Accurate patient alignment without unnecessary imaging using patient-specific 3D CT images synthesized from 2D kV images

Check for updates

Yuzhen Ding [1], Jason M. Holmes[1], Hongying Feng[1], Baoxin Li[2], Lisa A. McGee[1], Jean-Claude M. Rwigema[1], Sujay A. Vora[1], William W. Wong[1], Daniel J. Ma[3], Robert L. Foote[3], Samir H. Patel[1] & Wei Liu [1] ✉

Abstract

Background In radiotherapy, 2D orthogonally projected kV images are used for patient alignment when 3D-on-board imaging (OBI) is unavailable. However, tumor visibility is constrained due to the projection of patient's anatomy onto a 2D plane, potentially leading to substantial setup errors. In treatment room with 3D-OBI such as cone beam CT (CBCT), the field of view (FOV) of CBCT is limited with unnecessarily high imaging dose. A solution to this dilemma is to reconstruct 3D CT from kV images obtained at the treatment position.
Methods We propose a dual-models framework built with hierarchical ViT blocks. Unlike a proof-of-concept approach, our framework considers kV images acquired by 2D imaging devices in the treatment room as the solo input and can synthesize accurate, full-size 3D CT within milliseconds.
Results We demonstrate the feasibility of the proposed approach on 10 patients with head and neck (H&N) cancer using image quality (MAE: < 45HU), dosimetric accuracy (Gamma passing rate ((2%/2 mm/10%): > 97%) and patient position uncertainty (shift error: < 0.4 mm).
Conclusions The proposed framework can generate accurate 3D CT faithfully mirroring patient position effectively, thus substantially improving patient setup accuracy, keeping imaging dose minimal, and maintaining treatment veracity.

Plain language summary

Effective and accurate imaging guidance is critical for precise patient alignment, accurate tumor tracking, accurate delivery of radiation therapy and to protect organs that should not be irradiated. However, high-quality imaging guidance usually can only be provided following detailed imaging using a large amount of radiation. We propose a computational method that can generate the full size 3D images required as image guidance from X-Ray images. We demonstrated its utility using data from 10 people with head and neck cancer. Our proposed approach can be used by existing treatment machines to improve the accuracy of patient alignment and hence ensure more accurate treatment of patients.

Radiotherapy (RT) is a standard and a favored treatment modality for head and neck (H&N) cancer patients. Among the cutting-edge approaches in RT, intensity-modulated proton therapy (IMPT) distinguishes itself by precisely delivering maximum cell-killing energy to tumors while minimizing exposure to surrounding organs at risk (OARs)[1–4]. Despite its precision, IMPT remains highly susceptible to factors such as patient setup, proton beam range uncertainties, respiratory motion, and inter-fractional anatomical changes[5–11]. These uncertainties may potentially result in undertreatment of tumors or excessive exposure of surrounding OARs, leading to local recurrence and unexpected treatment-related adverse events (AEs)[12–15].

Image-guided patient alignment is an essential step for patient setup in most RT modalities, especially for the more uncertainty-vulnerable IMPT. During treatment, before delivering the prescribed dose, the therapists need to carefully adjust the patient's position and posture to align with the planning CT, on which the treatment plan was designed. Hence, accurate dose delivery depends on accurate patient setup, for which minimizing patient alignment error is vital. In RT, commonly employed on-board imaging (OBI) techniques are CT-on-rails (CToR), cone beam CT (CBCT), and orthogonal kV images, etc. CToR has the diagnostically equivalent imaging quality (same as planning CT) but requires effective transfer of the patient from the CT scanner to the position of treatment. During the

[1]Department of Radiation Oncology, Mayo Clinic, Phoenix, AZ, USA. [2]School of Computing and Augmented Intelligence, Arizona State University, Phoenix, AZ, USA. [3]Department of Radiation Oncology, Mayo Clinic, Rochester, MN, USA. ✉e-mail: Liu.Wei@mayo.edu

transfer, any uncertainty caused by either patients' movement or position/posture discrepancy can negate the effectiveness of image guidance. CBCT is used at the treatment position but has artifacts that could produce greater dosimetric errors in the correspondingly calculated dose distribution in the patient. Meanwhile, CToR and CBCT are relatively expensive, which impedes their widespread adoption, especially in low-income, rural areas In addition, the rather high imaging doses from both CToR and CBCT, which may have unknown, potential risks act as a barrier to frequent use, such as the promising daily online adaptive radiation therapy (ART)[16–18]. Thus, we aim to follow the dose management principle adopted by the diagnostic imaging community, which is summarized as the principle ALARA, i.e., as low as (is) reasonably achievable[19], improving the patient alignment accuracy while maintaining imaging dose as minimal as possible.

A recent study[20] proposed to synthesize virtual 3D CT from CBCT utilizing the commercially available software, partially addressing the drawbacks of CToR (additional alignment uncertainty due to indispensable position transfer and posture change) and CBCT (low image quality). However, the imaging dose is still not reduced and the process was complicated such that any errors accumulated during the procedure may ultimately lead to significant setup errors. Orthogonal kV image performs in a real-time manner with much less expense and a much lower imaging dose compared to CToR and CBCT. Nevertheless, image quality-wise, kV image only shows clear bony anatomies in a 2D x-ray projection(e.g. the middle image in Fig. 3), making it barely usable in online ART workflow except for patient alignment. Even for patient alignment, these 2D images often lack sufficient soft tissue details, leading to large patient setup uncertainties. The ad hoc handling of large patient setup uncertainties by using large target margins can lead to unnecessarily high doses to nearby OARs and thus unnecessary AEs. Overall, there is currently no imaging guide technique for radiation therapy that can simultaneously be real-time in efficiency, diagnostically equivalent in quality, and cheap and low dose for frequent and widespread use.

Lately, artificial intelligence (AI) has undergone rapid development and its application in radiation oncology has been growing quickly[21–24]. The feasibility of using AI models to inversely reconstruct the 3D CT from 2D clean and noise-free digital reconstructed radiography (DRR) images (e.g. the right image in Fig. 3) has been explored and validated[25–32]. Though conceptually inspiring, such a model itself is chronologically inapplicable in clinical practice since the DRR image is generated based on the planning CT or CToR images previously obtained. Only independent 2D images without prerequisite 3D images, such as kV images, are the practically meaningful inputs to such 2D-to-3D deep learning-based models. Although[26] attempted to test on the 2D X-Ray images in the reference stage, the images need to be converted to DRR first and the attempt lacked both quantitative and qualitative evaluation, making clinical application challenging.In addition, most existing deep learning-based 3D CT reconstruction approaches are tested and validated on small-size images only (typically $128 \times 128 \times 100$), which are far from clinically satisfactory.

More recently[33], explored synthesizing 3D CT from planar X-ray considering the fact that the CT scanners are rare in low and mid-resource settings due to their costs – a motivation that aligns with our work. Nevertheless, the DRR images were used rather than planar X-ray in the experiments, limiting the work to theoretical concepts. In other reported literature[34,35] and[36], all the 2D images used for training and testing were synthesized by ray-tracing from the 3D images, ensuring a strict one-to-one correspondence. Unfortunately, this pre-assumption does not hold in clinical scenario.

Therefore, as far as we are aware, we propose an innovative deep learning-based framework kV2CTConverter (Fig. 1a) composed of dual models built with hierarchical vision transformer (ViT) blocks[37,38]. The proposed framework will take kV images as the exclusive input to synthesize the corresponding full-size 3D CT within milliseconds, which can be used for reflecting 3D patient's position in time, thus achieving high-quality but almost "zero-dose" image-guided patient alignment. To our best knowledge, kV2CTConverter might be the one of the pioneering AI-based attempts that

utilizes only clinically available images (daily kV images) and timely reconstructs ready-to-use, full-size 3D CT, paving the way towards AI-based online ART. We believe this work will have broad impact upon image-guided interventional procedures such as radiation therapy and needle biopsy, especially for pediatric patients who are sensitive to imaging dose. In addition, it may help simplify the hardware of tomographic imaging systems. Also, the developed timely CT reconstruction with limited projections can benefit radiation therapy in low-income, rural areas, where many radiotherapy machines lack 3D OBI capability.

At the core of bridging the theoretical algorithm and the practical application, a hierarchical vision transformer is adopted and adapted to the medical images (i.e., kV images and 3D CT) with a dual-model setting and a data augmentation strategy termed as geometry property reserved shifting and sampling (GRSS) is also proposed. To our best knowledge, GRSS is an advanced yet easy-to-implement data augmentation method that takes the geometrical relation between the treatment couch and kV imaging source and detector into consideration. This strategy enables kV2CTConverter to take full advantage of the noisy but ultra-sparse 2D kV images to fulfill accurate 3D CT reconstruction while avoiding model overfitting. We validated the effectiveness of the kV2CTConverter using 10 independent patients with H&N cancer from three perspectives: (1) image quality evaluation of the synthesized CTs, (2) dosimetric evaluation of the dose distributions calculated using the synthesized CTs, and (3) robustness of the framework to random shifts that mimic patient alignment uncertainties during treatment. In both image quality and dosimetric evaluations, kV2CTConverter exhibited high accuracy. In random shift test, the kV2CTConverter achieved a minimum shift error of 0.4mm, which meets the clinical criterion.

Methods
10 H&N patients previously treated with IMPT were retrospectively selected for this study and the need to obtain informed consent was waived with approval from Mayo Clinic Arizona institutional review board (IRB#: 13-005709). We repeatedly trained independent patient-specific model for each of the selected 10 H&N patients included our study. For each patient, there are multiple pairs of kV images and CToR images. Also, for each patient, we used the leave-one-out strategy to evaluate the model performance and student t-test to measure the significance. To be specific, we randomly chose one pair of kV-CToR images for model validation and used the other pairs of kV-CToR images for model training. The number of raw images, the number of samples in dataset for training the primary model and secondary model used in this study along with the basic treatment plan information are listed in Table 1. For proton treatment, it is imperative to set up patients such that their anatomy matches the planning CT as closely as possible. This requires 1) precise patient positioning, and 2) ensuring that the patient's anatomy has not diverged from the planning CT. To monitor anatomical changes, verification CT scans are performed regularly (for this study, weekly verification CT scans were performed using the CToR), sometimes leading to a new treatment plan, a process known as ART. Precise patient positioning at our proton center is achieved using an orthogonal 2D kV imaging system. See Figure 3. This system compares the real-time orthogonal kV images with DRRs generated from the planning CT. After patients are initially aligned, orthogonal kV images are captured and rigidly registered to their corresponding DRRs to determine whether the patient position should be shifted/rotated. After a few iterations of this process, the patient is ready for treatment. As a result of these processes, a large amount of imaging and registration data are generated for each patient over the course of treatment. Of relevance for this study, these data include the initial planning CT, multiple verification CTs (weekly), rigid registration matrices generated by registering the verification CTs to the planning CT (stored as DICOM files), orthogonal kV images (for each treatment), and rigid registration matrices generated by registering the orthogonal kV images to the DRRs. In this study, we took advantage of the fact that patients were typically treated shortly after undergoing a CToR scan. This meant that for each CToR dataset (each patient in this study received a minimum of 3
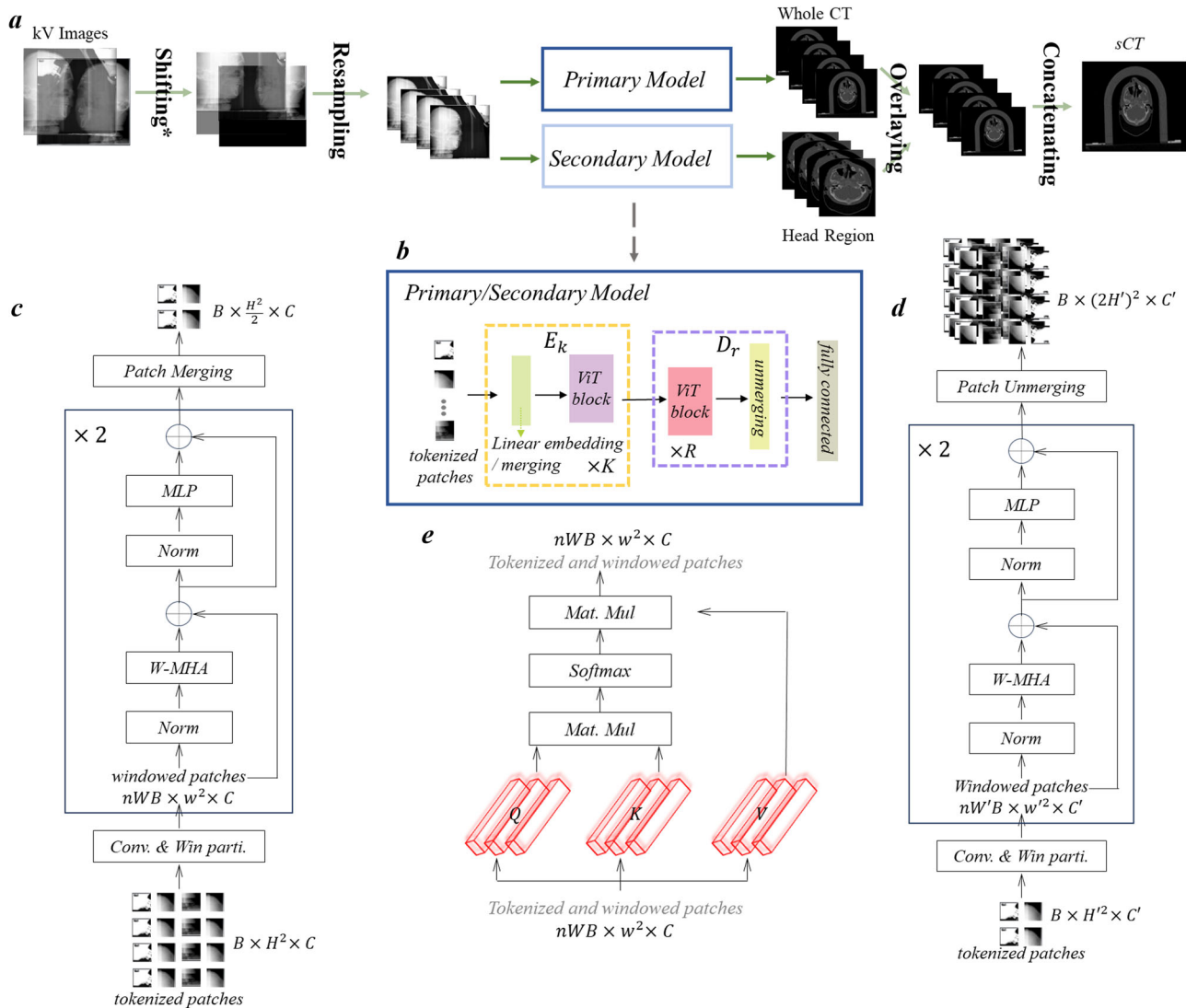
Fig. 1 | Overall architecture of kV2CTConverter. a Workflow of the proposed method. The raw kV images were augmented by GRSS to get adequate samples for model training. Then the processed images simultaneously went through dual models (i.e., primary model and secondary model) to generate the whole CT and the fractional CT that covered only the head region, respectively. Lastly, the full-size synthesized CT was achieved by overlaying and concatenating the outputs from two models according to their spatial relationship. **b** The model structure of both primary and secondary model. **c** The details of the hierarchical ViT blocks in the encoder E_k. **d** The details of the hierarchical ViT blocks in the decoder D_r. **e** The detailed illustration of the window-based Multi-Head Attention (W-MHA), the tokenized patches were first spat to nW non-overlapped windows of a size of $w \times w$ and the attention was only calculated on the windows instead of the whole inputs.

CToR scans), there was a corresponding pair of orthogonal kV images along with its corresponding rigid registration matrices. In this work, we have investigated the performance of a patient-specific model designed to generate synthetic CTs (sCTs) from a pair of 2D orthogonal kV images.

Data pre-processing

Firstly, the CToR and corresponding kV images were registered to the same coordinates with the stored rigid registration matrices to exclude any pre-existing positioning difference. Since the patients in this study were treated with a so-called bolus helmet in place during treatment. The bolus helmet is quite large in terms of volume and was therefore a considerable feature in all of the patient imaging. For this reason, two models were developed, the "primary" and "secondary" models. After registered to the same coordinate, the kV images and the CToR image were used for the training of both models as the input and as the ground truth respectively. While the primary model utilized the whole images of the CToR as reference, the secondary model was only provided with the images of the CToR within the head region (see Fig. 1). Over the course of this study, it was discovered that the

overall model performance was improved by introducing a secondary model that could focus specifically on the head region. The CToR datasets for each patient had a resolution of $512 \times 512 \times N$, where N was the number of the CT slices along the superior-inferior (S-I) direction (varies from patient to patient). CToR datasets were cropped to two different sizes, one size for the primary model and a smaller size for the secondary model. For the primary model, the CToR images were cropped to size $448 \times 336 \times 384$ to exclude the excessive regions outside the patient BODY with low density (i.e. air) as well as for the purpose of homogenizing the size of the datasets. For the secondary model, the CT images were cropped to size $M \times 224 \times 224$, where M indicates the minimum number of the voxels that covered the head region along the R-L direction, which varies from patient to patient. Finally, the corresponding kV images, initially having resolution of 1024×1024, were cropped to 1008×1008 accordingly.

Data augmentation

Converting kV images into a 3D CT is an extremely ill-constrained problem, going from 10^6 pixels to 10^8. This process requires accurate geometric

Table 1 | Statistical information for the 10 patients

Patient ID	kV(raw)	CToR(raw)	Primary	Secondary	Total dose(Gy)	Fractions
1	3	2	21,812	4824	60	30
2	2	2	14,544	3216	60	30
3	4	2	29,088	6432	60	30
4	3	2	21,812	4824	60	30
5	5	2	36,360	8040	60	30
6	3	2	21,812	4824	60	30
7	4	2	29,088	6432	60	30
8	4	2	29,088	6432	70	35
9	2	2	14,544	3216	60	30
10	4	2	29,088	6432	60	30

The second and third columns listed the number of raw kV images and CToR used in this study, respectively. The forth and fifth columns listed the total number of sample pairs in the dataset used for training the primary model and secondary model, respectively. The last two columns listed the total prescription dose and number of fractions for the patient.

Table 2 | MAE (in HU) results from all 10 test patients

Patient ID	kV2CTConverter	Primary model only	Difference
1	30.19	34.47	−4.28
2	28.93	34.00	−5.07
3	75.14	88.53	−13.39
4	40.44	47.85	−7.41
5	33.76	38.44	−4.68
6	39.83	41.81	−1.98
7	34.34	38.56	−4.22
8	46.78	50.34	−3.56
9	86.98	99.05	−12.07
10	29.45	35.47	−6.02
Average	44.58 ± 20.21	50.85 ± 23.40	−6.27 ± 3.71

The lower the MAE value is, the better the image quality is. The second column showed the results from kV2CTConverter while the third column was the results from only the primary model. The last column was the difference between the kV2CTConverter and the primary model only. The last row showcased the mean and standard deviation over all 10 results.

mapping from pixel to voxel and precise identification and localization of H&N voxels amidst background voxels (the air). Therefore, commonly used data augmentation techniques such as Gaussian noise and random image transformations may not be suitable. For example, the former might not generate sufficient data variations when the added noise magnitude is small, or it could obscure the H&N voxels with noise patterns when the added noise magnitude is large. The latter can complicate the pixel-voxel geometrical relationships. Moreover, it is essential that the framework can detect patient shifts and thereby improve patient alignment. For this reason, an innovative data augmentation strategy, the so-called **geometric property-reserved shifting and sampling (GRSS) data augmentation strategy** was proposed to the extent of our knowledge. Given the layout of the kV imaging system in the treatment room, as shown in Fig. 3, we noticed that for a shift of the CToR along the S-I direction, the kV images would shift by a factor of 1.5 in the same direction based on the kV imaging system geometry. Hence, we further augmented the kV-CT pairs by simultaneously moving CT along the S-I direction in steps of 0.1 mm /0.05 mm (0.15 mm/ 0.075 mm for kV images), ± 5mm in total for primary and secondary model respectively. In addition to the "property-reserved shifting" step, the shifted datasets were also downsampled. For the primary model, the CT images were downsampled with a factor of 4 along the R-L direction and a factor of 3 in the A-P and S-I directions, respectively. Correspondingly, the two orthogonal kV images were downsampled with a factor of 6 in both dimensions. Likewise, for the secondary model, the CT images were

downsampled with a factor of 2 along both the A-P and S-I directions, and the corresponding kV images were downsampled with a factor of 2 in both dimensions. Thus, a pair of initial CT and its corresponding kV images yields 36 additional shifted and downsampled CT-kV image pairs for the primary model dataset and 4 for the secondary model dataset. This method helps to avoid overfitting issues due to the limited number of training samples. In addition, it allowed for efficient model training since the size of each sample was less than 200 voxels along any direction. Finally, a high-resolution CT of full size ($512 \times 512 \times N$), desirable for clinical applications, was obtained by spatially stacking the small-size reconstructed CT generated by both the primary model and secondary model.

kV2CTconventer framework

The proposed framework, entitled kV2CTconventer (Fig. 1a), has dual models. Each has an asymmetric autoencoder-like architecture consisting of an encoder E_k and a decoder D_r with hierarchical ViT blocks as the basic building blocks. The overall architecture of both primary model and secondary model is shown in Fig. 1b. Specifically, both models consist of a patch embedding layer (a convolutional layer), an encoder E_k, a decoder D_r and a final fully connected layer. The patch embedding layer is used for projecting non-overlapping raw kV image patches to initial high-dimensional feature representations that serves as the input for the encoder E_k. Both E_k (Fig. 1c) and D_r (Fig. 1d) consist of multiple hierarchical ViT blocks, having a pattern of "layer normalization, window-based multi-head attention(W-MHA), layer normalization, multilayer perceptron (MLP), and patch merging/ unmerging layer". The W-MHA ((Fig. 1e)) calculates the attention within the windows only instead of the entire image, thus greatly reducing the computational complexity[38,39]. The patch merging layer in the E_k concatenating nearby 2×2 patches with a linear merging layer to obtain a hierarchical representation. Likewise, the unmerging layer in the D_r enlarged each patch by a factor of 2 along each dimension through a fully connected layer. Lastly, the final fully connected layer converts from the learned representations to the final output (i.e., the 3D sCT).

Training protocol. The proposed framework was implemented with the PyTorch deep learning library[40]. Moreover, distributed data parallel (DDP)[41] was employed to minimize memory usage and significantly accelerate the training speed. We used the AdamW optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and a cosine annealing learning rate scheduler with an initial learning rate of e^{-7} and 20 warm-up epochs. We used smooth L1[42] as the loss function to obtain a smoother loss curve, which is a combination of L1 and L2 loss. The batch size was set to 300 for both the primary model and the secondary model.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.
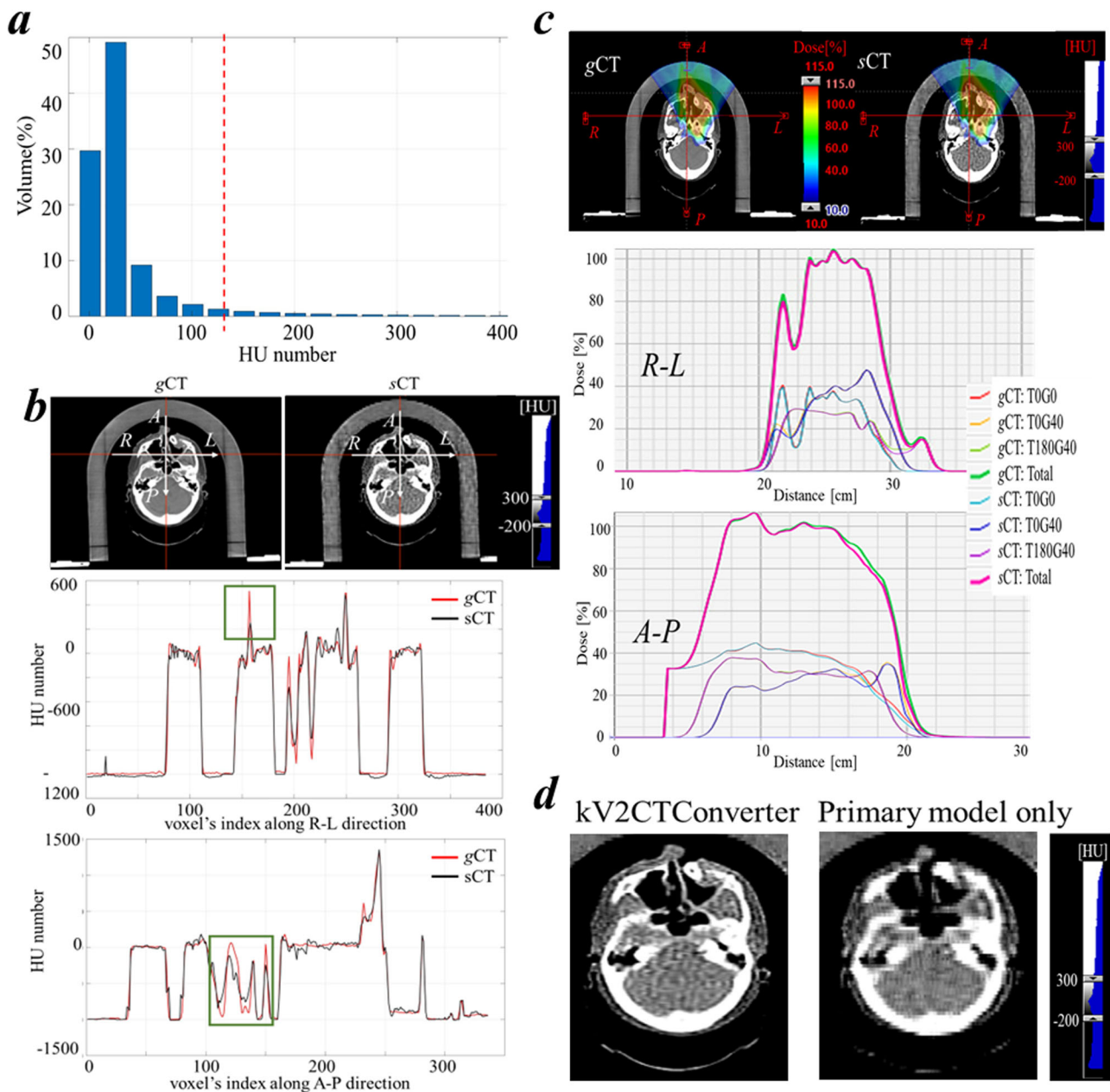
Fig. 2 | Experimental results from a typical patient. a CDVH of the sCT from a typical patient. **b.** HU number profile comparison between sCT and gCT in both R-L and A-P directions. The green box showed the area where the discrepancy was large. **c.** Dose profile comparison between the doses calculated on the sCT and on the gCT in both both R-L and A-P directions. **d** Enlarged head region comparison between the kV2CTConverter and primary model only.

Results

Image quality of the reconstructed 3D CT

We evaluated the sCT (In this paper, we use reconstructed CT and synthetic CT interchangeably.) with mean absolute error (MAE) and absolute per-voxel CT numbers (i.e., image intensity in terms of Housfield Unit (HU)) difference volume histogram (CDVH), compared to the CToR that was taken on the same day as the input kV images (i.e. ground-truth CT (gCT)). To measure the global similarity of the images, 3D Gamma analysis of the CT numbers comparing the sCT and gCT was done. The calculation was done twice, with sCT and gCT as the reference CT, respectively.

The second column in Table 2 listed the results given by kV2CTConverter and the third column showed the results given by the primary model only from 10 patients respectively in terms of MAE (in HU). On average, the kV2CTConverter achieved an MAE of 44.58 with a standard deviation (SD) of 20.21 and a p-value of 0.00023 given by the student t-test, indicating the sCT and gCT yielded a high agreement at the voxel level

and our proposed framework improve the MAE performance by a statistically significant amount compared to the primary model only.

Figure 2 a depicted the CDVH of one typical patient. It is clear that the majority of the voxels had a per-voxel CT number absolute difference of less 100 HU. Numerically, only 5% of the voxels had an absolute difference larger than 128HU. We also randomly selected one slice from the 3D CT image of one typical patient and conducted the HU number profile comparison to visualize the difference in both right-left (R-L) and anterior-posterior (A-P) directions between the gCT and sCT in Fig. 2b, where the red curve represented the gCT while the black one represented the sCT. We observed that the two curves were highly overlapped, showing the sCT matched well with gCT. The 3D Gamma passing rates of comparing sCT to gCT for all ten patients were calculated and the results were reported in Table 3. Regardless of reference CTs selected for calculation, our proposed approach achieved mean 3D Gamma passing rate of 98.95% (SD = 1.75%, p = 0.004), 98.91% (SD = 1.76%, p = 0.002) and 98.90%

Table 3 | 3D Gamma Passing rates (in %) comparing the CT numbers (HU) of sCTs to gCTs

Patient ID	3%3mm		3%2mm		2%2mm	
	sCT	gCT	sCT	gCT	sCT	gCT
1	99.75	98.95	99.74	98.89	99.74	98.89
	99.66	97.82	99.61	97.72	99.61	97.72
2	99.99	99.83	99.99	99.81	99.99	99.80
	99.98	99.04	99.97	98.94	99.97	98.94
3	99.16	97.09	99.14	97.03	99.14	97.03
	98.33	91.18	98.08	90.97	98.07	90.97
4	99.88	99.33	99.88	99.27	99.88	99.27
	99.69	98.09	99.50	98.02	99.51	98.03
5	99.99	99.74	99.99	99.68	99.99	99.68
	99.99	99.19	99.98	99.07	99.98	99.07
6	99.99	97.84	99.98	97.67	99.98	97.67
	99.88	96.82	99.79	96.74	99.79	96.74
7	99.99	99.71	99.99	99.65	99.99	99.65
	99.98	99.21	99.94	99.09	99.94	99.09
8	99.99	99.14	99.98	98.93	99.98	98.93
	99.94	98.10	99.90	97.96	99.90	97.96
9	95.94	92.94	95.93	92.85	95.92	92.85
	93.17	85.74	92.85	85.49	92.85	85.49
10	99.98	99.89	99.96	99.77	99.96	99.77
	99.34	98.88	98.98	97.79	98.97	97.79
kV2CT	99.46 ± 1.20	98.45 ± 2.03	99.46 ± 1.20	98.36 ± 2.04	99.46 ± 1.21	98.35 ± 2.04
	98.95 ± 1.75		98.91 ± 1.76		98.90 ± 1.97	
Primary	98.99 ± 2.00	96.41 ± 4.21	98.86 ± 2.08	96.18 ± 4.22	98.86 ± 2.08	96.18 ± 4.22
	97.70 ± 3.54		97.51 ± 3.58		97.51 ± 3.58	

Three criteria were considered with sCT and gCT as the reference, respectively. In each cell two values were reported, the 1st row represented the kV2CTConverter while the 2nd represented the primary model only. The last four rows showed the mean and standard deviation value over the 10 patients.

Table 4 | 3D Gamma Passing rates (in %) of the doses calculated on the sCTs and the gCTs using the same plan

Patient ID	3%3 mm		3%2 mm		2%2 mm	
	sCT	gCT	sCT	gCT	sCT	gCT
1	99.99	99.96	99.94	99.85	99.66	99.51
	93.18	90.73	88.36	86.62	85.45	83.44
3	99.36	99.67	98.54	98.95	97.28	97.70
	90.39	89.77	86.30	86.14	83.14	82.93
4	99.93	99.89	99.66	99.55	99.02	98.89
	96.04	93.25	92.41	90.21	89.87	87.38
5	99.99	99.99	99.94	99.97	99.82	99.83
	97.89	98.21	97.59	97.97	97.59	97.97
6	99.22	98.63	96.99	96.08	95.46	94.33
	93.05	95.19	88.18	90.96	84.92	87.75
7	99.94	99.93	99.55	99.56	98.04	98.14
	93.17	87.84	87.75	82.78	84.15	79.05
8	99.23	99.55	98.40	98.81	97.29	97.70
	95.05	94.27	91.74	91.30	88.66	88.17
9	99.31	99.09	97.34	97.22	94.70	94.65
	92.94	93.39	87.13	87.63	82.26	82.70
10	99.97	99.95	99.66	99.57	98.56	98.49
	91.39	90.67	88.21	88.54	86.74	85.06
kV2CT	99.67 ± 0.34	99.65 ± 0.45	98.95 ± 1.08	98.91 ± 1.27	97.87 ± 1.71	97.82 ± 1.89
	99.66 ± 0.39		98.93 ± 1.15		97.85 ± 1.76	
Primary	94.11 ± 2.57	93.04 ± 3.30	90.32 ± 3.84	89.69 ± 4.39	87.59 ± 4.83	86.65 ± 5.39
	93.57 ± 2.93		90.01 ± 4.03		87.12 ± 5.00	

Three criteria were considered with the dose calculated on the sCT and gCT as the reference, respectively. In each cell two values were reported, the 1st row represented the kV2CTConverter while the 2nd represented the primary model only. The last four rows showed the mean and standard deviation value over the 10 patients.

(SD = 1.97%, p = 0.002) with criteria of 3%3mm, 3%2mm and 2%2mm, respectively.

Dosimetric evaluation of the reconstructed 3D CT

We conducted forward dose calculation with Eclipse™ (version 15.6, Varian Medical System, Palo Alto, CA) on both the sCTs and gCTs using the same plan and compared the dose distributions using 3D Gamma analysis. The corresponding dose volume histogram (DVH) indices of targets and OARs were compared as well. Fig. 2c depicted a typical dose profile comparison between the doses calculated on the sCT and the dose calculated on the gCT in both R-L and A-P directions. We found that the dose calculated on the sCT was very close to that calculated on the gCT for every treatment field and all fields accumulated. Table 4 showed the 3D Gamma passing rate results where the calculation was done twice, with the dose calculated on the sCTs and the dose calculated on the gCTs as the reference dose, respectively. Our framework achieved high passing rates of 99.66% (SD = 0.39%, p = 1.015 $e-08$), 98.93% (SD = 1.15%, p = 1.386 $e-07$), and 97.85% (SD = 1.76%, p = 3.552$e-06$) on average with criteria of 3%3 mm, 3%2 mm, and 2%2 mm, respectively. Table 5 showcased the difference in DVH indices of CTV and 4 OARs between the doses calculated on the sCTs and the gCTs using the same plans, respectively. It is worth noting that only the DVH index differences were reported as the plans and dose-volume constraints varied from patient to patient. Hence, DVH index difference was considered to be a better sCT-vs-gCT similarity indicator than the DVH index absolute value itself. We noticed that the proposed framework achieved a negligible difference for D95% and D2% of CTV, indicating the equivalence of CTV coverage and hot spot control in both dose distributions. For the selected OARs, the DVH index differences were also small. All these results showcased that the doses calculated on the sCTs and gCTs using the same plan exhibited minimal disparity, indicating that sCT could serve as a viable substitute for verification CT for the purposes of plan evaluation or adaptive re-planning.

Shift robustness evaluation: mimicking patient setup uncertainty

As the primary application of the proposed framework is to generate accurate 3D CT for patient alignment during RT treatment, we conducted a comprehensive analysis of the robustness of the proposed framework to generate sCTs in the face of patient setup uncertainties. We performed random shifts of the kV images to simulate patient setup uncertainty. Given manually shifted kV images within ± 4.5 mm as input, on one hand, the model predicted the shifted sCT(ssCT). On the other hand, the shifted gCT (sgCT) could be calculated based on the geometrical relation between the treatment couch and kV imaging system (see Fig. 3 for the detailed treatment room layout). To obtain the shift error(SE), we first created a searching pool $S = sgCT + \delta, \delta \in [-1, 1]$, consisting of sgCT and its variances (shifting sgCT within ± 1mm with a step of 0.1 mm), total 21 candidates. Next, the MAE between each candidate sgCT in S and ssCT was calculated. Lastly, by linear search, the sgCT with δ_m that gave the minimum MAE was identified and the absolute value of δ_m was defined as SE. We reported the results in Fig. 4 and Table 6. The model yielded a mean SE of only 0.40 ± 0.16mm on average in the sCT robustness test mimicking daily clinic practice, the

Table 5 | DVH index difference between the doses calculated on the sCTs and the gCTs using the same plan

Patient ID	CTV (D95%)	CTV (D2%)	Brain stem (D0.01cc)	Parotid total (Mean)	Oral cavity (Mean)	Mandible (D0.03cc)
1	1.394	−0.97	27.1cGy	1.7cGy	43.1cGy	–
	−66.381	−2.96	2525cGy	19.9cGy	−140.7cGy	–
2	0.066	−0.08	44cGy	−7.3cGy	0.1cGy	2.4cGy
	−4.708	−0.25	148.4cGy	17cGy	−14.6cGy	35.1cGy
3	−3.25	0.172	−11cGy	–	–	–
	-32.466	2.5	53cGy	–	–	–
4	0.617	0.288	97.3cGy	10.1cGy	13.4cGy	13.2cGy
	-6.248	3.09	−961.6cGy	-8.2cGy	−19.2cGy	120.4cGy
5	0.229	−0.045	69.3cGy	0.6cGy	0	32cGy
	0.45	−0.21	112.2cGy	3.8cGy	1.2cGy	38cGy
6	0.35	−0.69	210.8cGy	–	–	–
	−5.832	1.56	−1078.9cGy	-	–	–
7	−1.786	0.31	0	0	−49cGy	−3cGy
	−15.927	0.33	0	0	−62.7cGy	128.8cGy
8	0.447	2.22	53.2cGy	-21.8cGy	2.8cGy	329.8cGy
	−0.652	2.91	91.6cGy	135.6cGy	57.3cGy	338.6cGy
9	0.094	−2.76	−7.6cGy	-105.2cGy	15.7cGy	22.8cGy
	2.583	4.44	188.1cGy	204cGy	105.3cGy	−21.4cGy
10	1.673	0.045	132.9cGy	18.7cGy	12.9cGy	-6.2cGy
	7.221	4.421	212.2cGy	36.2cGy	32.7cGy	15.7cGy

The CTV and 4 OARs were considered and the constraints associated with each structure were shown in the (). For each patient, two results were reported, the first row showed the results from kV2CTConverter, while the the second row was from the primary model only.
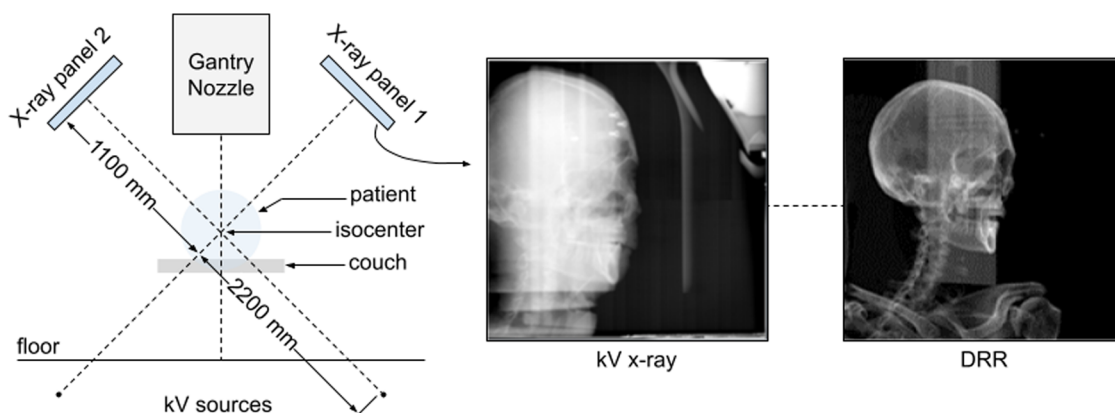
Fig. 3 | The orthogonal kV x-ray system used at our proton center and exemplary kV and DRR images. The orthogonal kV x-ray system is used for patient alignment at our proton center (left). An exemplary kV image (middle) captured by this system and its corresponding DRR image (right). Compared to the DRR image, the daily-used kV image is often noisy and contains unwanted artifacts from essential medical devices/accessories, such as dental implants or treatment couch attachments.

patient alignment tolerance for H&N patients is 2–3 mm clinically at our institution.

Ablation study

To investigate if training the secondary model indeed improved the quality of the sCTs, we conducted ablative experiments and reported the results in Tables 2 and 3. From the last row of Table 2, we observed that adding the secondary model reduced the mean MAE value(in HU) by 6.27 (SD = 3.71). Although not numerically significant, the visual details in the sCTs trained with both the primary and secondary models, especially in the nasal cavity, brain stem, cochleas, and the surrounding bones were much more distinct than those in the sCTs trained only with the primary model (Fig. 2d). Moreover, the MAE decreased by 38.8% when only calculated within the head region with the introduction of the secondary model. The observation from the 3D Gamma passing rates of the CT numbers of sCTs reported in Table 3 also followed the same trend. In addition, The 3D Gamma passing rates of the doses calculated using sCTs and the corresponding DVH index comparison further validated such an observation (Tables 4 and 5). This indicated that as a challenging task of sCT generation in this study, which is indispensable to sub-tasks, i.e., locating the target (finding the position of the H&N region in the CT images) and reconstructing the fine details of the patient anatomies, a practical solution is to train a framework consisting of dual models, each handling one sub-task.

Statistics and reproducibility

The details about experimental design and statistics used in model training and analysis performed in this study are given in the respective sections of methods and results.

Discussion

kV2CTConverter can effectively reconstruct accurate high-resolution 3D CT from two 2D kV images with arbitrary artefacts

This is the most prominent novelty in this proposed work. Specifically, to our best knowledge, the proposed framework might be one of the innovative initiatives to solely take the kV images and their corresponding CToR as the training and testing datasets, without referring to supplementary images such as DRRs. Moreover, different from the proof-of-concept results (i.e. cropped small-size sCT images) reported in the existing approaches[25–28,30–32], we can synthesize full-size sCTs, which can be directly applied in daily clinical practice. In terms of image quality, compared to the numerical results reported in[25,28], although theirs were calculated on small-size 3D lung CT (supposed to be easier as the structures were much larger and had more distinct boundaries), ours still outperformed them by 53.56% and 62.01% on average, respectively. We believe that such performance superiority has demonstrated the potential of the proposed framework in clinical applications.

kV2CTConverter can effectively handle patient shift uncertainty and generate sCT that reflects the patient position shift in milliseconds

To mimic patient position shift during RT treatment, which commonly happens, we conducted a random shift simulation. The proposed framework was able to generate sCTs reflecting their positions effectively with milliseconds and achieve a minimum SE of < 0.4 mm on average compared with the shifted real patient positions, which is much lesser than the clinic criteria, 2 mm or 3 mm for H&N cancer patients. The robustness gain is closely related to the proposed GRSS data augmentation strategy, which not only achieved plenty of data samples for model training but also enabled the framework to cope with random shifts of the patient position accurately.

The secondary model is essential for fine details recovery in the 3D CT reconstruction

From the experiment results reported in Tables 2–4, we observed that the secondary model indeed boosted the framework performance. Although ViT has outperformed other deep learning-based models in natural image

Fig. 4 | The histogram of shift error (SE) (in mm) from all 10 patients. The Y-axis shows the bin edges for the histogram. It is clear that the majority of the SEs are less than 0.4 mm, which is far smaller than the clinically acceptable patient alignment tolerance for H&N patients at our institution, set at 2-3 mm.

classification and reconstruction, it is still very challenging when employed for tasks related to medical images, as there is a great gap between medical images and natural images. Thus, we proposed a dual-model framework to enable the state-of-the-art ViT to be adapted to medical images, in which the primary model was dedicated to identifying the positions of structures of interest, and the secondary model focused on reasoning the 2D-3D relations and reconstructing the voxel-level fine details in 3D CT. Besides, such a dual-model configuration is intuitive, easy to implement, resource-efficient, and can be generalized to other medical imaging modalities, such as MRI, PET, etc.
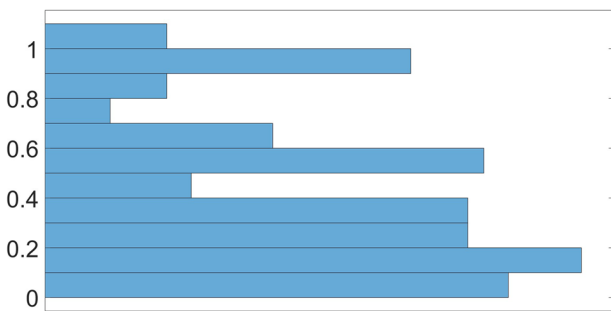
Outlook

From the green boxes shown in Fig. 2b, we noticed that the areas corresponding to the sinuses region, where locates the microscopic soft tissue surrounded by an air cavity with an irregular shape, was not well reconstructed (left green box). Moreover, the brain stem which has an irregular shape and has a similar CT number to the surrounding cerebral hemisphere matters, was not reconstructed with distinct boundaries either (right green box). Intuitively, we can build another model to focus on these two small regions respectively. However, different from the head region considered in the secondary model, where the shape is regular and the position is independent of the helmet, the properties of the two small regions are quite opposite. Thus, it may complicate the framework and double the computing resources. Fortunately, from a purely clinical point of view, those two small regions don't affect a lot: (1) if the tumor is located in those regions, it will either shrink the air cavity or exhibit a clear difference in CT numbers compared to the surrounding tissues, making it easier for the model to distinguish the tumor from nearby organs in either way; (2) if the tumor doesn't appear in those regions, the dosimetric constraints to those regions only possess a low priority when designing the treatment plan. In the future, we will investigate how to further improve the quality of the reconstructed CT regarding those two small regions.

Anatomical variations over the treatment is also critical that should be taken into consideration in RT. However, following our current clinical practice exactly (CToR and kV images on the first treatment day of the week, only kV images for the remainder of the week), the framework may not effectively reflect notable inter-fractional anatomical changes well since the long-term anatomical changes are only taken into account with weekly CToRs. Nevertheless, methodologically, no fundamental difference exists between the current training scheme and the scheme to account for long-term inter-fractional anatomical changes. If our framework is trained with more CToRs and the associated kV images along the treatment courses, our framework can predict long-term inter-fractional anatomical changes. We will extend our framework to account for long-term and possibly more significant inter-fractional anatomical changes in the future work.

Moreover, if CT with intravenous (IV) contrast are available, it may further improve the identification of the tumor and potentially improve the accuracy of the sCT. However, typically, on the treatment day in clinical practice, the patient will only have the kV images and/or regular CToR whereas the IV contrast is not used. Because the IV contrast will change the CT numbers of the tumor and consequently impact the accuracy of the dose delivery[43]. Therefore, we don't have kV images and the corresponding CT with IV contrast pairs available for the training of the model. We will leave it for further exploration.

In addition, this work is a feasibility study about the 2D to 3D imaging device which must be clinically commissioned (IMO). Thus, we will also consider using a phantom for an end-to-end test in the future for a more comprehensive assessment of the proposed framework.

Table 6 | The mean shift error (SE) (in mm) from all 10 patients. The lower the SE value is, the more robust the proposed framework is

Patient ID	1	2	3	4	5	6	7	8	9	10
SE	0.37	0.56	0.66	0.25	0.44	0.34	0.63	0.26	0.31	0.22

Data availability

Code availability

The interesting party can contact the corresponding author for the code sharing. A patent application based on this work has been filed by the Mayo Clinic Ventures Office and is currently pending. In accordance with Mayo Clinic institutional policy, sharing code associated with a filed patent requires approval from the Mayo Clinic Ventures Office. Additionally, all relevant legal agreements, including a materials transfer agreement, must be established between Mayo Clinic and the interested party.

References

1. Schild, S. E. et al. Proton beam therapy for locally advanced lung cancer: A review. *World J. Clin. Oncol.* **5**, 568 (2014).
2. Liu, W., Li, Y., Li, X., Cao, W. & Zhang, X. Influence of robust optimization in intensity-modulated proton therapy with different dose delivery techniques. *Med. Phys.* **39**, 3089–3101 (2012).
3. Liu, W. et al. Impact of respiratory motion on worst-case scenario optimized intensity modulated proton therapy for lung cancers. *Pract. Radiat. Oncol.* **5**, e77–e86 (2015).
4. Stuschke, M., Kaiser, A., Pöttgen, C., Lübcke, W. & Farr, J. Potentials of robust intensity modulated scanning proton plans for locally advanced lung cancer in comparison to intensity modulated photon plans. *Radiother. Oncol.* **104**, 45–51 (2012).
5. Lomax, A. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions. *Phys. Med. Biol.* **53**, 1043 (2008).
6. Deng, W. et al. A critical review of let-based intensity-modulated proton therapy plan evaluation and optimization for head and neck cancer management. *Int. J. Part. Ther.* **8**, 36–49 (2021).
7. Liu, C. et al. Small-spot intensity-modulated proton therapy and volumetric-modulated arc therapies for patients with locally advanced non-small-cell lung cancer: a dosimetric comparative study. *J. Appl. Clin. Med. Phys.* **19**, 140–148 (2018).
8. Matney, J. et al. Effects of respiratory motion on passively scattered proton therapy versus intensity modulated photon therapy for stage iii lung cancer: are proton plans more sensitive to breathing motion? *Int. J. Radiat. Oncol.* * *Biol.* * *Phys.* **87**, 576–582 (2013).
9. Matney, J. E. et al. Perturbation of water-equivalent thickness as a surrogate for respiratory motion in proton therapy. *J. Appl. Clin. Med. Phys.* **17**, 368–378 (2016).
10. Quan, E. M. et al. Preliminary evaluation of multifield and single-field optimization for the treatment planning of spot-scanning proton therapy of head and neck cancer. *Med. Phys.* **40**, 081709 (2013).
11. Shan, J., An, Y., Bues, M., Schild, S. E. & Liu, W. Robust optimization in IMPT using quadratic objective functions to account for the minimum MU constraint. *Med. Phys.* **45**, 460–469 (2018).
12. Ang, K. K. et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N. Engl. J. Med.* **363**, 24–35 (2010).
13. Yang, Y. et al. Exploratory investigation of dose-linear energy transfer (let) volume histogram (DLVH) for adverse events study in intensity modulated proton therapy (impt). *Int. J. Radiat. Oncol.* * *Biol.* * *Phys.* **110**, 1189–1199 (2021).
14. Yang, Y. et al. Exploratory study of seed spots analysis to characterize dose and linear-energy-transfer effect in adverse event initialization of pencil-beam-scanning proton therapy. *Med. Phys.* **49**, 6237–6252 (2022).
15. Yang, Y. et al. Empirical relative biological effectiveness (RBE) for mandible osteoradionecrosis (ORN) in head and neck cancer patients treated with pencil-beam-scanning proton therapy (PBSPT): a retrospective, case-matched cohort study. *Front. Oncol.* **12**, 843175 (2022).
16. Yan, D., Vicini, F., Wong, J. & Martinez, A. Adaptive radiation therapy. *Phys. Med. Biol.* **42**, 123 (1997).
17. Yan, D. et al. The use of adaptive radiation therapy to reduce setup error: a prospective clinical study. *Int. J. Radiat. Oncol.* * *Biol.* * *Phys.* **41**, 715–720 (1998).
18. Feng, H. et al. Gpu-accelerated Monte Carlo-based online adaptive proton therapy: A feasibility study. *Med. Phys.* **49**, 3550–3563 (2022).
19. on Radiation Protection, N. C. & (US), M. Ncrp report (National Council on Radiation Protection and Measurements, 1990).
20. Kaushik, S., Ödén, J., Sharma, D. S., Fredriksson, A. & Toma-Dasu, I. Generation and evaluation of anatomy-preserving virtual CT for online adaptive proton therapy. *Med. Phys.* **51**, 17–28 (2024).
21. Liu, C. et al. Artificial general intelligence for radiation oncology. *Meta-Radiol.* **1**, 100045 (2023).
22. Ding, Y. et al. Deep-learning based fast and accurate 3D CT deformable image registration in lung cancer. *Med. Phys.* **50**, 6864–6880 (2023).
23. Balagopal, A. et al. PSA-NET: Deep learning-based physician style-aware segmentation network for postoperative prostate cancer clinical target volumes. *Artif. Intell. Med.* **121**, 102195 (2021).
24. Jiang, Z. et al. 3d in vivo dose verification in prostate proton therapy with deep learning-based proton-acoustic imaging. *Phys. Med. Biol.* **67**, 215012 (2022).
25. Shen, L., Zhao, W. & Xing, L. Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. *Nat. Biomed. Eng.* **3**, 880–888 (2019).
26. Ying, X. et al. X2ct-gan: reconstructing ct from biplanar x-rays with generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10619–10628 (2019).
27. Bayat, A. et al. Inferring the 3D standing spine posture from 2D radiographs. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23*, 775–784 (Springer, 2020).
28. Lei, Y. et al. Deep learning-based real-time volumetric imaging for lung stereotactic body radiation therapy: a proof of concept study. *Phys. Med. Biol.* **65**, 235003 (2020).
29. Jiang, L. et al. Reconstruction of 3d ct from a single x-ray projection view using CVAE-GAN. In *2021 IEEE International Conference on Medical Imaging Physics and Engineering (ICMIPE)*, 1–6 (IEEE, 2021).
30. Maken, P. & Gupta, A. 2D-to-3D: A review for computational 3D image reconstruction from x-ray images. *Arch. Comput. Methods Eng.* **30**, 85–114 (2023).
31. Gao, Y. et al. 3DSRNET: 3D spine reconstruction network using 2D orthogonal x-ray images based on deep learning. *IEEE Trans. Instrum. Meas.* **72**, 1–14 (2023).
32. Zhang, C. et al. Xtransct: Ultra-fast volumetric CT reconstruction using two orthogonal x-ray projections via a transformer network. *Phys. Med. Biol.* **69**, 085010 (2024).
33. Sun, Y., Netherton, T., Court, L., Veeraraghavan, A. & Balakrishnan, G. Ct reconstruction from few planar x-rays with application towards low-resource radiotherapy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 225–234 (Springer, 2023).
34. Lei, Y. et al. Deep learning-based fast volumetric imaging using KV and MV projection images for lung cancer radiotherapy: A feasibility study. *Med. Phys.* **50**, 5518–5527 (2023).
35. Chung, H., Ryu, D., McCann, M. T., Klasky, M. L. & Ye, J. C. Solving 3D inverse problems using pre-trained 2D diffusion models. In

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 22542–22551 (2023).

36. Cafaro, A. et al. X2vision: 3D CT reconstruction from biplanar x-rays with deep structure prior. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 699–709 (Springer, 2023).

37. Dosovitskiy, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *Proceedings of the International Conference on Learning Representations* (2021).

38. Liu, Z. et al. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022 (2021).

39. He, K. et al. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16000–16009 (2022).

40. Paszke, A. et al. Automatic differentiation in pytorch. In *NIPS-W* (2017).

41. Li, S. et al. Pytorch distributed: Experiences on accelerating data parallel training. In *Proceedings of the VLDB Endowment*, **13**, 3005–3018 (2020).

42. Girshick, R. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, 1440–1448 (2015).

43. Bae, K. T. Intravenous contrast medium administration and scan timing at ct: considerations and approaches. *Radiology* **256**, 32–61 (2010).

Acknowledgements

Author contributions

W.L. proposed the initial idea and supervised the research. Y.D. designed and implemented the algorithm, carried out data processing, experimental and analysis work. J.H. extracted the raw data and advised on dosimetrical and shift robustness analysis. H.F. advised on dosimetrical and shift robustness analysis. B.L. advised on the framework design and results analysis. L.M., J.R., S.V., W.W., D.M., R.F. and S.P. provided the patient data, clinical guidance, and data interpretation. W.L., Y.D., J.H., H.F. and B.L. wrote the manuscript. All authors discussed the results, revised the draft manuscript, read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43856-024-00672-y.

Correspondence and requests for materials should be addressed to Wei Liu.

Peer review information *Communications Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.