HHS Public Access

Author manuscript

Phys Med Biol. Author manuscript; available in PMC 2021 April 23.

Published in final edited form as:

Phys Med Biol. ; 66(4): 045014. doi:10.1088/1361-6560/abd673.

Deep learning-augmented radiotherapy visualization with a cylindrical radioluminescence system

Mengyu Jia[1,3], **Xiaomeng Li**[1,3], **Yan Wu**[1], **Yong Yang**[1], **Priya Kasimbeg**[2], **Lawrie Skinner**[1], **Lei Wang**[1,*], **Lei Xing**[1,*]

[1]Department of Radiation Oncology, Stanford University, Palo Alto 94304, United States of America

[2]School of Engineering, Stanford University, Palo Alto 94304, United States of America

[3]These two authors contribute equally to the work

Abstract

This study aims to demonstrate a low-cost camera-based radioluminescence imaging system (CRIS) for high-quality beam visualization that encourages accurate pre-treatment verifications on radiation delivery in external beam radiotherapy. To ameliorate the optical image that suffers from mirror glare and edge blurring caused by photon scattering, a deep learning model is proposed and trained to learn from an on-board electronic portal imaging device (EPID). Beyond the typical purposes of an on-board EPID, the developed system maintains independent measurement with co-planar detection ability by involving a cylindrical receptor. Three task-aware modules are integrated into the network design to enhance its robustness against the artifacts that exist in an EPID running at the cine mode for efficient image acquisition. The training data consists of various designed beam fields that were modulated via the multi-leaf collimator (MLC). Validation experiments are performed for five regular fields ranging from 2×2 cm^2 to 10×10 cm^2 and three clinical IMRT cases. The captured CRIS images are compared to the high-quality images collected from an EPID running at the integration-mode, in terms of gamma index and other typical similarity metrics. The mean 2%/2 mm gamma pass rate is 99.14% (range between 98.6% and 100%) and 97.1% (ranging between 96.3% and 97.9%), for the regular fields and IMRT cases, respectively. The CRIS is further applied as a tool for MLC leaf-end position verification. A rectangular field with introduced leaf displacement is designed, and the measurements using CRIS and EPID agree within 0.100 mm ± 0.072 mm with maximum of 0.292 mm. Coupled with its simple system design and low-cost nature, the technique promises to provide viable choice for routine quality assurance in radiation oncology practice.

Keywords

radioluminescence imaging; super-resolution; perceptual loss

[*]Authors to whom any correspondence should be addressed. leiwang@stanford.edu and lei@stanford.edu.

1. Introduction

The current trend of external beam radiation therapy is moving toward increasingly precise and accurate delivery of highly conformal, organ-at-risk sparing dose distributions. In response to the increasing delivery complexity, pre-treatment quality assurance (QA) is requisite to ensure safe treatment delivery. Most dosimetric and geometrical QA depend on a megavoltage X-ray imaging modality that can accurately record the beam profile and intensity (Klein et al 2009). Electronic portal imaging device (EPID) is a popular imaging tool that has been exploited for various QA purposes and is already integrated in most modern linear accelerator (linac), referred to as on-board EPID. While on-board EPIDs maintain wide availability and high image quality, they do not provide independent measurement needed by certain QA tasks, e.g. end-to-end verification on image-guided treatment, 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. Detector array-based devices, such as ArcCHECK™ (Sun Nuclear, Melbourne, FL), extend the independent measurements with co-planar detection ability to allow verifications on linac gantry angle (Feygelman et al 2011). 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 application.

Camera-based radioluminescence imaging system (CRIS) provides a cost-effective alternative to EPID and has been sought after for geometric (Jenkins et al 2015,2016) and dosimetric verifications in radiation therapy (Frelin et al 2008, Guillot et al 2011, Cheon et al 2019). Benefitting from a modern imager sensor used, high acquisition rate and high measurement sensitivity could be readily reached, which is significant for beam visualization in real time with high dynamic range. However, CRISs suffer from severe optical scattering that leads to edge blurring and mirror-glare artifacts. To ameliorate the image quality, a variety of solutions have been exploited towards hardware and software optimizations. Collomb-Patton proposed to improve the image quality with a flat-field calibration approach, in which, a dosimetric film was overlaid on top of the scintillator sheet to acquire the reference image (Collomb-Patton et al 2009). However, this correction was limited in small field scenarios, and was advanced by the subsequent solutions based on deconvolution. To proceed with practical deconvolution operation, efforts have been made to conditionally simplify the kernel expression. Lee et al proposed a single-optical kernel strategy by assuming a spatially invariant and angularly isotropic scattering behavior in the detected radiation image (Lee et al 2018). Synthesizing EPID images from CRIS images is essentially a super-resolution (SR) problem in the computer vision field, which aims to restore a high resolution (HR) image from the low resolution (LR) counterpart coupled with diffuse-induced artifacts. SR algorithms have been greatly improved via various deep learning techniques, e.g. generative adversarial network (GAN) (Goodfellow et al 2014). A representative SR GAN is the one proposed by Wang et al (2018), referred to as enhanced SRGAN (ESRGAN). SRGAN benefits from joint improvements in network structure (a residual-in-residual dense block (RRDB) was proposed and used as the basic network building unit) and loss function design (a perceptual loss that measures the similarity distance in feature space was used to maintain high-frequency details). To focus on task-

specific structures and further improve the image quality, Biting *et al* recently proposed an edge-aware GAN (Ea-GAN) to enhance the cross-modality synthesis performance (Yu et al 2019). However, the attention mechanism is achieved using an analytical approach that may mistakenly focus on task-irrelevant structures such as the artifacts, thus degrade the image quality.

There are three major challenges in synthesizing high-quality EPID image from CRIS image using a typical SR network: (1) *The EPID images employed as the training ground truth contain synchronization artifacts, which are referred to as 'weak labels'*. The artifacts arise when an EPID running at full speed (~10 frame per second), namely the cine mode (Mooslechner et al 2013). With an EPID running in cine mode, multiple frames can be acquired for each treatment field (one frame per control point) as opposed to integration-mode, allowing for efficient collection of the training dataset. Furthermore, predictions on every control point in a treatment field encourage a comprehensive understanding of treatment delivery (Korreman et al 2009). (2) *There is a latent field shape discrepancy from mechanical limitation*, e.g. *reproducibility of* multi-leaf collimator (MLC) *leaf position, in the respective deliveries for EPID and CRIS imaging*. Therefore, the network has to learn abstractive features (e.g. the blurring behavior in the penumbra region) that are less sensitive to field size variations, instead of pixel-wise mapping. (3) *The learning network needs attention mechanisms to maintain high-fidelity beam visualization that determines the accuracy of related QA tasks*. For example, QA of MLC imposes sub-millimeter accuracy on the MLC leaf-end positions that are read from the acquired image (Klein et al 2009).

In this work, we developed a CRIS with deep-learning-based image processing for high-quality beam visualization in radiotherapy, which promises to be a clinical QA tool for reliable mechanical and dosimetric verifications. This system contains a cylindrical receptor that allows a co-planar detection fashion similar to an ArcCHECK™ to catch the latent errors from gantry angle and achieves an image quality comparable to an on-board EPID. To address the challenges above, task-aware perceptual modules are integrated into the deep learning network design by optimizing on critical structures that are highly related to the beam fidelity. Experiments were performed to visualize various beam shapes modulated via MLC. The imaging quality was validated by comparing to the benchmark EPID images collected in the integration-mode for both regular beam fields and three IMRT cases. Additional experiments were performed to explore the application on MLC leaf-end position verification. The proposed deep learning mode was compared to other state-of-art models in this application.

2. Method and materials

2.1. Proposed radioluminescence imaging system

2.1.1. Hardware design—Figure 1 shows the developed CRIS system. The inner surface of a 3D printed hollow cylinder is overlaid with a scintillator sheet (DRZ-plus™, MCI Optonix, Sedona, Arizona, USA), which is composed of three layers, i.e. a 6 μm thick protective layer (polyester), a 208 μm thick phosphor layer (Gd_2O_2S), and a 250 μm thick supporting layer (plastic). The scintillator sheet emits 545 nm light upon interaction with the megavoltage (MV) photons. The radiation-induced light from the scintillation coated layer is

reflected by a hemispherical mirror mounted at the far end of the cylinder and recorded by a CMOS camera (GS3-PGE-23S6M-C, Point Grey Research, Inc., Richmond, Canada) mounted at the other end of the cylinder. The spatial resolution of the CMOS camera is 1920 × 1200 pixel, resulting in a spatial resolution of ~0.45 mm in the region with the most distortion. By aligning the imaging center to that of the hemispheric mirror and the gantry isocenter, the system allows a co-planar detection for all the gantry angles theoretically. For proof-of-concept demonstration in this work, the scintillator sheet partially covers the inner surface with a range of 214°.

2.1.2. System calibration—The geometric distortion caused by hemispheric mirror was corrected via affine transformation, in which, the deformation field was measured by using a chessboard overlaid to the scintillator sheet. Additional calibrations include the typical dark-field and flood-field corrections (Van Nieuwenhove et al 2015). Five plastic fiducial points (see blue dots in figure 1) are distributed on surface of the phantom to work with the room laser system for alignment. Similar to ArcCHECK™ and other QA systems that contain a cylindrical or conical sensing receptor, the developed system is sensitive to any misalignment, which, in turn, encourages the potential for verification on radiation isocenter. Calibrated images still suffer from the blurring and mirror-glare issues caused by light scattering, which are mitigated using the deep learning model trained from the images collected from the cine-mode EPID.

2.2. Deep learning model for radioluminescence image enhancement

2.2.1. Overview of cRI-GAN—Figure 2 shows the pipeline of the cRI-GAN framework for radioluminescence image enhancement. The input images (x) were obtained from the CRIS; and the weak labels (y) were EPID images acquired in the cine-mode. The network consists of a generator (G), a discriminator (D) and three subnetworks responsible for deriving task-aware perceptual losses. We define that a leaf edge is composed of a leaf end and two inter-leaf boundaries. The task-aware perceptual modules take the generated images ($G(x)$) and the weak label (y) as the input, and generate the representations of task-specific structures in the feature space as the output, including leaf ends that measure the leaf position, leaf edges that determine the beam profile, and mirror glare regions that contain the dominant artifacts in CRIS images. These output features are used to quantify the corresponding perceptual loss functions. In particular, an adversarial loss ($\mathscr{L}_{adv}$) is constructed to keep the beam geometry, a style loss ($\mathscr{L}_{styl}$) is designed to specifically handle the accuracy of leaf ends, and a content loss ($\mathscr{L}_{cont}$) is used to suppress glare artifacts and prevent overfitting outside the primary beam. Finally, a weighted combination of these three losses forms the objective function. By automatically learning the task-specific features and minimizing the corresponding perceptual losses, a selective learning is formulated to preserve the high-resolution characteristics in the weak labels while eliminating both synchronization artifacts and mirror-glare artifacts.

2.2.2. Task-aware perceptual modules

Leaf-end awareness (Style loss): Difference between the leaf end and the inter-leaf boundary was investigated to automatically extract the leaf ends representations in feature

space. For a CRIS image shown in figure 3(a), the edge spread functions are found to be distinct (figure 3(b)) from the inter-leaf boundary, which could be physically explained by the curved face design of the leaf end structure (Macdonald et al 2020). Leveraging from the sensitivity of convolution operation on gradient variance, a deep neural network (DNN) is used to learn the characteristic scattering behavior so as to elicit the leaf end representation. The DNN is based on a pretrained VGG16 (Simonyan and Zisserman 2014) to reuse the latent shared feature information. Example feature maps of the arbitrary CRIS image extracted from the VGG16 are shown in figure 3(c), nominated as $\mathscr{F}_i(x)$, which are the integrations of the output of the convolutional layer before the ith max-pooling. Low-level features such as the leaf edges can be found in the shallow layers, e.g. $\mathscr{F}_1(x)$ to $\mathscr{F}_3(x)$. Proceeding towards deeper layers, higher-level features with more abstractive information are learned. In this procedure, the representation of dilated leaf end was found in the mid-level feature maps $\mathscr{F}_4(x)$. A style loss $\mathscr{L}_{styl}$ is then defined to minimize the perceptual similarity between $\mathscr{F}_4(y)$ and $\mathscr{F}_1(G(x))$, formulated as

$$\mathscr{L}_{styl} = \mathbb{E}[\|\boldsymbol{\mathscr{F}}_4(y) - \boldsymbol{\mathscr{F}}_4(\boldsymbol{G}(x))\|_1], \tag{1}$$

where $\|\cdot\|_1$, denotes *l1*-norm. It is noteworthy that, dilation on desirable structure is critical for the training stabilization and synthesis quality in SR problems (Rad et al 2019).

- **Leaf-edge awareness (Adversarial loss):** The leaf edges, which determines the basic beam profile, are used as the metric to evaluate the overall image synthesis. Along this line, the leaf edges representations are taken as input to the discriminator (D) for a real/fake identification. The leaf edge is represented via $\mathscr{F}_2(x)$, which outperforms other feature representations in structure fidelity and signal contrast. The adversarial loss is expressed as:

$$\mathscr{L}_{adv} = \mathbb{E}_{y\sim P_{\mathrm{EPID}(y)}}[\log\ D(\boldsymbol{\mathscr{F}}_2(y))] + \mathbb{E}_{x\sim P_{\mathrm{CRIS}(x)}}[\log(1 - \mathrm{D}(\boldsymbol{\mathscr{F}}_2(\boldsymbol{G}(x, c_n))))], \tag{2}$$

where $P_{\mathrm{CRIS}}(x)$ and $P_{\mathrm{EPID}}(y)$ is the distribution of the measurement from original CRIS image and the cine-mode EPID image, respectively. A joint regularization via $\mathscr{L}_{styl}$ and $\mathscr{L}_{adv}$ is formulated on the critical leaf end and leaf boundary.

Glare-region awareness (Content loss): The mirror-glare artifacts are visible in the dark regions on top of the image (see figure 6(a)), referred to as glare region. To preserve the penalties over the critical structures in $\mathscr{L}_{styl}$ and $\mathscr{L}_{adv}$, the glare-region is represented as the complement of the leaf edge, i.e. $\overline{\mathscr{F}}_2(= \mathscr{F}_2^{\max} - \mathscr{F}_2)$. The artifacts in the glare region are eliminated by the pixel-wise correction via the content loss

$$\mathscr{L}_{cont} = \mathbb{E}\{\|[y - \boldsymbol{G}(x)] \otimes \bar{\boldsymbol{F}}_2(y)\|_1\}, \tag{3}$$

where $\otimes$ denotes an element-wise product.

Total loss: The final objective function to optimize G and D can be formulated as

$$\begin{cases} \mathscr{L}_D = -\mathscr{L}_{adv} \\ \mathscr{L}_G = \mathscr{L}_{adv} + \lambda_c \mathscr{L}_{cont} + \lambda_s \mathscr{L}_{styl} \end{cases}, \tag{4}$$

where λ_c and λ_s are used to weight the content loss and style losses with respect to the adversarial loss.

2.2.3. Network details—Figure 4(a) shows the architecture of generator *G*, which contains two convolution blocks with a stride size of two for down-sampling, fifteen RRDB (Wang et al 2018), two transposed convolution blocks with a stride size of two for up-sampling, and one convolution layer followed by Tanh activation. A global skip connection is introduced, which makes the network to focus on learning the residual correction to the original CRIS image, encouraging faster training process and better network generalizability (Kupyn et al 2018). Figure 4(b) illustrates the framework of discriminator *D*, where the top layers are transferred from a pertained VGG16, and the subsequent layers include three convolution blocks with a stride size of two for down-sampling and a convolution layer followed by two dense layers with Tanh and Sigmoid activations. In every convolution block, there are two convolution layers followed by a batch-normalization layer and a LeakyReLU activation with $\alpha = 0.2$ (Radford et al 2015). While the top layers are frozen, the subsequent layers remain trainable to yield high-level features for semantic evaluation on the generated images.

At each epoch, G was updated once followed by five-times *D* updates. The Adam optimizer was used with $\beta_1 = 0.9$, $\beta_1 = 0.999$ and $\epsilon = 10^{-8}$ (Kingma and Ba 2014). The batch size was set as 4. The learning rate is initialized as 10^{-4} for both *G* and *D*, and linearly decayed after half the training epochs. In all experiments, $\lambda_c = 0.5$ and $\lambda_{styl} = 1$ were set empirically. A large λ_c showed negative effects on both the image synthesis quality and training stability, which might be explained by the adversities from the pixel-wise minimization in $\mathscr{L}_{cont}$. The whole framework was built on PyTorch with an NVIDIA TITANV GPU. The training time of the network was around twelve hours and the inference time was 0.05 s per image. To compromise on the learning speed and GPU memory, all the images were normalized and resized to 320 × 320, corresponding to a pixel size 0.44 mm for the regions with the most distortion. During the training process, the samples of CRIS and EPID images were fed in random order.

2.3. Dataset collection

A Varian linear accelerator (linac) (2100CD, Varian Medical Systems, Palo Alto, California) equipped with a Millennium MLC was used to deliver 6 MV x-ray beams at 600 MU min^{-1}. The linac was calibrated in accordance with the AAPM TG142 (Klein et al 2009). The MLC consists of two banks of 60 leaves: the central 40 leaves of each bank are 0.5 cm in width (at the isocenter plane) and the outer 20 leaves are 1.0 cm in width. The on-board EPID was set in a SSD of 100 cm and the gantry angle was set to 0°. Flat-field and dark-field calibrations have been performed for the EPID running in both cine mode and integration mode. The axis of the CRIS cylinder was centered to the isocenter. The measured dataset was divided for network training and validation.

The training datasets were collected from the CRIS and the EPID running in cine-mode. The planned beam shapes are shown in figure 5, where the circular and comb-like shapes are designed for learning the inter-leaf and intra-leaf information, respectively. Image pairs were acquired for 12 MLC rotations angles from 0° to 180° at step size of 15°. Random rotations within a range of 15° were simulated in image post-processing as complementary data augmentation. To demonstrate the capability of synthesizing high-quality EPID images, test images were acquired in the integration mode, where artifacts were already eliminated using the linac built-in software. The test dataset consists of regular fields and three clinical IMRT cases delivered at the gantry angle fixed to 0°. The regular fields include a set of MLC-shaped square fields: 2×2 cm^2, 4×4 cm^2, 6×6 cm^2, 8×8 cm^2 and 10×10 cm^2. The clinical IMRT cases include a brain, a lung, and aprostate case, with the number of fields (and step-and-shot segments) being 6 (136), 6 (142), and 7 (77), respectively. The ranges of field sizes are 8.0×7.5 cm^2–13.0×8.0 cm^2, 8.0×6.5 cm^2–11.5×7.0 cm^2, and 9.0×6.5 cm^2–11.1×6.5 cm^2, for the brain, lung, and prostate cases, respectively. The enhanced CRIS images and EPID images were registered by aligning the intensity centroids in the 10×10 cm^2 field images. Furthermore, pixel intensity of the enhanced CRIS images was calibrated to that of the EPID image for a 10×10 cm^2 field, which is based on the assumed linear relation between the intensities measured from CRIS and EPID. In reality, the Gd_2O_2S-based scintillator materials used have been widely accepted for dose linearity.

2.4. Evaluation metrics

In validation, the synthesized EPID images were compared to the ground truth EPID images that were acquired in the integration-mode with higher image quality than those from the cine-mode EPID. The similarity is quantized using mean squared error (MSE), peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), gamma index (γ) (Low and Dempsey 2003), and pixel deviation (σ). σ is defined as the relative difference of pixel intensity between the CRIS and EPID images, which is calculated as the ratio between the absolute difference and the maximum intensity of the EPID image. The gamma analysis was performed for absolute intensity or fluence comparison. EPID images shown in the result section were normalized for display purpose only. Additionally, signal-to-noise ratio (SNR), which was calculated following the method in (Mooslechner et al 2013), is used to evaluate the performance of synchronization artifacts suppression. Furthermore, the gamma passing rates were calculated (γ_{pass}) using various gamma criteria of 1% (global intensity)/1 mm (distance-to-agreement) and 2%/2 mm with a low-intensity cut-off threshold value of 10%. The pass criteria were that more than 90% pixels should have a γ less than one.

3. Experiments and results

3.1. Verification of images captured for regular fields

An evaluation was performed for a set of square fields with side lengths of 2,6,8 and 10 cm. We assess the suppressions of the artifacts as well as the improvements of image quality.

3.1.1. Suppression of mirror-glare and synchronization artifacts—The mirror glare is caused by the inter-reflections between the mirror and the phosphor screen, and is aggravated with the reduced distance in between. In our case, the glare artifacts are observed

on top of the image (see figure 6(a)), where the detected photons have experienced more scatterings due to the shorter distance. With an increase in beam size, the deviation inside the glare artifact regions increases, e.g. up to 40% pixels have a pixel deviation (σ) more than 10% for a 10 × 10 cm^2 field. In the EPID images acquired in the cine-mode, there are obvious synchronization artifacts manifested as the gridding shadow, accounting for up to 2% pixel deviation depending on the beam size. These artifacts are distinguished from the vertical inter-leaf leakage stripes in the ground truth EPID images, which are caused by the gaps between the leaves. In the enhanced CRIS images obtained using cRI-GAN, glare artifacts are effectively suppressed. Comparing to the original CRIS image captured for a 10 × 10 cm^2 field, the number of pixels in the enhanced counterpart that have a pixel deviation (σ) larger than 2% reduces to 3.5% and the maximum σ reduces to 4%. SNRs are calculated for the EPID images (acquired in both the cine-mode and the integration-mode) and for the enhanced CRIS images in table 1. It can be seen that the noise in the enhanced CRIS images is substantially reduced as compared to that in the cine-mode EPID and is even lower than that in the ground truth integration-mode EPID. The SNR values of the enhanced CRIS images have a mean increase ratio of 133.5% (range from 26.7% to 221.3%) with respect to those of the cine-mode EPID images. In comparison, the mean increase ratio of the integration-mode EPID images versus the cine-mode EPID images is 118.6% (range from 40.1% to 240.6%), which is in agreement with the previous studies (Roberts et al 2008, Mooslechner et al 2013). The lower degree of improvement in the 10 × 10 cm^2 field could be attributed to the larger area of the glare artifact region as presented in the given field of view.

3.1.2. Quantitative analysis—The improvements of image quality achieved in the resultant CRIS images are further confirmed in terms of MSE, PSNR, SSIM, and γ_{pass}, as listed in table 2. Substantial improvements can be found in MSE and PSNR values. The mean MSE reduction ratio (i.e. the percentage reduction averaged over various field sizes) and the mean PSNR increase ratio are 88.6% and 35.6% respectively. In comparison, SSIM improvements are moderate (mean of 7.8%). Figure 6(b) shows the inter-leaf leakages that are recovered after image enhancement, demonstrating the capability of the algorithm in preserving true details. Deviations with $\gamma \approx 1$ are found surrounding the primary beam from the gamma map in figure 6(e). The synthesis quality is further evaluated via γ_{pass} with 1%/1 mm and 2%/2 mm criterion. The gamma pass rate (1%/1 mm) decreases with increasing field size due to the increasing area of the glare artifact region. The mean γ_{pass} for the processed CRIS images are 83.2% (1%/1 mm) and 99.1% (2%/2mm), compared to 62.1% (1%/1 mm) and 84.8% (2%/2 mm) for the original images. Caused by the glare artifacts, γ_{pass} for the 10 × 10 cm^2 original field is 19.3% (1%/1 mm) and 42.3% (2%/2 mm).

3.2. Verification of images captured for IMRT cases

We further explored a preliminary application on patient-specific QA for three IMRT cases by comparing the captured CRIS images with those collected from the calibrated EPID running in integration mode. As an example, comparisons among the original CRIS image, enhanced CRIS image, and corresponding EPID images are shown in figure 7 for the fourth field in the prostate case. The CRIS image is an integration of multiple frames captured for every control point in an IMRT field, and the image enhancement was performed frame by

frame. The glare artifact is visible on top of the original CRIS image, accounting for a pixel deviation of 2.6%. The glare artifact is less obvious comparing to that in figure 6(a), since multiple small beam fields that were involved in the integrated result are relatively far from the top regions. The original CRIS image also shows blurred edges and missing details on the inter-leaf leakages which was enhanced in the figure (b). In the gamma map in figure 7(d), deviations ($\gamma \approx 1.0$) can be mainly found in the penumbra regions surrounding the primary beam, which agrees with the deviation distribution for regular field case as shown in figure 6(e). Additional deviations can be found both inside the primary beam region and the region pointed with a black arrow in figure 7(d). By checking the settings in TPS, the pointed region is covered by the secondary collimator jaws. The gamma analysis results of the IMRT cases using 2%/2 mm criterion are summarized in table 3, where the last column shows the averaged gamma passing rates over the whole fields in each case. The passing rates for every single fields range from 92.9% to 99.5%, and the mean passing rates are all above 95%.

3.3. Application on MLC leaf-end positioning

To investigate the measurement sensitivity with respect to leaf end position, twenty leaf pairs with various displacement errors (in a sinusoid pattern with an amplitude of 0.3 mm) were used to deliver a field of $100.0 \times 50.6\ mm^2$ as depicted in figure 8. Synthesized images were interpolated with an upscaling factor of 5 to achieve a sub-pixel positioning accuracy (the original pixel size is 0.44 mm). The leaf end positions were determined as those corresponds to 50% maximum value on edge of the rectangular field. The 0-mm position (indicated with a dark dashed line in figure 8(a)) is aligned to the center axis of the field of view that is calibrated via the fiducial points located on the outer surface of the phantom. The leaf position obtained from the enhanced CRIS images and the on-board EPID images acquired using the integration-mode as well as the values in the treatment planning system (TPS) are compared in figure 8(b). The discrepancy between the measurements from CRIS and EPID (a mean of 0.099 mm ± 0.072 mm with 0.292 mm maximum) is smaller than the difference between CRIS measurement and TPS calculation (a mean of 0.391 mm ± 0.292 mm with 0.495 mm maximum). It is noteworthy that, the MLC motion is not perfectly consistent in two deliveries. In accordance with AAPM TG 142 report, MLC leaf position repeatability has a tolerance of ± 1 mm (Klein et al 2009). A standard deviation of 0.19 mm from ten measurements was reported in our monthly QA. Further quantitative investigation on the difference between CRIS and EPID images is demonstrated in the histogram (see figure 8(c)), which shows the percentage of pixels that have a difference above the specified values. Only 7% leaf pairs show position deviations higher than 0.17 mm between those measured by the CRIS images and the EPID images.

3.4. Ablation analysis of cRI-GAN

Ablation analysis is performed to demonstrate the advantages of task-aware perceptual modules by removing one or two task-specific modules from the classic cRI-GAN design. The advantages are also demonstrated by comparing to modified cRI-GANs with traditional losses including a pixel-based style loss

$$\widehat{\mathscr{L}}_{styl} = \mathbb{E}[\|y - G(x)\|_1], \tag{5}$$

and a pixel-based adversarial loss

$$\widehat{\mathscr{L}}_{adv} = \mathbb{E}_{y \sim P_{\mathrm{EPID}(y)}}[\log D(y)] + \mathbb{E}_{x \sim P_{\mathrm{CRIS}(x)}}[\log(1 - D(G(x, c_n)))]\,. \tag{6}$$

The derived models with various combinations of losses are listed in table 3. The same data were used as in figure 6, and all the trainings were terminated upon 400 epochs. Quantitative analysis of SNR, PSNR, SSIM and γ_{pass} are given in table 4. In all the network models, the classic cRI-GAN ($\mathscr{L}_{adv} + \mathscr{L}_{cont} + \mathscr{L}_{styl}$) and the conventional GAN ($\widehat{\mathscr{L}}_{adv}$) show the best and worst results, respectively, demonstrating the importance of additional constraints (in conjunction with the adversarial loss) as well as the effect of the perceptual loss. Among the three modules, $\mathscr{L}_{styl}$ contributes most, whereas $\mathscr{L}_{adv}$ and $\mathscr{L}_{cont}$ make similar contributions. For example, the combination of $\widehat{\mathscr{L}}_{adv} + \mathscr{L}_{cont} + \mathscr{L}_{styl}$ yields similar results as $\mathscr{L}_{adv} + \mathscr{L}_{styl}$, slightly inferior to the classic model. The limited improvement of $\mathscr{L}_{adv}$ could be explained by the fact that the features ($\mathscr{F}_2$) exported from the pretrained VGG16 layers could be learned to certain extent by the discriminator (D) upon sufficient training. The contribution of $\mathscr{L}_{cont}$ is limited by $\mathscr{L}_{styl}$ and $\mathscr{L}_{adv}$, since the target structure (mirror-glare region) in $\mathscr{L}_{cont}$ is partially included in the dilated leaf ends and leaf edges in $\mathscr{L}_{styl}$ and $\mathscr{L}_{adv}$, respectively. Furthermore, $\mathscr{L}_{cont}$ is assigned with a small weight (λ_c) as shown in equation (4), which is based on the concern that a pixel-based minimization might adversely affect image synthesis via over smoothing (Ledig et al 2017).

The training curves of these modified models are compared in figure 9, where the three cases with the top performance (i.e. $\mathscr{L}_{adv} + \mathscr{L}_{styl}$, $\widehat{\mathscr{L}}_{adv} + \mathscr{L}_{cont} + \mathscr{L}_{styl}$ and the classic cRI-GAN) are presented. The classic model shows the highest convergency accuracy and the most stable training process, which could be attribute to the minimization of the three perceptual losses calculated on low- and mid-level feature maps with a reduced dimension and compressed effective pixels. The peaks in the curves could be explained by the small training batch size (4), which is limited by the GPU memory.

3.5. Comparison with other state-of-the-art deep networks

The proposed cRI-GAN is compared with two state-of-the-art GANs containing task-specific attention modules. One is the edge-aware GAN (Ea-GAN), which incorporates the edge information into its generator and discriminator to enhance synthesis quality (Yu et al 2019). The other is the enhanced super-resolution GAN (ESR-GAN), which achieves high visual quality by exploiting improved network architecture, adversarial loss and perceptual loss (Wang et al 2018). Since these works are not intended for beam visualization, we did not directly run the models. Instead, we reserved their key ideas while adapting the same network backbone (including the structures of generator and discriminator) and training strategies (including the training data and optimization method) as ours for fair comparison.

An experiment was performed with the data collected for an arbitrary field shape as shown in figure 10. The quantitative results are presented in table 5. In the original CRIS image, adversities from light scattering are obvious, including edge blurring and mirror-glare artifact located on top of the image. In ESR-GAN, a steep edge gradient and moderate inter-leaf leakage recovery are obtained, yet the penumbra regions are totally lost. In the perceptual loss of ESR-GAN, the high-level features were the output from the last convolution layers in the last block of a pretrained VGG19. While the highly abstractive information neglects the details to some degree in the low-dose regions, i.e. those outside the primary beam. In comparison, Ea-GAN gives a better result with much more recoveries in the low-dose regions, but with limited suppression on the glare artifacts indicated with a red arrow. This could be explained by the edge extraction manner in Ea-GAN, which is based on a conventional gradient-based approach. In this way, the edges were extracted not only from the desirable targets, but also from the artifacts, including mirror-glare artifacts and synchronization artifacts. The subsequent minimization between EPID images and CRIS images would lead to unexpected recovery of the overlapped textures between the mirror-glare artifacts and the synchronization artifacts. In cRI-GAN, the leaf edge is extracted from a DNN (a VGG16 pretrained on ImageNET), which has the potential to identify the desirable target through sufficient training. The feature maps extracted from CRIS images ($\mathcal{F}_2(x)$) or EPID images ($\mathcal{F}_2(y)$) showed that limited weights were assigned to the edges of the artifacts. Furthermore, the DNN-extracted edges have an dilation effect on the boundaries (see figure 10), which has been demonstrated necessary in GAN-based SR applications to benefit the convergence and synthesis quality (Rad et al 2019).

4. Discussion and conclusion

We have developed a low-cost CRIS for high-quality beam visualization in external beam radiotherapy. The cylindrical receptor design provides a pathway that allows co-planar measurement and detects the errors caused by the gantry angle. Moreover, the independent measurement potentially enables an end-to-end verification of the image-guided treatment. The primary cost of the system depends on both the CMOS camera and the scintillator sheet, which are typically much cheaper than array detector-based 2D and especially 3D devices such as ArcCHECK™ and Delta4™. For a proof-of-concept demonstration, the system is currently demonstrated using a fixed gantry angle due to the limited size of the scintillator sheet used. Once the scintillator sheet is extended to cover the entire inner surface of the cylindrical receptor, a co-planar detection could be achieved by aligning the imaging center to that of the hemispheric mirror and the gantry isocenter. In that case, the training data and deep learning model used here will still be valid due to the radially symmetric design. Any latent asymmetric factors caused by the camera-lens components could be removed via performing a flat-field correction.

To enhance the robustness of cRI-GAN against the latent noise in training data, task-aware perceptual modules were incorporated into the developed network architecture. A selective learning is formulated to avoid the disturbance from artifacts existing in the training images. Compared to pixel-oriented learning strategies widely adapted in traditional SR models, learning in feature domain enables to extract abstractive information that is less sensitive to

field size variations due to the limited mechanical consistency. We also demonstrated that the networks with loss functions in feature domain (namely, perceptual loss) achieve higher convergency accuracy and more stable training process. The validations of applying DNN-based the attention mechanisms on critical beam structures were verified by comparing to that using analytical method (e.g. the edge extraction in Ea-GAN). The outperformance of cRI-GAN over several state-of-the-art SR networks was validated in terms of general similarity metrics and gamma passing rate.

The design ideas of the proposed deep learning model can be applied to generic networks. First, a joint regularization over hierarchical perceptual losses is likely to outperform an individual regularization. As demonstrated in our application, a variety of features that represent the critical structures or textures were extracted at different network layers and made their own contribution to the corresponding task objectives in the subsequent joint optimization. Second, different task-specific structures can be represented in the feature maps extracted from the shallow layers in a VGG16 network, which was pretrained with irrelevant dataset for the classification purpose. While this is only validated in the specific application, we believe that feature maps extracted using pretrained networks can be utilized for desirable representations so as to form task specific losses.

As a demonstration of our system, we performed our experiments on a Varian linac equipped with a Millennium MLC. Results were compared to those collected from an EPID running in integration mode, which generally provides images with higher quality than those from a cine-mode EPID. Quantitative analysis was performed using general similarity metrics and gamma index. Improvements on artifacts compression, spatial resolution and details recovery were demonstrated for various regular fields ranging from 2×2 cm^2 to 10×10 cm^2 and three IMRT cases. Our results agree well with those from an integration-mode EPID with a mean gamma passing rate (2%/2 mm) of 99.1% and 97.1% for the regular fields and three IMRT cases, respectively. In the gamma maps (see figure 6(e) and figure 7(d)), failures are mainly found surrounding the primary beam, which could be attributed to both the limited reproducibility of MLC leaf positioning and accuracy of predictions in the penumbra regions. Gamma failures are also observed inside the beam region as shown in figure 7(d), where the imaging results for each IMRT field is an integration over all the controls points, overlapping the beam and penumbra regions. Current training datasets were collected with the secondary collimator jaws retracted, and thus the influence from the jaw position was not learned by cRI-GAN. Since the jaws follow MLC in the IMRT treatment plans, differences were found in the low-dose regions (see the pointed region in figure 7(d)). However, this difference is rather limited (a maximum gamma index of ~0.8) and mostly appears in the cut-off regions (less than 10% threshold) defined by typical gamma evaluations. This adversity maybe more pronounced for those scenarios with extremely low doses. Additional training could be performed in the future with the effect of the jaw setting taken into consideration.

We further designed an experiment to apply our system as a tool for MLC leaf-end verification. The positions read from the images acquired by our system are very consistent to those from an EPID (0.099 ± 0.072 mm). The superiority of the proposed deep learning network is confirmed with comparisons to other networks and ablation experiments.

Benefitting from the cylindrical geometry and high-fidelity beam visualization, the developed system could be a useful tool for various machine QA, patient-specific QA and MLC QA where high-quality beam visualization is required.

Despite the promising results achieved in this study, limitations exist. First, the mirror-glare artifacts are not completely suppressed. As observed from figure 6(b), some artifacts remain on top of the synthesized image although the percentage deviation is as low as 0.1%. This was caused by the supervised learning that aims at maximizing the similarity between the input image and the ground truth in an end-to-end training way. Consequently, the original structures and textures including those of the artifacts could be more or less inherited in the enhanced CRIS images due to the noise in the ground truth or cine-mode EPID images. The second pitfall is caused by the slightly unpaired training data. Limited by the practical setup, the acquisitions of EPID images and CRIS images were conducted in separate experiments, where MLC motions were not perfectly consistent.

To mitigate both problems, an unsupervised network architecture with the incorporation of task-specific modules could be employed to facilitate more reliable image domain translation and allow for learning with unpaired data in the future. For example, the cycle consistent loss, which is widely used in unsupervised image-to-image translation networks, enforces the network to learn a mapping from the target domain (EPID images) to the source domain (CRIS images) (Zhu et al 2017, Choi et al 2018). In this way, the network will be trained to catch all the details in the CRIS images, including the mirror-glare artifacts, which, in turn, could benefit the image synthesis. By incorporating the learned task-specific features into the cycle consistent losses, better results could be expected. For both the CRIS and the EPID, flat-field corrections are required to remove the inherent characteristics of the detectors such as the pixel sensitivity. Ideally, the corrections can be performed for an absolute flat radiation field. In practice, additional corrections can be applied for the non-uniformity of the calibration radiation field. In this way, once the CRIS is calibrated and trained in one machine, it would reflect the true characteristics of the beam delivered by other machines. This study is mostly focusing on the imaging side without consideration of the additional corrections mentioned.

In summary, a novel CRIS is developed with the image quality comparable to that of an EPID running at the integration mode. The cylindrical receptor design enables independent measurement with co-planar detection ability. Benefiting from the high-quality image, low cost and streamlined data collection, the system promises to be a practical tool that provides reliable measurement for dosimetric and geometrical verifications in radiation oncology practice.

Acknowledgments

This work was partially supported by NIH/NCI (1R01CA176553 and 1R01CA227713) and a Faculty Research Award from Google Inc.

References

Cheon W, Kim SJ, Kim K, Lee M, Lee J, Jo K, Cho S, Cho H and Han Y 2019 Feasibilityof two-dimensional dose distribution deconvolution using convolution neural networks Med. Phys 46 5833–47 [PubMed: 31621917]

Choi Y, Choi M, Kim M, Ha J-W, Kim S and Choo J 2018 Stargan: Unified generative adversarial networks for multidomain image-to-image translation Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition pp 8789–97

Collomb-Patton V, Boher P, Leroux T, Fontbonne J-M, Vela A and Batalla A 2009 The dosimap, a high spatial resolution tissue equivalent 2d dosimeter for linac qa and imrt verification Med. Phys 36 317–28 [PubMed: 19291971]

Feygelman V, Zhang G, Stevens C and Nelms BE 2011 Evaluation of a new VMAT QA device, or the 'X' and 'O' array geometries J. Appl. Clin. Med. Phys 12 146–68

Frelin A-M, Fontbonne J-M, Ban G, Colin J, Labalme M, Batalla A, Vela A, Boher P, Braud M and Leroux T 2008 The dosimap, a new 2d scintillating dosimeter for iMrt quality assurance: Characterization of two čerenkov discrimination methods Med. Phys 35 1651–62 [PubMed: 18561640]

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative adversarial networks (PDF) Proc. of the Int. Conf. on Neural Information Processing Systems (NIPS 2014) pp 2672–80

Guillot M, Beaulieu L, Archambault L, Beddar S and Gingras L 2011 A new water-equivalent 2d plastic scintillation detectors array for the dosimetryof megavoltage energy photon beams in radiation therapy Med. Phys 38 6763–74 [PubMed: 22149858]

Jenkins CH, Naczynski DJ, Shu-Jung SY and Xing L 2015 Monitoring external beam radiotherapy using real-time beam visualization Med. Phys 42 5–13 [PubMed: 25563243]

Jenkins CH, Naczynski DJ, Shu-Jung SY, Yang Y and Xing L 2016 Automating quality assurance of digital linear accelerators using a radioluminescent phosphor coated phantom and optical imaging Phys. Med. Biol 61 L29–37 [PubMed: 27514654]

Kingma DP and Ba J 2014 Adam: a method for stochastic optimization arXiv:1412.6980

Klein EE, Hanley J, Bayouth J, Yin F-F, Simon W, Dresser S, Serago C, Aguirre F, Ma L, Arjomandy B et al. 2009 Taskgroup 142 report: quality assurance of medical accelerators Med. Phys 36 4197–212 [PubMed: 19810494]

Korreman S, Medin J and Kjaer-Kristoffersen F 2009 Dosimetric verification of RapidArc treatment delivery Acta Oncol. 48 185–91 [PubMed: 18777411]

Kupyn, O; Deblurgan: blind motion deblurring using conditional adversarial networks; Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition; 2018.

Ledig C et al. 2017 Photo-realistic single image super-resolution using a generative adversarial network Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition pp 4681–90

Lee M, Ding K and Yi B 2018 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 63 20NT03

Low DA and Dempsey JF 2003 Evaluation of the gamma dose distribution comparison method Med. Phys 30 2455–64 [PubMed: 14528967]

Macdonald RL, Alasdair S and Thomas C G 2020 Systems and methods for planning and controlling the rotation of a multileaf collimator for arc therapy US Patent 10,525,283

Mooslechner M, Mitterlechner B, Weichenberger H, Huber S, Sedlmayer F and Deutschmann H 2013 Analysis of a free-running synchronization artifact correction for mv-imaging with asi: H flat panels Med. Phys 40 031906 [PubMed: 23464321]

Rad M, Bozorgtabar B, Marti U, Basler M, Ekenel H and Thiran J 2019 Srobb: targeted perceptual loss for single image super-resolution Proc. of the IEEE Int. Conf. on Computer Vision pp 2710–9

Radford A, Metz L and Chintala S 2015 Unsupervised representation learning with deep convolutional generative adversarial networks arXiv:1511.06434

Roberts D, Hansen V, Niven A, Thompson M, Seco J and Evans P 2008 A low z linac and flat panel imager: comparison with the conventional imaging approach Phys. Med. Biol 53 6305–19 [PubMed: 18936518]

Simonyan K and Zisserman A 2014 Very deep convolutional networks for large-scale image recognition arXiv:1409.1556

Van Nieuwenhove V, De Beenhouwer J, De Carlo F, Mancini L, Marone F and Sijbers J 2015 Dynamic intensity normalization using eigen flat fields in x-ray imaging Opt. Express 23 27975–89 [PubMed: 26480456]

Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Qiao Y and Change Loy C 2018 Esrgan: enhanced super-resolution generative adversarial networks Proc. of the European Conf. on Computer Vision (ECCV) pp 0–0

Yu B, Zhou L, Wang L, Shi Y, Fripp J and Bourgeat P 2019 Ea-gans: edge-aware generative adversarial networks for crossmodality mr image synthesis IEEE Trans. Med. Imaging 38 1750–62 [PubMed: 30714911]

Zhu J-Y, Park T, Isola P and Efros AA 2017 Unpaired image-to-image translation using cycle-consistent adversarial networks Proc. of the IEEE Int. Conf. on Computer Vision pp 2223–32

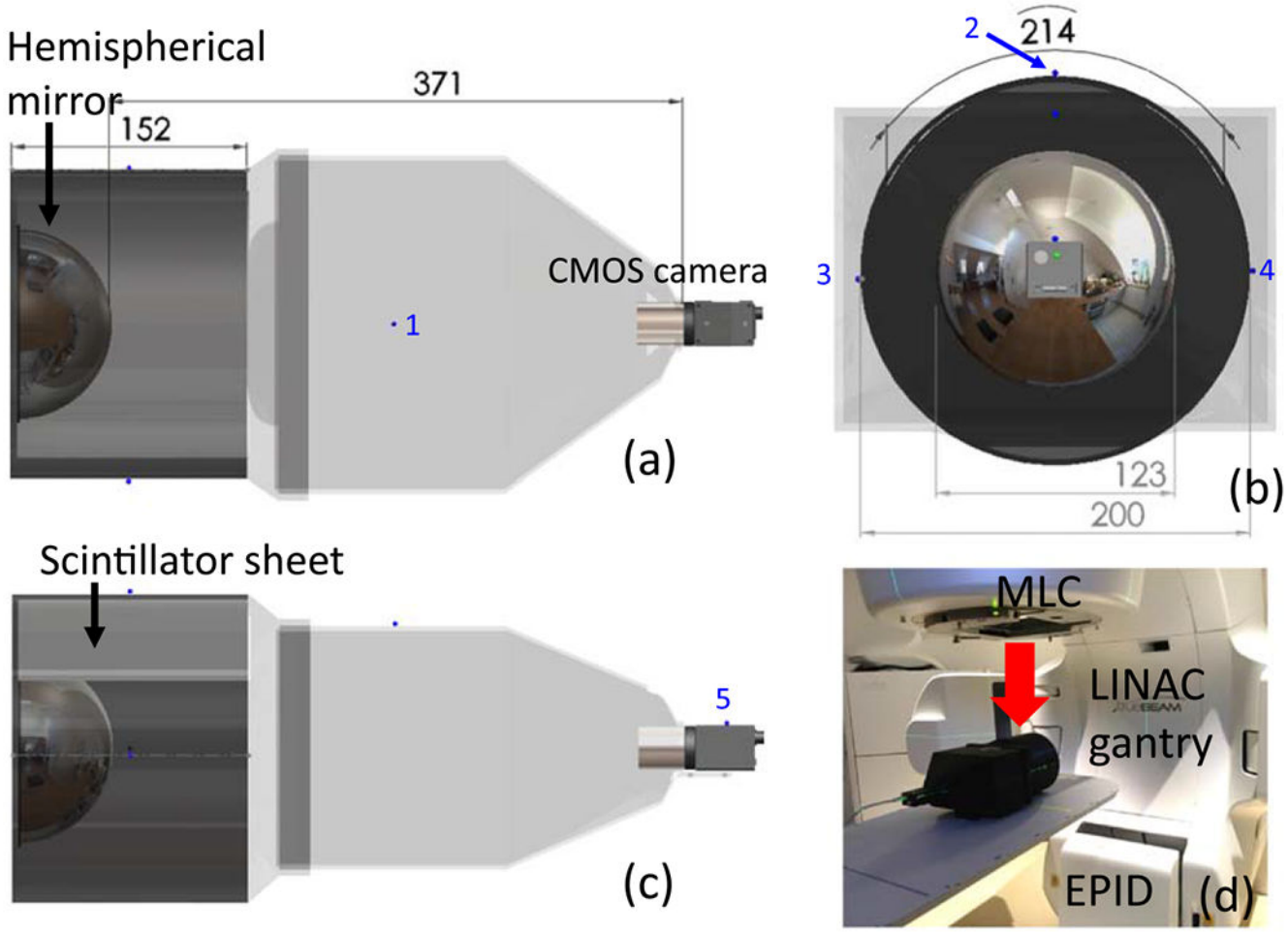

Figure 1.

A schematic diagram of our CRIS in (a) anteroposterior view, (b) axial view, and (c) lateral views. This device consists of a scintillator sheet, a hemispheric mirror and a digital camera. The scintillator sheet is overlaid to the inner surface of a cylinder; and the hemispheric mirror reflects the emitted visible light to the camera. Dimensions are indicated in a unit of millimeter. Panel (d) shows a photo illustrating the practical setup of the system during data acquisition. Five fiducial points denoted with blue dots are distributed on both the phantom and the camera.

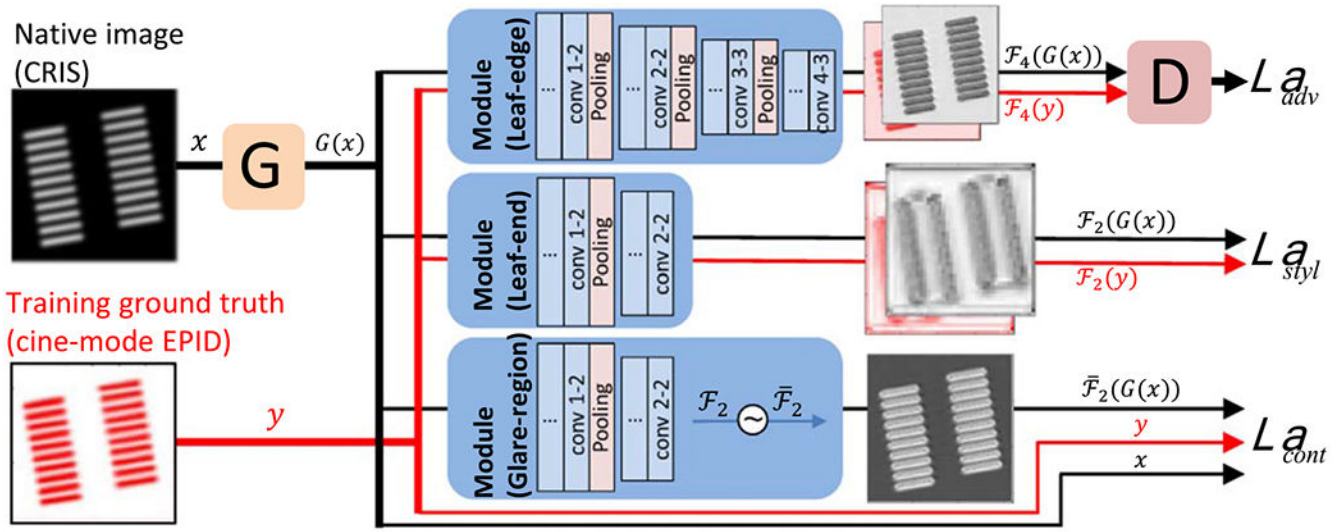

Figure 2.

Illustration of the proposed cRI-GAN framework for radioluminescence image enhancement. The whole framework consists of a generator (G), a discriminator (D) and three task-aware perceptual modules. The module networks pretrained for image classification were used to extract discriminative features between the weak labels and the generated images. The leaf-edge module imports low-level feature to discriminator for adversarial learning; leaf-end module outputs mid-level features for style learning; and the glare-region module outputs a complementary feature map, together with original x and y for content learning. By jointly optimizing on these modules, noise is removed from the synthesized images while high resolution is preserved.

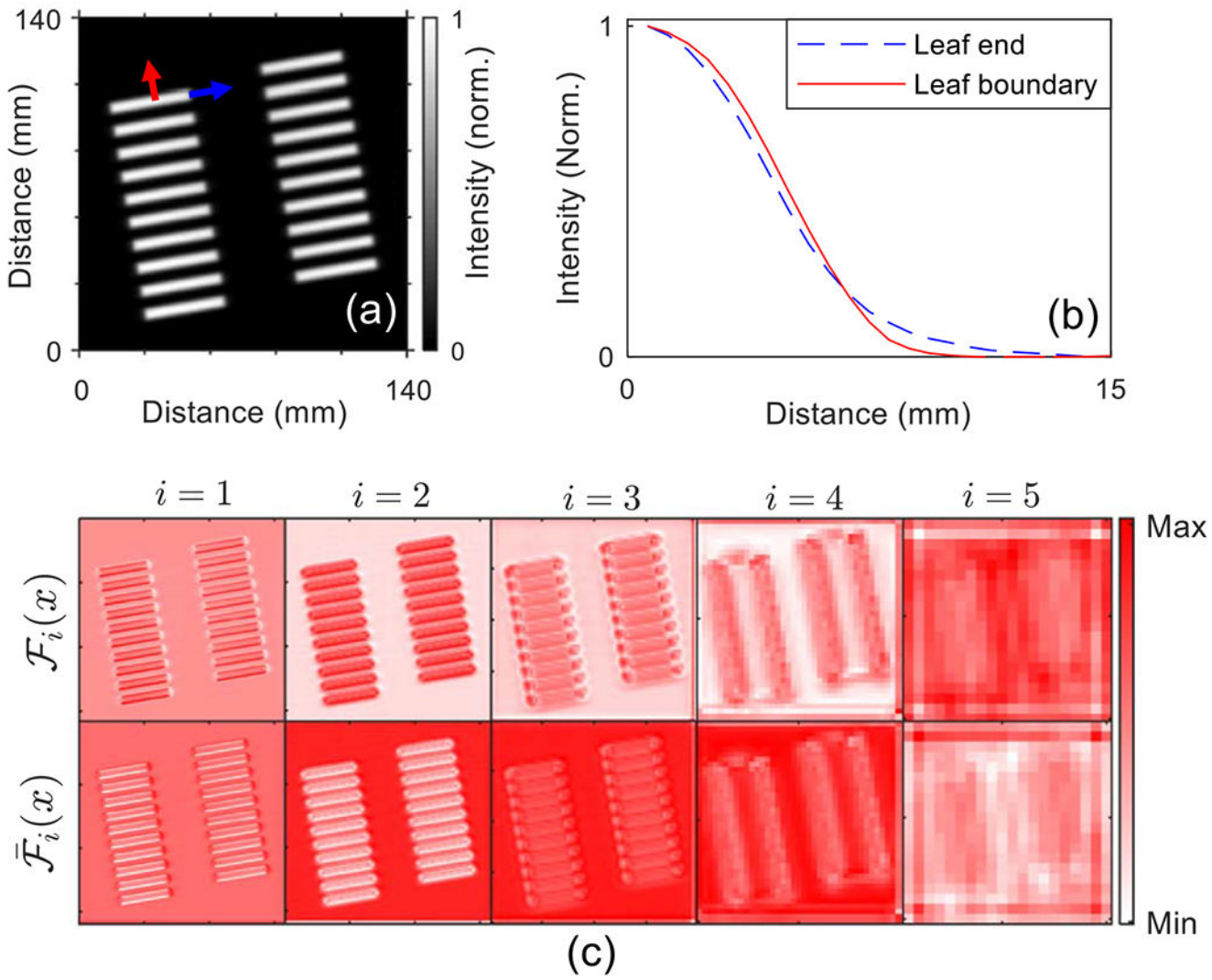

Figure 3.

(a) An example CRIS image, (b) edge spread functions along the leaf end and inter-leaf boundary indicted with blue and red arrows in (a), and (c) feature maps ($\mathcal{F}_i(x)$) showing the representations on the desired structure. $\mathcal{F}_i(x)$ is the integration over the output from the last convolutional layer before ith maxpooling in a pretrained VGG16.

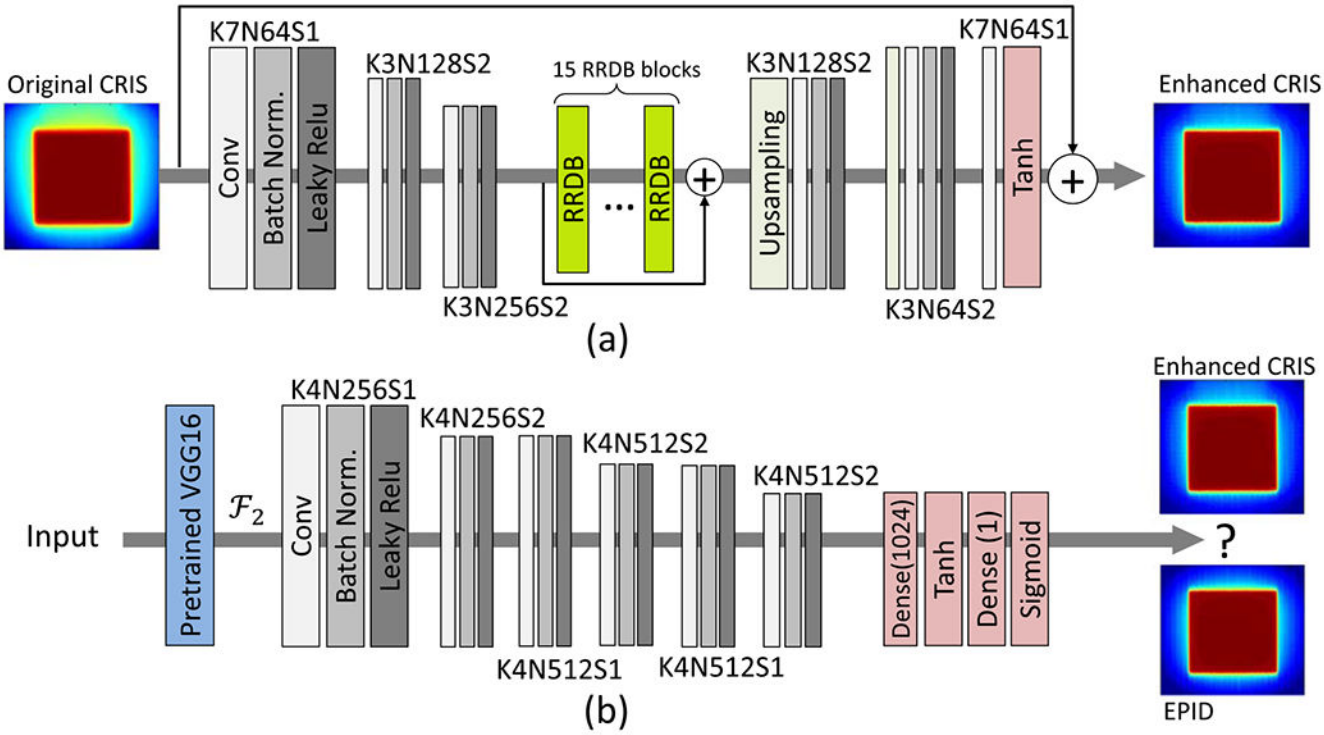

Figure 4.

Architecture of the (a) generator and (b) discriminator subnetworks with corresponding kernel size (K), number of filters (N) and stride size (S) indicated for each convolution layer.

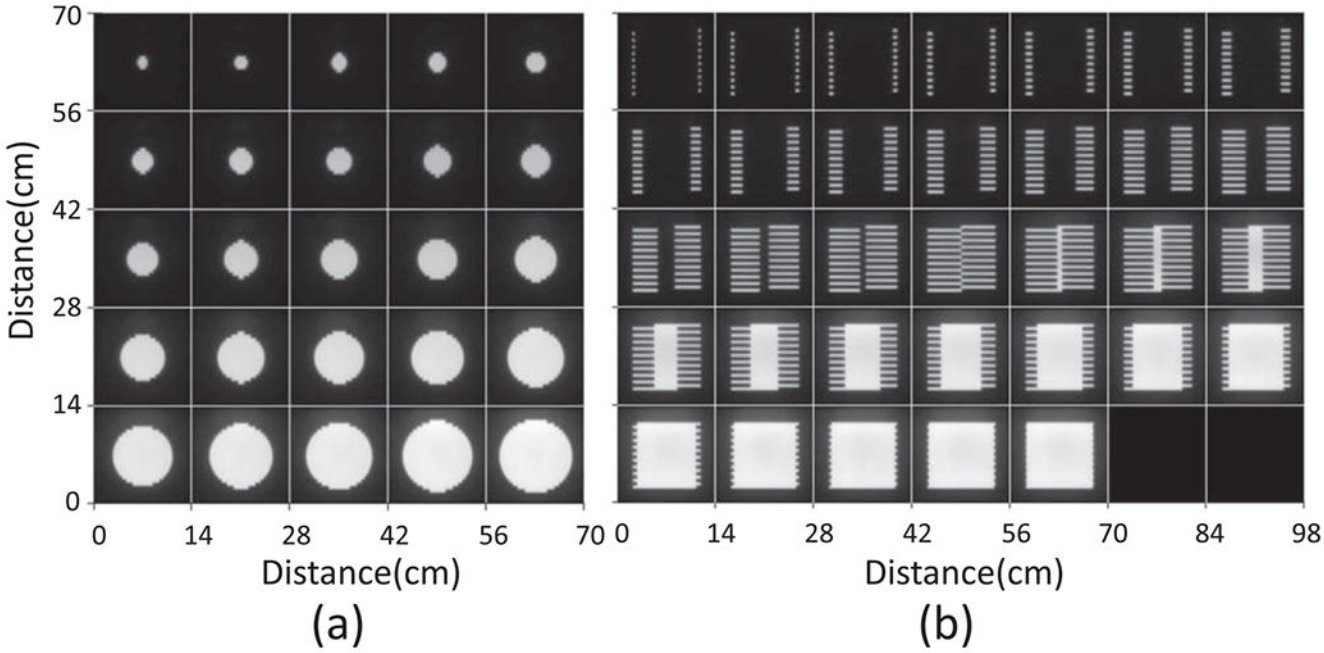

Figure 5.

Images with (a) circular and (b) comb-like shapes for composing the training datasets. The two kinds of beam shapes are intended for learning the intra-leaf and inter-leaf information, respectively.

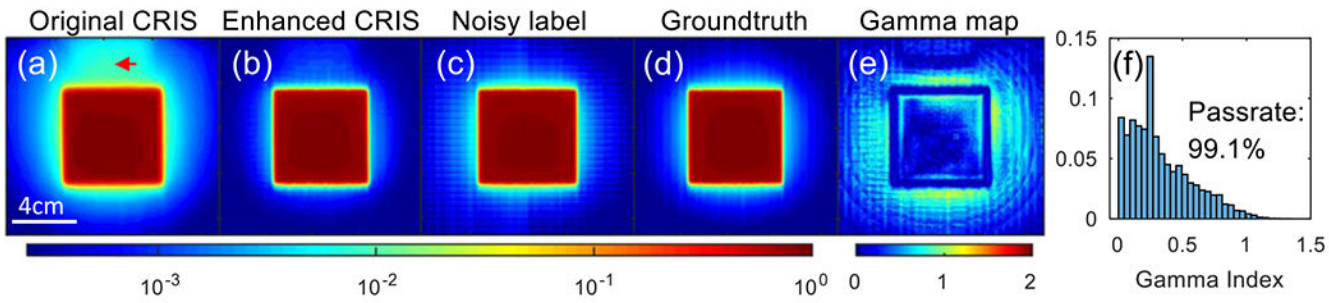

Figure 6.

Demonstration of a CRIS enhanced image using cRI-GAN for a 6 × 6 cm^2 beam case. (a) an original CRIS image, and (b) the enhanced CRIS image, (c) the EPID image acquired in the cine-mode, (d) the EPID image acquired in the integration-mode, (e) the gamma map (2%/2 mm), and (f) gamma histogram. Images are displayed in logscale.

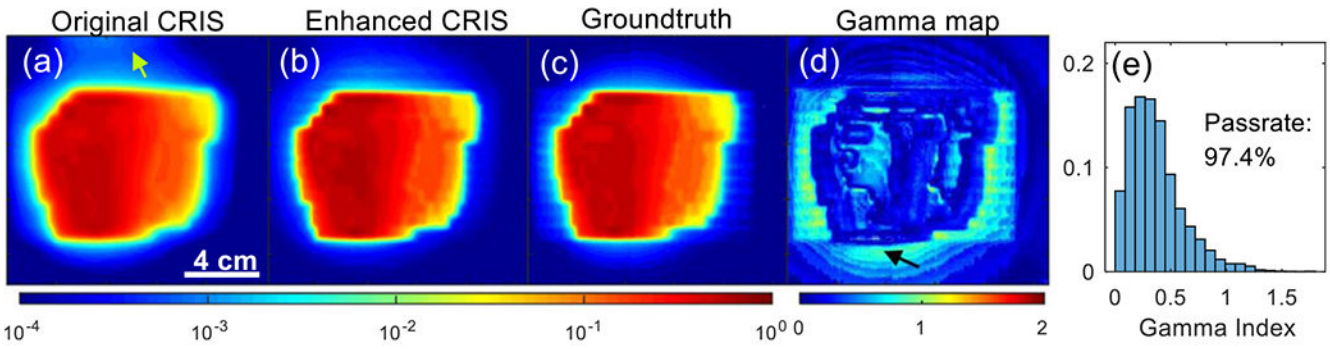

Figure 7.

Verification of enhanced CRIS image captured for a prostate IMRT case (Field 4). (a) The original CRIS image, (b) the enhanced CRIS image, (c) the EPID image acquired in the integration-mode, (d) the gamma map (2%/2mm), and (e) gamma histogram. Images are displayed in logscale.

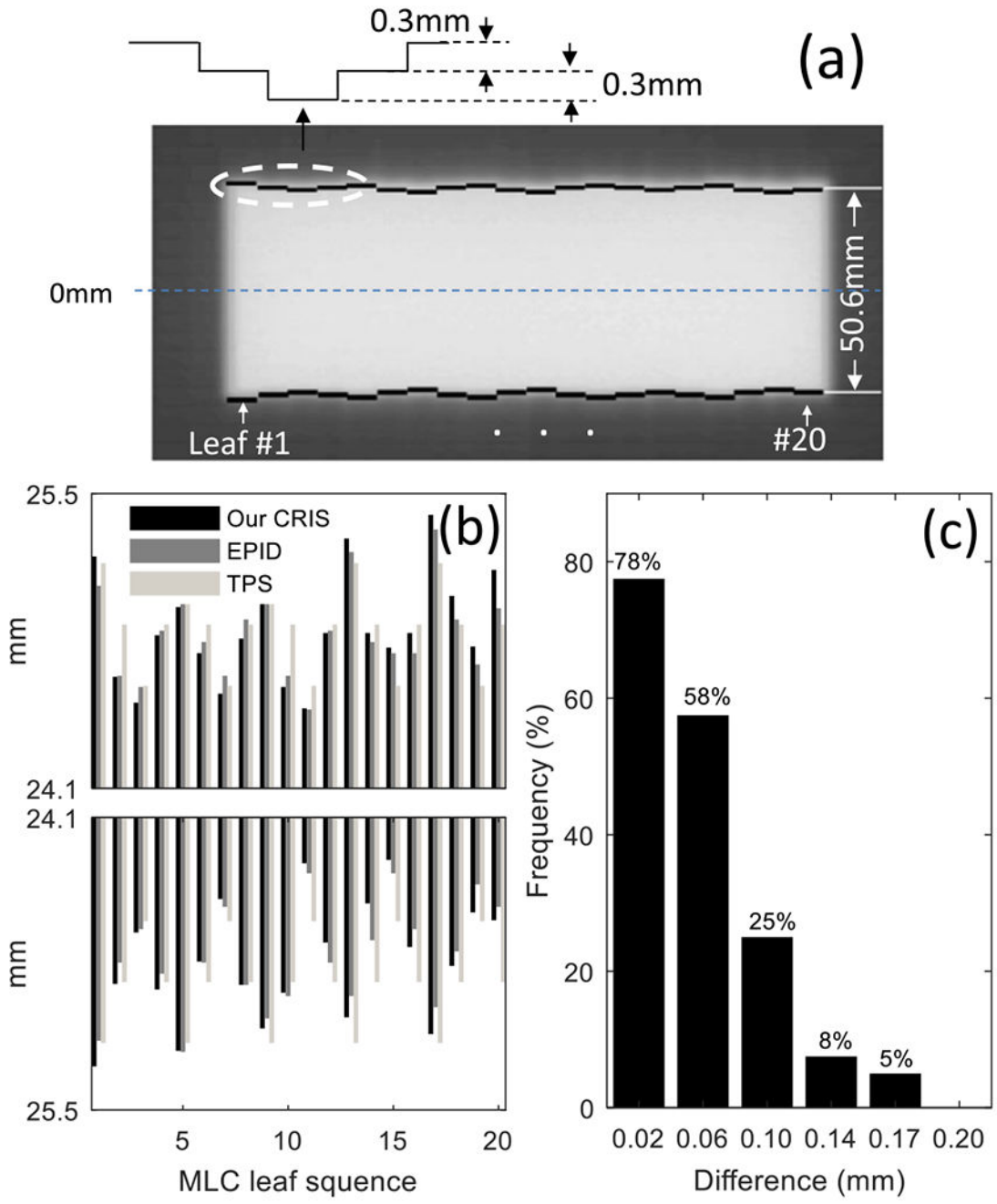

Figure 8.

Verification of MLC leaf end positioning. (a) Geometric description of a rectangular beam with the black lines indicating the planned leaf end positions, (b) comparison between the leaf positions measured by CRIS and EPID (integration-mode) versus TPS planned values, and (c) a histogram showing the difference between CRIS and EPID measurement. Small displacements were designed to test the detection sensitivity.

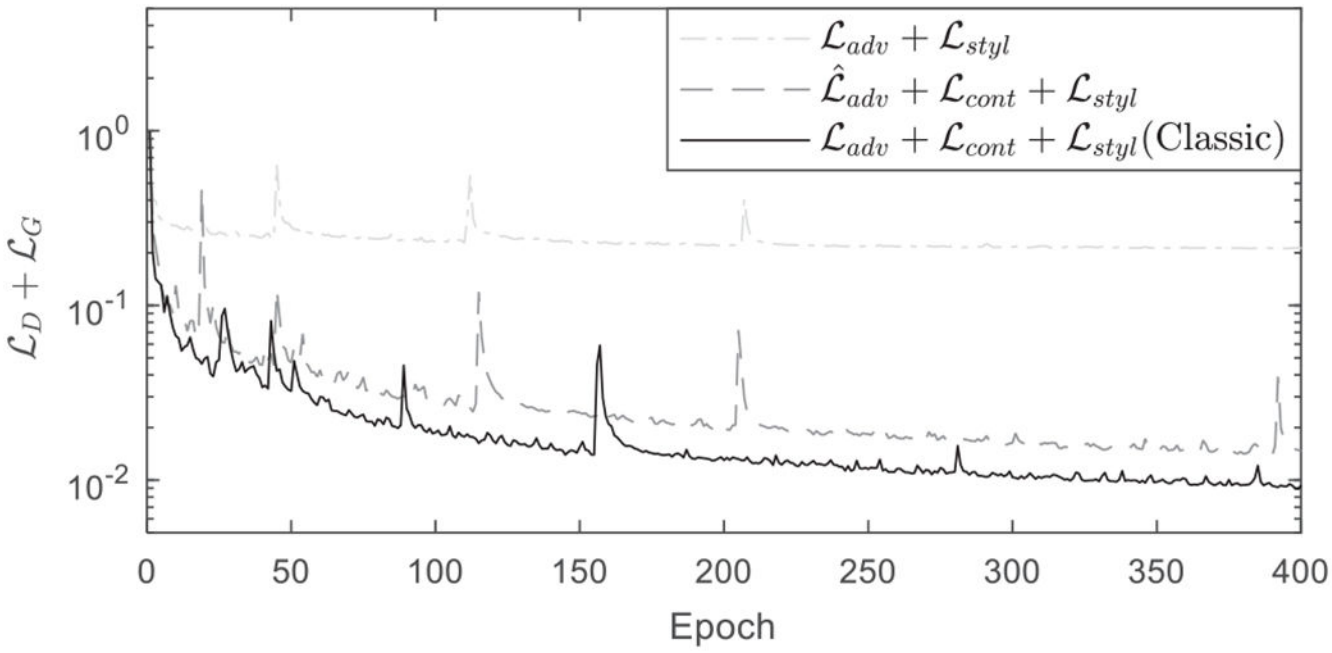

Figure 9.
Training curves of the classic cRI-GAN versus modified versions.

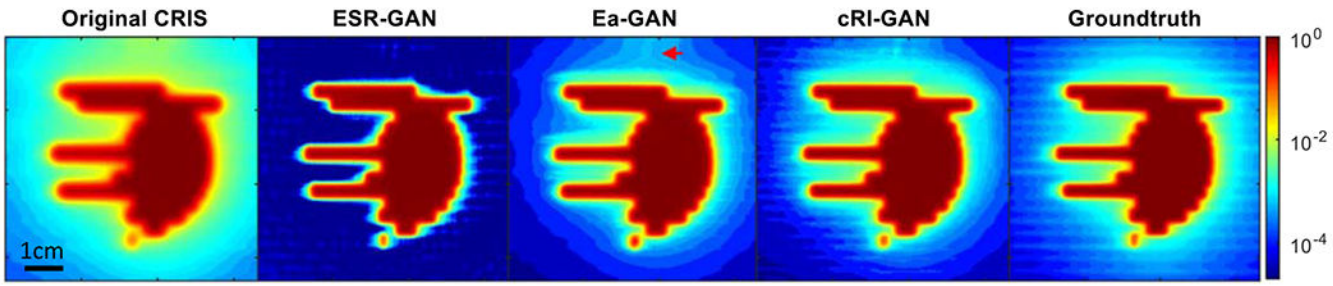

Figure 10.

Comparison of images generated by ESR-GAN, Ea-GAN and our cRI-GAN. The proposed cRI-GAN outperforms the other models in noise suppression and preserving structures of leaf leakage and penumbra region.

Table 1.

SNRs of the images acquired in the cine-mode EPID, the integration-mode EPID and our enhanced CRIS for a set of square fields.

Field sizes	EPID (Cine)	EPID(Integration)	CRIS
2×2 cm^2	16.40	**55.87**	51.31
4×4 cm^2	25.75	77.51	**82.76**
6×6 cm^2	33.29	56.52	**60.11**
8×8 cm^2	41.47	89.00	**93.74**
10×10 cm^2	50.14	**70.25**	63.52

Table 2.

Quantitative evaluations on the enhanced and original CRIS images (in bracket) for a set of square fields.

Field sizes	MSE	PSNR (dB)	SSIM (dB)	γ_{pass} (2%/2 mm)	γ_{pass} (1%/1 mm)
2 × 2 cm^2	4.01 (13.60)	42.13 (36.81)	0.9900 (0.9581)	100% (98.2%)	96.9% (85.8%)
4 × 4 cm^2	3.44 (58.50)	42.83 (30.46)	0.9904 (0.8238)	98.9% (97.7%)	84.1% (77.6%)
6 × 6 cm^2	10.82 (90.63)	37.81 (28.56)	0.9849 (0.8958)	99.1% (97.0%)	82.3% (71.5%)
8 × 8 cm^2	12.71 (191.84)	37.11 (25.30)	0.9935 (0.8328)	98.6% (88.7%)	79.4% (56.5%)
10 × 10 cm^2	21.96 (251.23)	34.73 (24.13)	0.9939 (0.8977)	99.1% (42.3%)	73.1% (19.3%)

Table 3.

Gamma analysis (2%/2 mm) for three IMRT cases.

	Field 1	Field 2	Field 3	Field 4	Field 5	Field 6	Field 7	Mean
Prostate	98.4%	99.5%	97.6%	97.4%	95.0%	99.4%	98.1%	97.9%
Brain	96.8%	97.9%	97.5%	94.7%	92.9%	97.9%	N/A	96.3%
Lung	92.8%	98.3%	98.2%	97.4%	96.3%	99.4%	N/A	97.1%

Table 4.

Quantitative results of ablation experiments.

Loss comb.	MSE	SSIM	PSNR (dB)	γ_{pass} (1%/1mm)
$\mathscr{L}_{adv}$	1894.25	0.4209	13.2	46.2%
$\mathscr{L}_{adv} + \mathscr{L}_{styl}$	49.84	0.9565	31.16	96.4%
$\mathscr{L}_{adv} + \mathscr{L}_{cont}$	138.89	0.90	26.70	84.5%
$\widehat{\mathscr{L}}_{adv}$	1059.61	0.3693	7.88	46.3%
$\mathscr{L}_{adv} + \widehat{\mathscr{L}}_{styl}$	149.52	0.8940	26.44	77.4%
$\widehat{\mathscr{L}}_{adv} + \mathscr{L}_{cont} + \mathscr{L}_{styl}$	44.69	0.9566	31.63	90.5%
$\mathscr{L}_{adv} + \mathscr{L}_{cont} + \mathscr{L}_{styl}$ (Classic)	**30.41**	**0.9750**	**33.30**	**98.8%**

$\widehat{\mathscr{L}}$ means the loss is calculated in the image domain instead of the feature domain.

Table 5.

Quantitative evaluations of the images in figure 9 processed using different models.

	MSE	SSIM	PSNR (dB)	γ_{pass}(1%/1 mm)
Raw CRIS	180.33	0.090	25.93	70.3%
ESR-GAN	168.34	0.6215	26.70	77.7%
Ea-GAN	64.39	0.9313	30.90	86.1%
cRI-GAN (ours)	**57.85**	**0.9531**	**32.58**	**95.2%**