
ΔPDD_{meas}	1.19%	1.54%	0.73%

Note: Deviations are averaged for those calculated from depths of 1.5, 5, and 10 cm.

TABLE 3 Gamma analysis on predicted dose profile relative to TPS calculations (γ_{TPS}) and measurements (γ_{meas}) for various regular fields

		2×2 cm^2	4×4 cm^2	6×6 cm^2	8×8 cm^2	10×10 cm^2
X-direction	γ_{TPS}	100%	100%	100%	100%	100%
	γ_{meas}	89.8% (93.8%)	96.4% (100%)	92.9% (96.4%)	89.3% (96.5%)	96.4% (96.4%)
Y-direction	γ_{TPS}	100%	100%	100%	100%	100%
	γ_{meas}	100% (100%)	85.7% (96.3%)	85.7% (96.1%)	82.1% (94.9%)	81.6% (93.9%)

Note: Gamma indices are calculated with 1%/1 mm and 2%/1 mm (in bracket) criteria.

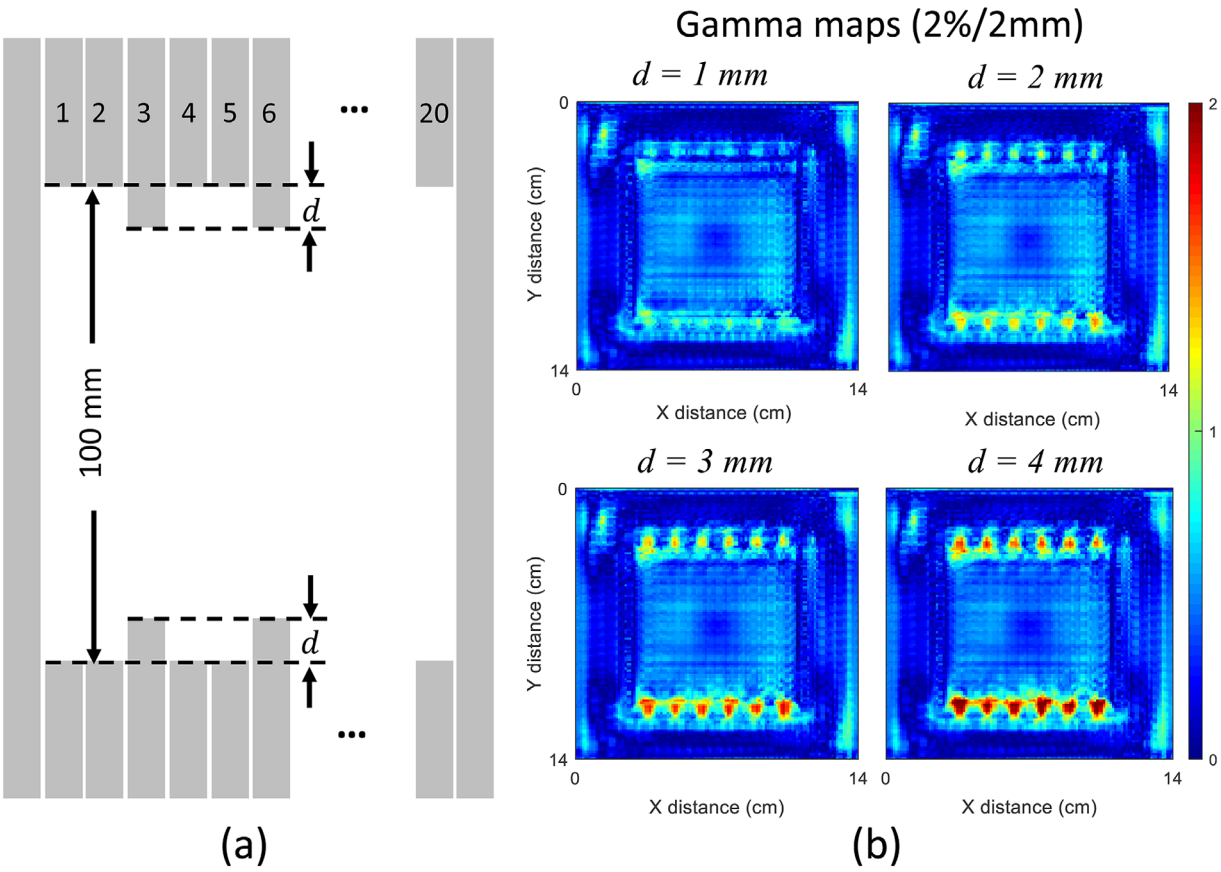

FIGURE 9 Gamma analysis for failure deliveries. (a) Geometric description of the field with introduced MLC leaf displacement errors noted by d, and (b) the resultant 2%/2 mm gamma distributions for $d = 1, 2, 3$, and 4 mm, respectively

delivery with $d = 0$. The gamma distributions (2%/2 mm) are calculated in Figure 9b. It can be found that the abnormal dose regions caused by erroneous leaf positionings are identified and highlighted with larger gamma index ($\Upsilon > 1$). The 2%/2 mm gamma pass rates are 95.6% ($d = 1$ mm), 93.6% ($d = 2$ mm), 89.8% ($d = 3$ mm), and 86.8% ($d = 4$ mm), yielding differences of 0.3%, 2.3%, 6.1%, and 9.1%, respectively, comparing to the normal delivery result (95.9%). And the 3%/3 mm gamma pass rates are 98.3% ($d = 1$ mm), 98.2% ($d = 2$ mm), 97.1% ($d = 3$ mm), and 95.1% ($d = 4$ mm), yielding differences of 0.1%, 0.2%, 1.3%, and 3.3%, respectively.

4.2 | Mirror-glare artifacts

In CRIS, mirror-glare effects were found to be a major confounding artifact,[5,7] as seen in the radioluminescence images (Figures 6a and 7a). Mirror glare artifacts are caused by the interreflections between the mirror and the phosphor screen, and are aggravated with a reduced distance in between. In our study, glare artifacts are observed on top of the raw image, where the distance between the hemispherical mirror and the scintillator sheet is smaller than other places. The artifacts can be reduced by changing the irradiation field far away from the top sensitive regions of the mirror, which explains the inverse relation between the γ_{pass} and the field sizes. The artifacts are eliminated in the dose maps predicted using fGAN. In fGAN, a domain translation is enhanced via the cycle consistency supervision, which enforces attention to both the high-frequency structures and low-dose contents. Using $\mathcal{L}_{\text{cyc}}$, a radioluminescence image is reconstructed from the dose map by importing a source domain label c_{src}. In this process, the generator learns the representations of the raw image contaminated with the glare artifacts, which in turn benefits the artifacts removal. Apart from the glare artifacts, γ_{pass} at the depth of 1.5 cm is always higher than that at 10 cm, which might be explained by the fact that the trained nonlinearities in the network fail to accurately approximate the stronger diffuse effect at a larger depth.

4.3 | Advantages, limitations, and ongoing works

Using the proposed method, promising passing rates were achieved in measurements with regular open

fields. Slight degradation was observed in the IMRT cases, which could be explained by the limited dose prediction accuracy for some highly irregular field shapes and sizes. By examining the delivery, we noticed that the segments with a leaf gap less than 0.2 cm generally led to limited dose accuracy. For example, the prediction for the third field of the brain case has a gamma passing rate of only 91.8% (2%/2 mm), in which 40% segments contain one or more leaf pairs with a gap less than 0.2 mm. This could be attributed to the fact that the model was not trained for such a small field size. Additionally, current training datasets were collected with the secondary collimator jaws retracted, and thus the influence from the jaw position was not learned by fGAN. As the jaws follow MLC in the IMRT treatment, errors were found in the low-dose regions (see Figure 7b,c). However, the errors were very limited (a mean gamma index of ∼0.5 is found in Figure 7e) and mostly appear in typical cutoff regions (those with dose less than 10% threshold). This adversity may be more pronounced for those scenarios with extremely low dose. Additional training could be performed in the future, with the effect of jaw setting taken into consideration.

As a proof-of-concept demonstration, the gantry angle was fixed at 0° for all the experiments due to the limited size (150 × 150 mm^2) of the scintillator sheet. However, the key techniques that were demonstrated for both image restoration (as schematized in Figure 2) and image–dose conversion at a fixed gantry angle remain applicable for other angle scenarios benefiting from the consistent beam's eye-view design. As for applications on verifying dynamic treatments such as VMAT, a salient and unique feature that potentially challenges the current image-to-dose conversion method is the integration effect over the dynamic fields in every single frame. Given the consistency of good linearity of the Gd_2O_2S-based scintillator material in dose response and structural invariability of the VMAT measurement in feature space, the CNN-based dose engine that specifically deals with abstractive features promises to be effective with no need for retraining. The measurement might fail at low-dose regions when the temporal resolution (5 fps currently) is too low to support a reasonable signal-to-noise ratio (SNR) or signal-to-background ratio (SBR).

In practice, several latent limitations are found: (i) CMOS cameras will experience image degradation following exposure to radiation. While no such degradation was noted in the cameras used throughout this study, further work will be necessary to fully investigate the usable lifetime of both the phantom and camera used in the system. (ii) Dose accuracy might be degraded for the treatment delivery with very large fields. As demonstrated in Section 2.1, the image quality has a reduction against the distance to the mirror center due to the hemispheric distortion. (iii) Prediction on absolute dose distribution in a specific water phantom is based on the assumption of a good consistency between the TPS setting and actual beam energy. In reality, dose measurement in most imaging-oriented QA devices (e.g., EPID and ArcCHECK) is subject to this assumption. However, the above weakness does not eliminate the value of such devices as the whole treatment procedure involves much more than LINAC beam performance. Other factors include dose calculation, imaging guidance, MLC performance, etc. Any performance change in those components will affect the QA agreement. Furthermore, LINAC energy is usually very stable, and could be considered sufficiently checked if it is included in monthly and daily QA.

The proposed dose engine has the potential for full-depth prediction in a specific water phantom. To this end, fGAN needs to be trained for sufficient number of depths within the interrogated range to fulfill a 3D interpolation that enables volume rendering of predicted dose maps. Preliminary experiments show that a dense depth distribution of ∼1 mm space within 3 cm range is necessary to follow the high-dose variations in the buildup region, while a coarse distribution with ∼10 mm space is reasonable outside this region. To allow for absolute dosimetry, the linearity of the detector response relative to dose rate is an important measure. We did not include the linearity study about the Gd_2O_2S-based scintillator material in this work, because it has been studied extensively in another study we performed on real-time beam visualization.[16]

5 | CONCLUSION

In this work, a novel dosimetric verification system is developed using a radioluminescence imaging system. The system involves a cylindrical sensing receptor to allow for a coplanar detection and maintains a relatively high spatial resolution. Given the specific geometric design that complicates the dose response function, a data-driven learning strategy is proposed to enable reliable, practical, and robust dose measurement without explicitly investigating the imaging physics. The proposed deep learning model enables flexible domain transformation from radioluminescence image to corresponding dose maps at multiple depths, potentially offering a way for full 3D dosimetry. The proposed system is validated by comparing to TPS results and measurements in regular fields ranging from 2 × 2 to 10 × 10 cm^2 and in three clinical IMRT cases. This study renders a way for further advancing the volumetric dosimeters in terms of spatial resolution and economic cost, promising to enhance the armamentarium of treatment plan verification.

ACKNOWLEDGMENTS

The authors acknowledge the funding supports from the National Cancer Institute (1R01CA223667 and 1R01CA227713).

CONFLICT OF INTEREST

The authors declare that there is no potential conflict of interest.

DATA AVAILABILITY STATEMENT

The data are available upon request from the authors.

REFERENCES

1. Han B, Ding A, Lu M, Xing L. Pixel response-based EPID dosimetry for patient specific QA. *J Appl Clin Med Phys*. 2017;18(1):9-17.
2. Bissonnette JP, Balter PA, Dong L, et al. Quality assurance for image-guided radiation therapy utilizing CT-based technologies: a report of the AAPM TG-179. *Med Phys*. 2012;39(4):1946-1963.
3. Bailey DW, Kumaraswamy L, Bakhtiari M, et al. EPID dosimetry for pretreatment quality assurance with two commercial systems. *J Appl Clin Med Phys*. 2012;13(4):82-99.
4. Feygelman V, Zhang G, Stevens C, Nelms BE. Evaluation of a new VMAT QA device, or the "X" and "O" array geometries. *J Appl Clin Med Phys*. 2011;12(2):146-168.
5. Frelin AM, Fontbonne JM, Ban G, et al. The DosiMap, a new 2D scintillating dosimeter for IMRT quality assurance: characterization of two Čerenkov discrimination methods. *Med Phys*. 2008;35(5):1651-1662.
6. Goulet M, Archambault L, Beaulieu L, Gingras L. High resolution 2D dose measurement device based on a few long scintillating fibers and tomographic reconstruction a. *Med Phys*. 2012;39(8):4840-4849.
7. Lee M, Ding K, Yi B. A single-optical kernel for a phosphor-screen-based geometric QA system (RavenQA) as a tool for patient-specific IMRT/VMAT QA. *Phys Med Biol*. 2018;63(20):20NT03.
8. Cheon W, Kim SJ, Kim K, et al. Feasibility of two-dimensional dose distribution deconvolution using convolution neural networks. *Med Phys*. 2019;46(12):5833-5847.
9. Bruza P, Andreozzi JM, Gladstone DJ, et al. Online combination of EPID & Cherenkov imaging for 3-D dosimetry in a liquid phantom. *IEEE Trans Med Imaging*. 2017;36(10):2099-2103.
10. Wendling M, Louwe RJ, McDermott LN, Sonke JJ, van Herk M, Mijnheer BJ. Accurate two-dimensional IMRT verification using a back-projection EPID dosimetry method. *Med Phys*. 2006;33(2):259-273.
11. Brost E, Watanabe Y. Space-variant deconvolution of Cerenkov light images acquired from a curved surface. *Med Phys*. 2019;46(9):4021-4036.
12. Alhazmi A, Gianoli C, Neppl S, et al. A novel approach to EPID-based 3D volumetric dosimetry for IMRT and VMAT QA. *Phys Med Biol*. 2018;63(11):115002.
13. Liu H, Li F, Park J, et al. Feasibility of photon beam profile deconvolution using a neural network. *Med Phys*. 2018;45(12):5586-5596.
14. Yuan Y, Liu S, Zhang J, Zhang Y, Dong C, Lin L. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2018. https://doi.org/10.1109/CVPRW.2018.00113
15. Sun C, Pratx G, Carpenter CM, et al. Synthesis and radioluminescence of PEGylated $Eu3^{+}$-doped nanophosphors as bioimaging probes. *Adv Mater*. 2011;23(24):H195-H199.
16. Xing L, Naczynski DJ, Jenkins C. Visualizing radiation therapy beam in real-time in the context of patient's anatomy. U.S. patent 9,604,077. March 28, 2017.
17. Jenkins CH, Naczynski DJ, Shu-Jung SY, Yang Y, Xing L. Automating quality assurance of digital linear accelerators using a radioluminescent phosphor coated phantom and optical imaging. *Phys Med Biol*. 2016;61(17):L29-L37.
18. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. *Generative Adversarial Networks*. Thesis. Cornell University; 2014. https://arxiv.org/abs/1406.2661v1
19. Donahue C, Lipton ZC, Balsubramani A, McAuley J. *Semantically Decomposing the Latent Spaces of Generative Adversarial Networks*. Thesis. Cornell University; 2017. https://arxiv.org/abs/1705.07904v3
20. He Z, Zuo W, Kan M, Shan S, Chen X. Attgan: facial attribute editing by only changing what you want. *IEEE Trans Image Process*. 2019;28(11):5464-5478.
21. Choi Y, Choi M, Kim M, Ha JW, Kim S, Choo J. Stargan: unified generative adversarial networks for multi-domain image-to-image translation. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018. https://doi.org/10.1109/CVPR.2018.00916
22. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: a largescale hierarchical image database. Paper presented at: 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009. https://doi.org/10.1109/cvprw.2009.5206848
23. Yu B, Zhou L, Wang L, Shi Y, Fripp J, Bourgeat P. Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis. *IEEE Trans Med Imaging*. 2019;38(7):1750-1762.
24. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. Paper presented at: Proceedings of the IEEE International Conference on Computer Vision; 2017. https://doi.org/10.1109/ICCV.2017.244
25. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. Cornell University; 2015. https://arxiv.org/abs/1511.06434v2
26. Kingma DP, Ba J. Adam: a method for stochastic optimization; 2014. https://arxiv.org/abs/1412.6980v9
27. Klein EE, Hanley J, Bayouth J, et al. Task Group 142 report: quality assurance of medical accelerators. *Med Phys*. 2009;36(9):4197-4212.
28. Low DA, Dempsey JF. Evaluation of the gamma dose distribution comparison method. *Med Phys*. 2003;30(9):2455-2464.

How to cite this article: Jia M, Yang Y, Wu Y, Li X, Xing L, Wang L. Deep learning-augmented radioluminescence imaging for radiotherapy dose verification. *Med. Phys.* 2021;48:6820–6831. https://doi.org/10.1002/mp.15229