Title

Deep match: A zero-shot framework for improved fiducial-free respiratory motion tracking

Authors

Xu, Di
Descovich, Martina
Liu, Hengjie
et al.

Publication Date

2024-05-01

DOI

10.1016/j.radonc.2024.110179

Radiotherapy and Oncology 194 (2024) 110179

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Deep match: A zero-shot framework for improved fiducial-free respiratory motion tracking

Di Xu [a], Martina Descovich [a], Hengjie Liu [b], Yi Lao [b], Alexander R. Gottschalk [a], Ke Sheng [a,*]

[a] Radiation Oncology, University of California, San Francisco, United States
[b] Radiation Oncology, University of California, Los Angeles, United States

ARTICLE INFO

Keywords:
SBRT
CyberKnife
Xsight Lung Tracking
Template matching
Deep learning
Computer vision

ABSTRACT

Background and purpose: Motion management is essential to reduce normal tissue exposure and maintain adequate tumor dose in lung stereotactic body radiation therapy (SBRT). Lung SBRT using an articulated robotic arm allows dynamic tracking during radiation dose delivery. Two stereoscopic X-ray tracking modes are available – fiducial-based and fiducial-free tracking. Although X-ray detection of implanted fiducials is robust, the implantation procedure is invasive and inapplicable to some patients and tumor locations. Fiducial-free tracking relies on tumor contrast, which challenges the existing tracking algorithms for small (e.g., <15 mm) and/or tumors obscured by overlapping anatomies. To markedly improve the performance of fiducial-free tracking, we proposed a deep learning-based template matching algorithm – Deep Match.

Method: Deep Match consists of four self-definable stages – training-free feature extractor, similarity measurements for location proposal, local refinements, and uncertainty level prediction for constructing a more trustworthy and versatile pipeline. Deep Match was validated on a 10 (38 fractions; 2661 images) patient cohort whose lung tumor was trackable on one X-ray view, while the second view did not offer sufficient conspicuity for tumor tracking using existing methods. The patient cohort was stratified into subgroups based on tumor sizes (<10 mm, 10–15 mm, and >15 mm) and tumor locations (with/without thoracic anatomy overlapping).

Results: On X-ray views that conventional methods failed to track the lung tumor, Deep Match achieved robust performance as evidenced by >80 % 3 mm-Hit (detection within 3 mm superior/inferior margin from ground truth) for 70 % of patients and <3 mm superior/inferior distance (SID) ~1 mm standard deviation for all the patients.

Conclusion: Deep Match is a zero-shot learning network that explores the intrinsic neural network benefits without training on patient data. With Deep Match, fiducial-free tracking can be extended to more patients with small tumors and with tumors obscured by overlapping anatomy.

Introduction

With the advancements of diagnostic algorithm and the increase of lung screening initiatives, a growing number of patients with lung cancer are being diagnosed at earlier stages, where surgery or radiation therapy can be offered with curative intent. Notably, patients with smaller tumors (<=15 mm) have reported a significantly higher 5-year survival rate of >95 % compared to those with larger tumors (5-year survival rate >79 % for >15 mm tumors) [1]. In recent years, stereotactic body radiation therapy (SBRT) has gained increased popularity due to improved local control and reduced toxicity compared to conventional radiation therapy, and becomes the standard-of-care for surgically ineligible NSCLC (Non-small cell lung cancer) patients [2]. SBRT has demonstrated remarkable outcomes in localized stage I NSCLC, with a promising two-year survival rate of 70 % and a two-year local control rate of 91 % [3]. Moreover, studies suggest that SBRT is applicable to a broader range of tumors beyond medically inoperable cases [4], and it may provide comparable survival outcomes to surgical resection [5]. To achieve adequate lung tumor coverage and minimize exposure to normal organs, motion management plays a pivotal role in lung SBRT.

* Corresponding author at: Department of Radiation Oncology, University of California, UCSF Box 1708, 1600 Divisadero St., Suite H-1031, San Francisco, CA 94115, United States.
E-mail address: ke.sheng@ucsf.edu (K. Sheng).
URL: https://shenglab.ucsf.edu, https://radonc.ucsf.edu/physics-division (K. Sheng).

https://doi.org/10.1016/j.radonc.2024.110179
Received 23 October 2023; Received in revised form 24 January 2024; Accepted 16 February 2024
Available online 24 February 2024

Current motion management strategies, including internal target volume (ITV), gated radiotherapy, and dynamic tumor tracking, have been employed to account for respiratory motion and enhance treatment precision while reducing exposure to healthy tissues. The CyberKnife® Robotic Radiosurgery System (Accuray Incorporated, Sunnyvale, CA) offers frameless dynamic SBRT for lung tumors, specifically attuned for real-time target and motion tracking, providing enhanced treatment accuracy and efficacy [6].

The CyberKnife system manages respiratory motion using the Synchrony® Respiratory Tracking System (Accuray Incorporated, Sunnyvale, CA), which establishes a correlation between the surface infrared light-emitting diodes and periodic stereoscopic X-rays of the lung tumor and then uses the hybrid system to guide robotic tracking of the moving tumor [7]. In specific, lung tumor tracking is achieved by matching stereoscopic live X-rays acquired periodically throughout treatment (typically every 60 s) to their corresponding digitally reconstructed radiographs (DRRs) generated from the treatment planning computed tomography (CT) images [6]. Two tracking algorithms are currently available for lung SBRT – fiducial marker tracking and Xsight™ (Accuray Incorporated, Sunnyvale, CA) Lung Tracking (XLT).

Fiducial Marker tracking requires 1–3 bronchoscopically, endovascularly, or percutaneously implanted radiopaque fiducial markers in or adjacent to the target volume several days before treatment planning [8]. The radio-opaque fiducials are used as reference markers to localize the target during patient alignment and treatment delivery [7]. Research has shown that fiducial markers provide unambiguous contrast in X-rays, and their tracking is generally robust [9–11]. Fiducials are not without limitations. Besides reported fiducial migration [7], fiducial implantation is an invasive and resource-demanding procedure associated with non-negligible medical risk, including pneumothorax.

Alternatively, the intrinsic radiographical contrast between lung tumors and lung parenchyma makes fiducial-less XLT possible for certain lung patients. Evidence suggests that such X-ray visible tumors are typically above 15 mm across all dimensions and located in the peripheral or apex regions [6,7]. Tumors smaller than this or obscured by dense structures (such as the mediastinum, heart, ribcage, or spine) cannot be reliably detected and tracked with existing XLT algorithms. XLT registers live X-rays to its corresponding DRRs solely using pixel intensities from the tumor and its surrounding soft tissues. The algorithm consists of two steps. The first step is spine registration for global patient alignment (position and orientation), while the second step is direct tumor registration for tumor position localization. The step-two soft-tissue registration algorithm has evolved through three versions. The original XLT released in 2006 performed a block-matching search of the most similar region in live X-rays based on the pre-defined tumor "matching window" extracted from DRRs. Pattern intensity [12] was used to determine image similarity [7]. Original XLT was revised in 2009 with a more rigidly defined DRR "matching window" (within 20 mm of delineated tumor volume), $\pm 10^{\circ}$ in-plane "matching window" rotation to compensate tumor respiratory rotation on X-ray images, and automatic live X-ray enhancement to match the X-ray intensity histogram to that of DRR [13]. The most recent release in 2011 divided the tumor template into overlapping patches, registered the template patches independently to the live X-rays using normalized cross-correlation (NCC) similarity measurements, and used a weighted sum of 2D patches to calculate the overall registration for the whole tumor template [6]. However, both algorithms have specific limitations that restrict the broader applicability of CyberKnife in certain SBRT cases.

Prior to treatment, a correlation model between internal XLT-measured tumor position and external optical surface marker position is built using a series 8–12 X-rays acquired at multiple phases of the breathing cycle. During the treatment delivery, the infrared signal combined with the correlation model is used to predict the internal target location 115 ms in the future, so that the radiation beam can be redirected to the actual target location in real-time with a 100 % duty cycle. Additional X-ray images are taken every ~60 s to verify and update the correlation model throughout the treatment [6].

Utilizing the stereoscopic X-rays at fixed angles, XLT has 2-view and 1-view tracking modes. The latter is used when reliable and consistent tumor detection is unachievable in one of the two views. A larger planning target margin in the direction perpendicular to the 1-view plane is used to account for the out-of-plane tumor motion. A retrospective analysis shows that the third-generation XLT algorithm achieved 2-view and 1-view tracking for 64 % and 81 % of cases, respectively [6]. In other words, 36 %–19 % of lung cancer patients eligible for SBRT did not receive the full benefit of dynamic tumor tracking due to low tumor conspicuity in X-ray images, resulting in greater normal tissue or compromised tumor doses. It is, therefore, essential to advance algorithms for lung tumor detection and tracking of smaller and less conspicuous tumors.

Emerging deep learning (DL) neural networks (NNs), including convolutional neural networks (CNNs), Transformers, etc., have revolutionized conventional medical imaging processing tasks, including classification [14], objection detection [15], segmentation [16], and template matching (TM) [17]. TM finds local matches in an image to a provided template image, essentially the lung tumor tracking problem. NNs are superior to their conventional counterparts in part attributed to their learning ability from large training datasets. In other words, DL-based TM frameworks are data-hungry [17,18], but the available data are often insufficient in both quantity and quality. The data scarcity challenge is exacerbated because of the data imbalance between normal anatomies vs. anomalies, e.g., tumors [19]. Providentially, recent research shows that an NN generator can capture low-level image statistics for denoising, super-resolution, and inpainting tasks before training on domain-specific data [20].

To this end, we explore the feasibility of designing a training-free DL-based TM framework named Deep Match to improve the current CyberKnife XLT tracking.

Methods and materials

Data cohort

The study was conducted under Institute Review Board (IRB) - approval (IRB # 20-32527) and included 10 de-identified patients (38 fractions in total) who underwent lung SBRT with 1-view XLT between 2015 and 2023. Detailed specification of the data cohort is listed in Table 1. Both seed X-ray images for building the correlative prediction model and treatment delivery are included.

Data preparation

Coordinate system transformation

Although CyberKnife imaging apparatus has been previously reported, a description of details pertinent to the current study is essential. As shown in Fig. 1 **(a)**, the two orthogonal X-ray sources installed on the ceiling and two amorphous silicon panel detectors tilted 45° from the

Table 1
Patient profiles in the data cohort.

Patient	Size range (mm^3)	Trackable view	Fractions	Total frames
P1	$> 15^3$	B	5	263
P2		A	5	344
P3	$(10^3, 15^3)$	A	5	178
P4		A	3	73
P5		A	2	113
P6		A	2	524
P7		B	1	586
P8	$< 10^3$	A	5	250
P9		B	5	285
P10		B	5	61
Total			38	2661

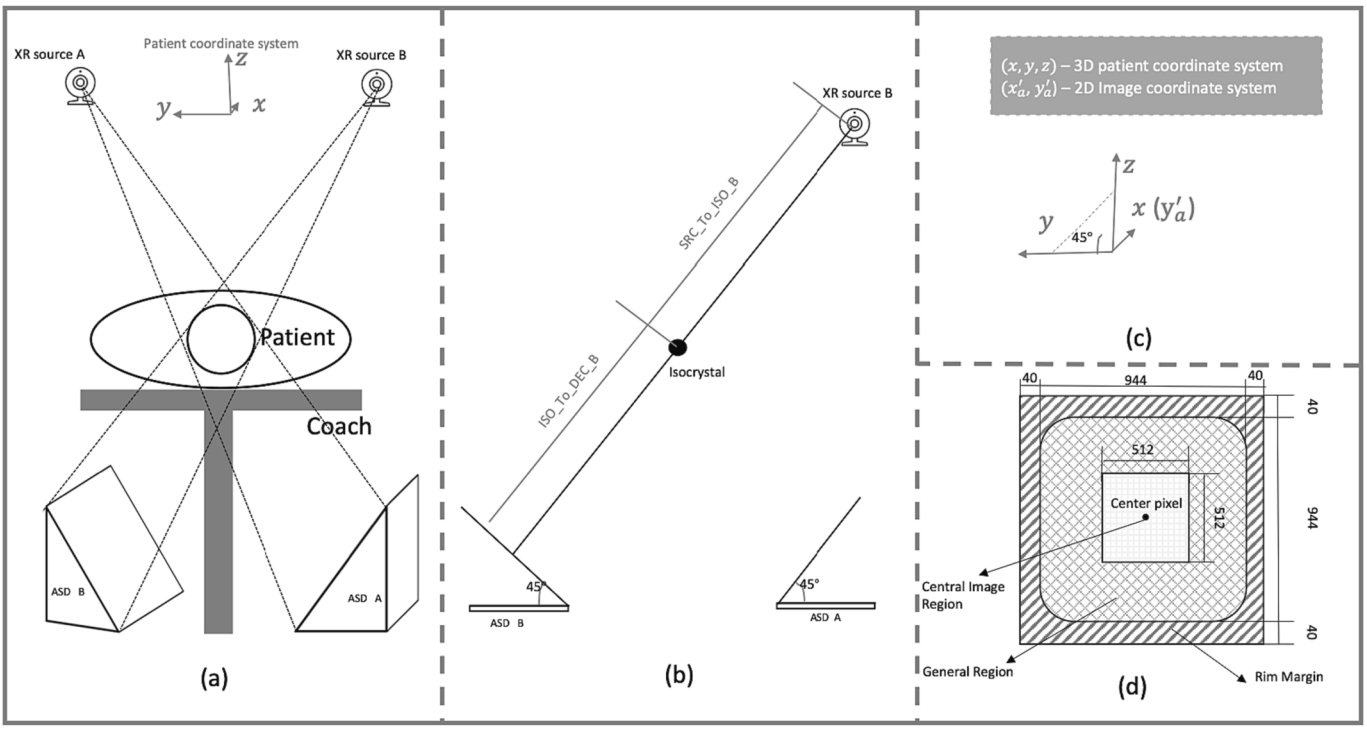

Fig. 1. Diagram of CyberKnife imaging system. **(a)** A schematic of treatment setup. (b) Patient coordinate system geometry. (c) Coordinate transformation between the patient coordinate system and 2D imaging coordinate system (View A). (d) Definition of detector imaging regions.

floor consistute the main hardware components of the CyberKnife image guidance system. Although in modern CyberKnife systems the detectors are embedded horizontally in the floor, images are reconstructed in the 45° geometry. Two CyberKnife coordinate systems related to the current work are illustrated – the three-dimensional patient coordinate system (3D-PCS) and the two-dimensional imaging coordinate system (2D-IMS). The 3D-PCS is based on a supine patient and has $+$X representing the inferior direction, $+$Y representing the left direction, and $+$Z representing the anterior direction. 3D-PCS is of millimeter (mm) unit. The origin of 3D-PCS is the machine isocenter, which coincides with the center of the imaging system. The 2D-IMS is used to define the geometry of the live X-ray and DRR images, which were generated in the Precision planning system for CK. 2D-IMS is of the absolute pixel unit. The origin of 2D-IMS is at the center of the detector/image. To transform 3D-PCS into 2D-IMS, coordinate projection, magnification, and unit transformation were provided. Coordinate projection is explained in **Eq. (1)** and Fig. 1 **(c)**. X direction of 3D-PCS is $Y^{'}$ in 2D-IMS, whereas $X^{'}$ requires projection from Y and Z. Coordinate magnification is shown in **Eq. (2)** and Fig. 1 **(b)**. Lastly, the CyberKnife X-ray detector has a detection area of $40cm \times 40cm$ divided into 1024×1024 pixels with pixel spacing of $0.39 \times 0.39mm$ (Fig. 1 **(d)**). Unprocessed (1024×1024 pixels) and processed (512×512 pixels) live X-ray modalities are generated per treatment delivery, with processed having better imaging contrast. Thus, we chose processed X-ray as our image source, and our unit transformation follows **Eq. (3)**.

$$Y^{'} = \left((-1)^{n} \bullet cos45^{\circ} \quad (-1)^{n+1} cos45^{\circ} \right) \begin{pmatrix} Y \\ Z \end{pmatrix}; X^{'} = X \tag{1}$$

$$n = \begin{cases} 0, View\ \ B \\ 1, View\ \ A \end{cases}$$

where X, Y, Z represent 3D-PCS coordinate system and $X^{'}, Y^{'}$ represents intermediate transformation results to 2D-IMS.

$$\begin{pmatrix} X^{''} \\ Y^{''} \end{pmatrix} = \frac{SRC2ISO_{\mathscr{V}} + ISO2DEC_{\mathscr{V}}}{SRC2ISO_{\mathscr{V}}} \begin{pmatrix} X^{'} \\ Y^{'} \end{pmatrix} \tag{2}$$

where X'', Y'' represents intermediate transformation results to 2D-IMS, $SRC2ISO$ refers to source to isocenter distance, $ISO2DEC$ refers to isocenter to detector distance, and $\mathscr{V}$ refers to view A or B. $SRC2ISO_A$, $ISO2DEC_A$, $SRC2ISO_B$, $ISO2DEC_B$ at our facility is specified at 2315.6 mm, 1420 mm, 2319.1 mm, and 1419 mm.

$$\begin{pmatrix} X^{''} \\ Y^{''} \end{pmatrix} = \frac{L_P}{L_D} \begin{pmatrix} X^{''} \\ Y^{''} \end{pmatrix} \tag{3}$$

where L_P is the number of pixels on the detector and L_D is the absolute length of the detector $L_P = 512, L_D = 200mm$ in this case.

Data processing

DRR template is cropped from DRR images of size $W \times H = 512 \times 512$ with tumor bounding boxes (BBoxes) provided from machine log files. BBoxe sizes are approximately $20mm \times 20mm$ and are provided with 3D-PCS format. Coordinate system transformation is applied to BBoxes before cropping the template from DRR.

Processed live X-ray images of size, $W \times H = 512 \times 512$ are selected as search inputs. Tumor locations generated from the CyberKnife Synchrony trackable view throughout treatment delivery are utilized as our template matching ground truth. Tumor coordinates (defined as the centroid of the tumor tracking volume) are stored in the 3D-PCS format in the ModelPoints.log file. Coordinate system transformation is applied to align the tumor coordinates with the live X-ray imaging system.

Sanity check

To verify the accuracy of our data preparation process, we tested our data processing workflow on 5 patients treated with Synchrony tracking based on a single fiducial. As exemplified in Fig. 2, the coordinates of the fiducial extracted from the ModelPoints.log file and converted into the imaging coordinate system agree well with visible fiducials on the live treatment delivery X-rays from both views, validating the correctness of

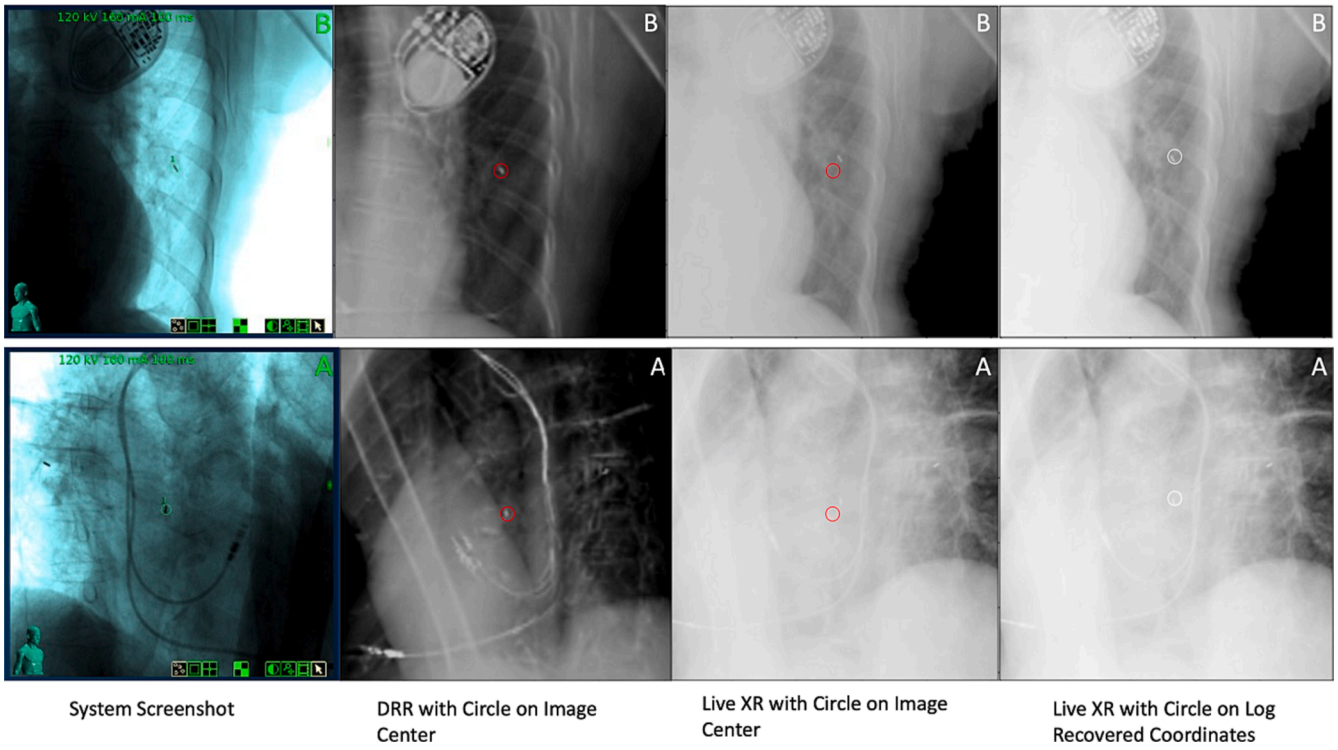

Fig. 2. Data preparation sanity check on fiducial tracking case. Red circles are the image center, and white circles are the log recovered coordinates (circle diameter = 10 pixels/3.9 mm). The image center on DRR and live XR is calculated from the center of gross tumor volume on the DRR image. The DRR image center is then set as the origin for 2D IMS and the isocenter per treatment delivery. White circle on the fourth column is calculated via moving the log recovered fiducial coordinates from the origin (red circle on the third column). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

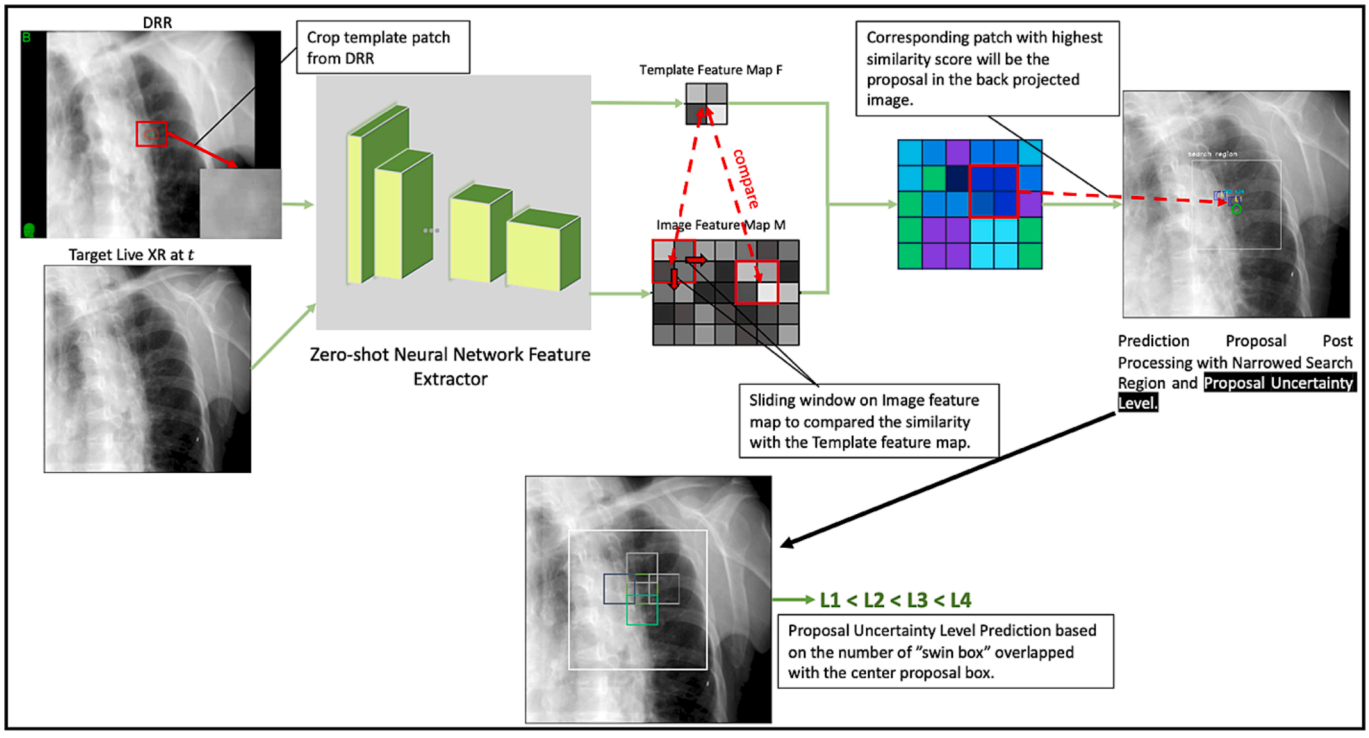

Fig. 3. Illustration of the Deep Match algorithm.

model geometry.

Deep match algorithm

Inspired by Kim et al. [21], Deep Match, as shown in Fig. 3, is defined to localize the most similar patch within the live X-ray image at time t $I_t \in R^{w\times h}$ given DRR template $T \in R^{m\times n}$ generated per treatment planning, where w and h represent the width and height of the image and m and n correspond to the width and height of the template. Deep Match has stages of NN feature extraction, similarity measurements and proposal generation, location refinement, and uncertainty level prediction. Deep Match is essentially a versatile pipeline, having each stage plug-and-playable. We will demonstrate this framework with VGG-16 [22] as a feature map extractor, a combined similarity metric of NCC and pattern intensity (PI) [12], and the swin window mechanism as proposal uncertainty level prediction.

Deep neural network feature extraction

The feature extractor in Deep Match directly uses ImageNet [23] pretrained weights to construct a data-efficient framework, without further task-specific model fine-tuning. Specifically, VGG-16 is used to extract feature maps from both the DRR template and the input images. While the pretrained VGG network may not grasp task-specific contextual information, it still captures crucial low-level statistics through local and translational invariant filtering operations inherent to any NN. These statistics effectively capture pixel neighborhood relationships at multiple scales, leading to improved template searching compared to using raw images [20]. Following Kim et al. [21], we designed our VGG network to be scale adaptive, where rescaling the template or images into a particular size as standard CNN methods is no longer needed. In this way, we avoid tumor deformation and degraded structural information caused by image rescaling.

Similarity measurements

As NCC is commonly used as a similarity metric in natural imaging matching, and studies show that PI is a strong artifact-insenstive metric for medical imaging registration [12]. We designed our similarity measurement to be a fusion of the two in the format shown in **Eqs. (4)–(7)**. Since S is measured in feature map domain, we will back-project the BBox location found on feature map to that in the original image domain for proposal generation.

$$S = \alpha S_{NCC} + (1-\alpha) S_{PI} \tag{4}$$

$$S_{NCC} = \frac{< F_T, \widetilde{F}_{I_t} >}{|F_T| \left| \widetilde{F}_{I_t} \right|} \tag{5}$$

$$\Delta F(i,j) = \widetilde{F}_{I_t}(i,j) - F_T(i,j) \tag{6}$$

$$S_{PI} = \sum_{i,j} \frac{\sigma^2}{\sigma^2 + (\Delta F(i,j) - \Delta F(i,j-1))^2} + \frac{\sigma^2}{\sigma^2 + (\Delta F(i,j) - \Delta F(i-1,j))^2} + \frac{\sigma^2}{\sigma^2 + (\Delta F(i,j) - \Delta F(i-1,j-1))^2} + \frac{\sigma^2}{\sigma^2 + (\Delta F(i,j) - \Delta F(i-1,j+1))^2} \tag{7}$$

where α is a self-defined hyperparameter (0.5 currently used), F_{I_t} and F_T are the NN-processed feature maps of live X-ray images and DRR template, $\widetilde{F}_{I_t} = F_{I_t}(i:i+h_{F_T}-1, j:j+w_{F_T}-1)$ is a feature map patch extracted from F_{I_t} with a width of w_{F_T} and h_{F_T} equal to the size of F_T, ΔF is the pixel intensity difference, and σ^2 is a self-defined weight constant. According to Fu et al. [7], 20 is used for σ^2.

Location refinement

For more robust tracking, proposal refinement was conducted. First, to enable more precise proposal coordinates, a similarity layer-back-projected weighted sum of the original location in a 3×4 patch was applied to the proposal BBox coordinates, as shown in **Eq. (8)**. Second, a user-defined search window was specified to exclude unreasonable BBox proposals generated. A range of 25 mm from the center of the image is currently used, which is wide enough to include irregular respiratory motions.

$$\begin{pmatrix} \bar{x}(i,j) \\ \bar{y}(i,j) \end{pmatrix} = \frac{\sum_{u=-1}^{1}\sum_{v=-2}^{1} S(i+u, j+v) \bullet \left[\begin{pmatrix} x(i,j) \\ y(i,j) \end{pmatrix} + v \bullet \prod_{k=1}^{l-1} s^k \right]}{\sum_{u=-1}^{1}\sum_{v=-2}^{1} S(i+u, j+v)} \tag{8}$$

where s^k is the stride of *kth* layer in the NN feature extractor ($s = 2$ for all layers), and l is the total layers.

Uncertainty level prediction

As shown in the second row of Fig. 3, we defined a "swin window" mechanism to determine localization confidence. The "swin window" mechanism performs repeated detection box shifts along the up/down/left/right directions by half the box length. The more shifted window BBox proposals overlapped with the non-shifted BBox proposal, the higher prediction confidence. As shown in Fig. 3**, L4** represents the highest level of confidence (four directional shifted regions are located by Deep Match) whereas L1 represents the lowest level (one directional shifted region is located by Deep Match). Considering the efficiency of the matching algorithm, we shifted windows in four directions – top/down/left/right with four levels of confidence.

Model evaluation

CyberKnife Tracking System needs to register information from both views to determine tumor 3D motion. As mentioned, CyberKnife XLT 1-view tracking cannot determine motion perpendicular to the imaging plane. Live X-ray imaging geometry shows that the superior-inferior tumor direction, which tends to be the largest component of respiratory motion, is detected in both X-ray views. Therefore, the superior-inferior target location in the trackable view was used as the ground truth for evaluation. The superior-inferior-difference (SID) between the target location proposed by Deep Match to the ground truth was calculated. We organized four evaluation metrics – 1.5 mm-Hit, 3 mm-Hit, 5 mm-Hit, and SID. The Xmm-Hit is a prediction accuracy metric that calculated the images with BBox predicted within X mm SID tolerance over total images. SID was evaluated in 3D-PCS, and the algorithm was evaluated patient-wise.

Benchmark algorithm and model training

Since XLT algorithms are close-sourced, we coded second and third-generation XLT (XLT 2nd Gen and XLT 3rd Gen) according to literature [6,7] for benchmark purposes. The performance of our implementation was validated against the commercial XLT.

All the experiments are parallelly run on a $4 \times$ *RTXA*6000 GPU clusters with a batch size of 4×1.

Results

The comparison of our XLT implementations, Deep Match, and the commercial XLT, is shown in Table 2**.** On the trackable views, all the models can achieve >90 % of 3 mm-Hit and ~1.5 mm SID on most patients (except XLT 2nd Gen on P7), with Deep Match slightly outperforming the benchmark models - ~2% higher on 3 mm-Hit and ~0.2 mm lower on SID.

The results from nontrackable views are shown in Fig. 4 and Table1 nontrackable view. From Table 2, we can observe that XLT 2nd Gen and

Table 2
Experimental results on Deep Match and replicated XLT second- and third-generation (XLT 2nd Gen and XLT 3rd Gen) algorithms. Trackable views refer to XLT 3rd Generation trackable views and vice versa for the category of nontrackable views. 1.5/3/5mm-Hit refers to the percentage of predictions within 1.5/3/5 mm superior/inferior margin from ground truth.

View	Patient	Deep Match				XLT 2nd Gen		XLT 3rd Gen	
		1.5 mm-Hit	*3 mm-Hit*	*5 mm-Hit*	*SID(mm)*	*3 mm-Hit*	*SID(mm)*	*3 mm-Hit*	*SID(mm)*
Trackable	P1	89.04 %	98.37 %	100.00 %	1.23 ± 1.05	96.55 %	1.32 ± 1.06	97.62 %	1.28 ± 1.05
	P2	92.00 %	95.00 %	100.00 %	1.19 ± 1.04	93.99 %	1.48 ± 1.11	94.86 %	1.37 ± 1.06
	P3	80.33 %	98.31 %	98.31 %	1.09 ± 1.03	96.07 %	1.36 ± 1.18	96.07 %	1.38 ± 1.17
	P4	90.55 %	91.94 %	92.57 %	1.08 ± 0.81	90.38 %	1.48 ± 0.43	91.94 %	1.53 ± 0.67
	P5	92.00 %	97.21 %	100.00 %	1.19 ± 1.25	100.00 %	1.13 ± 1.09	95.00 %	1.03 ± 1.22
	P6	91.25 %	94.86 %	99.37 %	1.41 ± 1.05	92.67 %	1.63 ± 1.21	93.04 %	1.60 ± 1.16
	P7	94.36 %	97.45 %	98.91 %	0.95 ± 0.80	88.54 %	1.01 ± 0.98	91.45 %	1.24 ± 1.05
	P8	88.26 %	98.25 %	99.37 %	1.36 ± 1.04	95.23 %	1.69 ± 0.95	96.58 %	1.74 ± 0.91
	P9	89.23 %	95.19 %	96.45 %	1.22 ± 1.01	93.33 %	1.37 ± 1.09	93.70 %	1.36 ± 1.22
	P10	91.32 %	97.61 %	97.61 %	1.26 ± 1.07	91.89 %	1.35 ± 0.67	92.57 %	1.44 ± 1.05
Nontrackable	P1	93.54 %	99.62 %	100.00 %	2.13 ± 1.38	0.00 %	–	0.00 %	–
	P2	31.34 %	66.79 %	80.00 %	1.96 ± 1.16	47.38 %	2.21 ± 2.04	48.89 %	2.59 ± 1.99
	P3	71.35 %	91.60 %	96.60 %	2.22 ± 1.41	27.00 %	10.66 ± 3.74	0.00 %	14.63 ± 1.55
	P4	87.50 %	95.83 %	95.83 %	1.67 ± 1.24	72.22 %	1.98 ± 1.78	1.38 %	4.91 ± 2.98
	P5	48.00 %	89.00 %	100.00 %	1.81 ± 1.24	27.00 %	4.37 ± 3.73	0.00 %	14.16 ± 3.22
	P6	30.66 %	52.93 %	83.40 %	2.40 ± 1.42	25.98 %	3.44 ± 1.92	38.67 %	$3 .27 \pm 1.93$
	P7	66.36 %	90.00 %	97.63 %	1.32 ± 1.07	50.55 %	3.21 ± 1.89	47.82 %	3.39 ± 2.18
	P8	76.71 %	85.94 %	89.16 %	1.58 ± 1.20	0.00 %	13.40 ± 2.04	0.00 %	12.74 ± 0.81
	P9	37.69 %	70.90 %	83.58 %	1.80 ± 1.15	14.93 %	3.42 ± 4.08	29.10 %	3.09 ± 3.32
	P10	48.94 %	89.36 %	100.00 %	1.92 ± 1.24	76.60 %	2.48 ± 1.72	72.97 %	2.39 ± 1.07

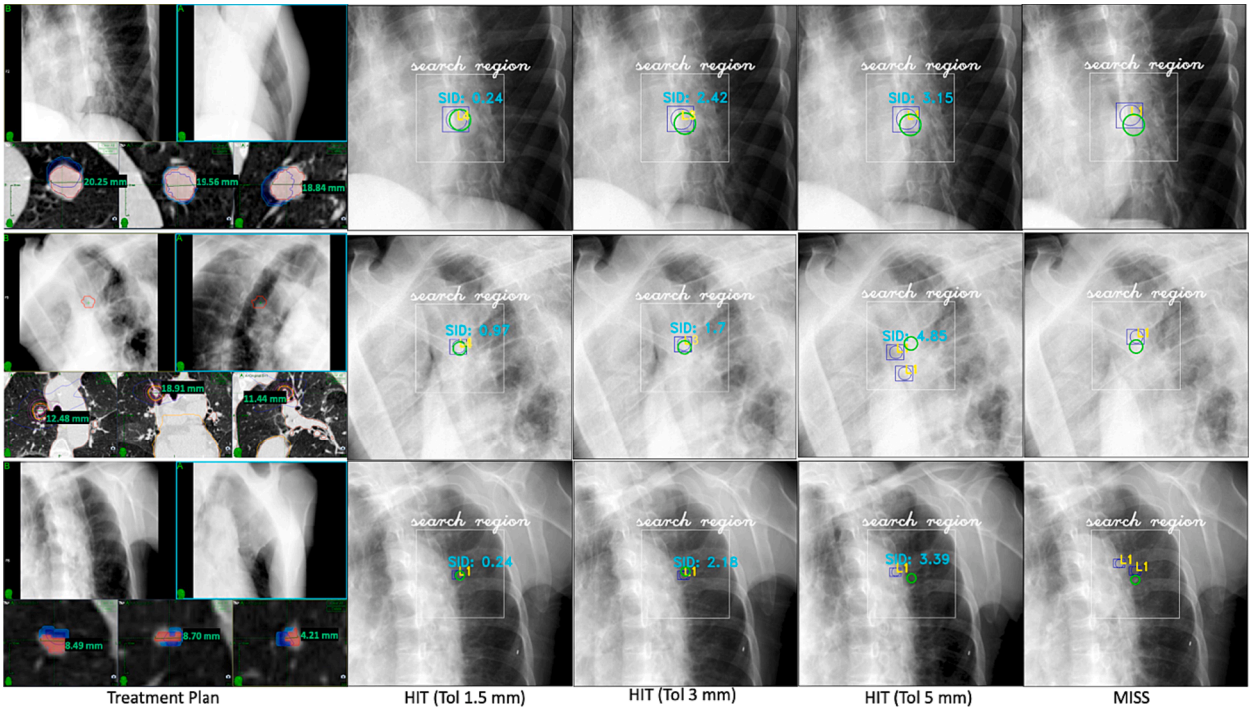

Fig. 4. Deep Match tracking results demonstration from three individual patients on 3rd Generation, patient #2 (P2), patient #3 (P3) and patient #8 (P8). P2 represents the XLT patient group with tumor size greater than $15mm \times 15mm \times 15mm$, P3 represents the XLT patient group with tumor size between $10mm \times 10mm \times 10mm$ and $15mm \times 15mm \times 15mm$, and P8 represents the XLT patient group with tumor size smaller than $10mm \times 10mm \times 10mm$.

XLT 3rd Gen achieve 0 % 3 mm-Hit (P1, P3, P5, and P8) or over 5 mm SID margin (P2, P5, and P8), which is consistent with the fact that these lesions could not be tracked by the CyberKnife system during patient treatment. >80 % patients achieve 3 mm-Hit (except P2, P6, and P9) and 100 % patients achieves <3 mm SID with ~1 mm standard deviation with Deep Match. Fig. 4 shows three representative patients – P2, P3, and P8. P2 has a tumor size $>15mm \times 15mm \times 15mm$ and the target is superimposed with the spine. P3 has tumor size between $10mm \times 10mm \times 10mm$ and $15mm \times 15mm \times 15mm$ and the target is superimposed with cardiovascular structures. P8 has tumor size $<10mm \times 10mm \times 10mm$with no anatomy overlapping issue. We can see that Deep Match can generate high confidence BBox proposal (L3, L4) on these patients. Although some images have more than one BBox proposal, the BBoxes with the highest confidence level proposal consistently agree with the ground truth position.

Discussion

Effective management of internal organ motion is of paramount importance in radiotherapy. The importance has elevated with

hypofractionation and SBRT, where high fractional doses and steep dose gradients are observed. Although surface markers can be correlated with internal organ motion for lung cancer, the correlation can be unreliable without periodic direct visualization of the lung tumor during treatment. Among the overwhelming majority of radiotherapy equipment, X-rays remain the only imaging method for real-time internal anatomy monitoring. While implanted fiducials can be accurately tracked, the implantation procedure is demanding and medically contraindicated by some patients. For robotic tumor tracking, fiducial-free X-ray tracking is attractive to exploit the intrinsic tumor lung parenchyma contrast. Yet, existing tracking algorithms based on raw imaging features are limited in locating small tumors and tumors obscured by dense overlapping structures in the X-rays [6]. The tracking performance could benefit from the recent DL revolution leading to numerous networks outperforming conventional image processing methods. However, most DL networks require large, annotated datasets to train, a condition that cannot be met for the specific study.

Therefore, we proposed Deep Match in this work. Deep Match is a zero-shot versatile DL-based template matching pipeline consisting of four stages – NN feature extractor, feature map similarity measurements and tracking proposal, proposal refinement, and tracking confidence prediction. Each stage of Deep Match can be substituted by user self-definition. We demonstrate the effectiveness of Deep Match using VGG16 as feature extractor, PI and NCC fusion as similarity measurements, weighted local box proposal adjustments and respiratory motion margin defined searching window as local refinement, and swin window as prediction uncertainty level measurements. Deep Match is validated on a CyberKnife 1-view patient cohort with 10 patients with a total of 38 fractions (2661 images).

For all trackable views, Deep Match identified the SI target location within 1.5 mm of the ground truth location (1.20 mm average, range 0.95–1.41 mm), which is below the overall system margin measured in end-to-end tests for motion tracking treatments [20].

For all non-trackable views, Deep Match identified the SI target location within 2.5 mm of the ground truth (1.88 mm average, range 1.32–2.40 mm), which is below the target relative distance along the alignment center of the superior-inferior (SI) axis in the 2 projections (dxAB) threshold used clinically. The target dxAB threshold defines the maximum difference in the SI target location extracted from the two x-ray images. The default value is 2.5 mm. Large values of dxAB might indicate the presence of rotation in patient setup.

For lung SBRT patients, a 5 mm planning target margin is typically used. The 1.5 mm-, 3 mm- and 5 mm-Hit were calculated to evaluate the prediction accuracy of Deep Match. Deep Match achieved an average 5 mm-Hit of 98.3 % (range 92.6–100 %) on the trackable views and 92.6 % (range 80–100 %) on the nontrackable views.

With the introduction of Deep Match, 1-view-only patients are potentially eligible for 2-view tracking with a tightened treatment margin and reduced toxicity. Additionally, since Deep Match can detect smaller and obscured tumors and targets, patients previously excluded from XLT direct tumor tracking could benefit from fiducial-free SBRT [6].

Unlike most current DL medical imaging applications, Deep Match, requiring no network training, is data-independent. Essentially, any NN digests image I as a parametrization $f_\theta(I)$ process with standard structure and alternating filtering operations (i.e., linear convolution, upsampling, nonlinear activation functions, etc). The local and translational invariant nature of convolutions and the sequential usage of such operators effectively capture the pixel neighborhood relationships at multiple scales. Without task-specific training and fine-tuning, the data-efficient framework constructed using ImageNet pretrained weights may not comprehend contextual information, such as lung tumor morphology. Albeitly, low-level statistics in NN processed feature maps contribute to an improved template searching task compared to using raw images [20]. Therefore, while Deep Match may sacrifice some performance compared to models trained and fine-tuned on specific data, it gains the advantage of being more generalizable and less prone to overfitting. Another distinct benefit of Deep Match is the ability to quantify the tracking uncertainty with a "swing window" mechanism, which is crucial for the safe clinical adoption of a DL product [24] by alerting the operator of reduced tumor tracking confidence for necessary treatment pauses.

The current study is based on a specific robotic radiotherapy platform, but its applications are generalizable to other X-ray-based radiotherapy platforms, including the widely available C-arm gantry systems [9,25] with an X-ray onboard imager. One potential advantage of applying this method to C-arm is the additional flexibility of acquiring X-rays from different angles and avoiding dense overlapping structures, thus improving tracking of small tumors.

Nevertheless, the current work can be improved and expanded in several areas. From the algorithm perspective, improved tracking may be attained with techniques such as rib or spine removal [26]. As a zero-shot method, Deep Match network parameters are not task-optimal. Better performance can be expected by fine-tuning the NN feature extractor based on chest X-rays for better precision in future work [27]. Lastly, improving soft-tissue visualization on x-ray images is an important aspect to consider, as the physician is ultimately responsible for verifying the correct target alignment by visual inspection of the live camera images and DRRs. For this purpose, dual-energy X-ray imaging has been proposed to enhance soft-tissue visibility and increase tracking accuracy, especially in small lung targets [28]. From the data perspective, for validation, the current study is performed on the available 1-view tracking patients in the database, which is limited in both the data size and available ground truth dimension. Overcoming both challenges requires a larger dataset of patients with low lung tumor conspicuity, yet the ground truth tumor 3D coordinates are available. To curate such a database in future studies, one can conceivably perform imaging inpainting [29] to remove fiducials from 2 to view fiducial tracking patients without impacting the rest of thoracic anatomy.

Conclusion

An improved X-ray fiducial-less lung tumor tracking method – Deep Match – is tested on patients with small and obscured tumors. Markedly improved performance was observed compared with the commercial system using conventional object detection methods, achieving high tumor tracking accuracy on previously untractable targets. The proposed method may further push the current tractable tumor limit to below 15 mm, thereby broadening the patient cohort that stands to benefit from fiducial-free SBRT.

Funding statement

NIH R01CA188300.
DOD W81XWH2210044.

Data Availability Statement for this Work

Research data are stored in an institutional repository and will be shared upon request with the corresponding author.

CRediT authorship contribution statement

Di Xu: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Martina Descovich:** Data curation, Methodology, Resources, Validation, Visualization, Writing – original draft. **Hengjie Liu:** Formal analysis, Investigation, Methodology, Writing – original draft. **Yi Lao:** Methodology. **Alexander R. Gottschalk:** Validation, Visualization. **Ke Sheng:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The study is supported in part by NIH R01CA188300 and CDMRP W81XWH2210044.

References

[1] Lyons G, Quadrelli S, Chimondegy D, Iotti A, Silva C. Tumor size and survival in lung cancer, stage IA. Medicina (B Aires) 2008;68:23–30.
[2] National Institutes of Health. SEER stat fact sheets: Lung and bronchus cancer. Rockville, MD: National Institutes of Health; 2016.
[3] Soldà F, Lodge M, Ashley S, Whitington A, Goldstraw P, Brada M. Stereotactic radiotherapy (SABR) for the treatment of primary non-small cell lung cancer; Systematic review and comparison with a surgical cohort. Radiother Oncol 2013; 109:1–7. https://doi.org/10.1016/j.radonc.2013.09.006.
[4] Song AJ, Evans N, Cowan S, et al. Stereotactic body radiation therapy (SBRT) for patients with stage I non-small cell lung cancer is applicable to more tumors than sublobar resection. J Thorac Dis 2021;13:1576–83. https://doi.org/10.21037/jtd-20-2001.
[5] Fernando HC, Timmerman R. American College of Surgeons Oncology Group Z4099/Radiation Therapy Oncology Group 1021: A randomized study of sublobar resection compared with stereotactic body radiotherapy for high-risk stage I non–small cell lung cancer. J Thorac Cardiovasc Surg 2012;144:S35–8. https://doi.org/10.1016/j.jtcvs.2012.06.003.
[6] Kilby W, Naylor M, Dooley JR, Maurer CR, Sayeh S. A Technical Overview of the CyberKnife System. In: Handbook of Robotic and Image-Guided Surgery. Elsevier; 2020. p. 15–38. https://doi.org/10.1016/B978-0-12-814245-5.00002-5.
[7] Fu D, Kahn R, Wang B, et al. Xsight Lung Tracking System: A Fiducial-Less Method for Respiratory Motion Tracking. In: Urschel HC, Kresl JJ, Luketich JD, Papiez L, Timmerman RD, Schulz RA, editors. Treating Tumors That Move with Respiration. Berlin Heidelberg: Springer; 2007. p. 265–82. https://doi.org/10.1007/978-3-540-69886-9_26.
[8] Casutt A, Kinj R, Ozsahin EM, von Garnier C, Lovis A. Fiducial markers for stereotactic lung radiation therapy: review of the transthoracic, endovascular and endobronchial approaches. Eur Respir Rev 2022;31:210149. https://doi.org/10.1183/16000617.0149-2021.
[9] Nuyttens JJ, Prévost JB, Praag J, et al. Lung tumor tracking during stereotactic radiotherapy treatment with the CyberKnife: Marker placement and early results. Acta Oncologica 2006;45:961–5. https://doi.org/10.1080/02841860600902205.
[10] Le QT, Loo BW, Ho A, et al. Results of a phase I dose-escalation study using single-fraction stereotactic radiotherapy for lung tumors. J Thorac Oncol 2006;1:802–9.
[11] Whyte RI, Crownover R, Murphy MJ, et al. Stereotactic radiosurgery for lung tumors: preliminary report of a phase I trial. Ann Thorac Surg 2003;75:1097–101. https://doi.org/10.1016/s0003-4975(02)04681-7.
[12] Penney GP, Weese J, Little JA, Desmedt P, Hill DL, Hawkes DJ. A comparison of similarity measures for use in 2-D-3-D medical image registration. IEEE Trans Med Imaging 1998;17:586–95. https://doi.org/10.1109/42.730403.
[13] Jordan P, West J, Sharda A, Maurer C. SU-GG-J-24: retrospective clinical data analysis of fiducial-free lung tracking. Med Phys 2010;37:3150. https://doi.org/10.1118/1.3468248.
[14] Qin J, Pan W, Xiang X, Tan Y, Hou G. A biological image classification method based on improved CNN. Eco Inform 2020;58:101093. https://doi.org/10.1016/j.ecoinf.2020.101093.
[15] Xu D, Ma M, Cao M, et al. Mask R-CNN assisted 2.5D object detection pipeline of 68Ga-PSMA-11 PET/CT-positive metastatic pelvic lymph node after radical prostatectomy from solely CT imaging. Sci Rep 2023;13:1696. https://doi.org/10.1038/s41598-023-28669-y.
[16] Xu D, Ma TM, Savjani R, et al. Fully automated segmentation of prostatic urethra for MR-guided radiation therapy. Med Phys 2023;50:354–64. https://doi.org/10.1002/mp.15983.
[17] Li L, Han L, Ding M, Cao H, Hu H. A deep learning semantic template matching framework for remote sensing image registration. ISPRS J Photogramm Remote Sens 2021;181:205–17. https://doi.org/10.1016/j.isprsjprs.2021.09.012.
[18] Wang S, Wang H, Zhou Y, et al. Automatic laser profile recognition and fast tracking for structured light measurement using deep learning and template matching. Measurement 2021;169:108362. https://doi.org/10.1016/j.measurement.2020.108362.
[19] Willemink MJ, Koszek WA, Hardell C, et al. Preparing medical imaging data for machine learning. Radiology 2020;295:4–15. https://doi.org/10.1148/radiol.2020192224.
[20] Ulyanov D, Vedaldi A, Lempitsky V. Deep Image Prior. Published online 2017. doi: 10.48550/ARXIV.1711.10925.
[21] Kim J, Kim J, Choi S, Hasan MA, Kim C. Robust template matching using scale-adaptive deep convolutional features. In: 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE; 2017: 708-711. doi: 10.1109/APSIPA.2017.8282124.
[22] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:14091556 [cs]*. Published online April 10, 2015. Accessed February 17, 2022. http://arxiv.org/abs/1409.1556.
[23] Deng J, Dong W, Socher R, Li LJ, Kai Li, Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2009:248-255. doi: 10.1109/CVPR.2009.5206848.
[24] Patel VL, Shortliffe EH, Stefanelli M, et al. The coming of age of artificial intelligence in medicine. Artif Intell Med 2009;46:5–17. https://doi.org/10.1016/j.artmed.2008.07.017.
[25] Kaur M, Luce J, Lehmann M, et al. Effect of scattered megavoltage x-rays on markerless tumor tracking using dual energy kilovoltage imaging. *J Appl Clin Med Phys*. Published online April 18, 2023:e13993. doi: 10.1002/acm2.13993.
[26] Xu D, Xu Q, Nhieu K, Ruan D, Sheng K. An efficient and robust method for chest X-ray rib suppression that improves pulmonary abnormality diagnosis. Diagnostics 2023;13:1652. https://doi.org/10.3390/diagnostics13091652.
[27] Xian Y, Lampert CH, Schiele B, Akata Z. Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly. IEEE Trans Pattern Anal Mach Intell 2019;41:2251–65. https://doi.org/10.1109/TPAMI.2018.2857768.
[28] Menten MJ, Fast MF, Nill S, Oelfke U. Using dual-energy x-ray imaging to enhance automated lung tumor tracking during real-time adaptive radiotherapy. Med Phys 2015;42:6987–98. https://doi.org/10.1118/1.4935431.
[29] Guillemot C, Le Meur O. Image inpainting: overview and recent advances. IEEE Signal Process Mag 2014;31:127–44. https://doi.org/10.1109/MSP.2013.2273004.