Development and evaluation of ultrasound image tracking technology based on Mask R-CNN applied to respiratory motion compensation system

Lai-Lei Ting[1#], Ming-Lu Guo[2#], Ai-Ho Liao[3,4], Sen-Ting Cheng[2], Hsiao-Wei Yu[1,5], Subramaninan Ramanathan[6], Hong Zhou[7], Catherin Meena Boominathan[8], Shiu-Chen Jeng[1,9], Jeng-Fong Chiou[1,10,11], Chia-Chun Kuo[1,12,13], Ho-Chiao Chuang[2^]

[1]Department of Radiation Oncology, Taipei Medical University Hospital, Taipei, Taiwan; [2]Department of Mechanical Engineering, National Taipei University of Technology, Taipei, Taiwan; [3]Graduate Institute of Biomedical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan; [4]Department of Biomedical Engineering, National Defense Medical Center, Taipei, Taiwan; [5]School of Biomedical Engineering, College of Biomedical Engineering, Taipei Medical University, Taipei, Taiwan; [6]Department of Chemical Technology, Faculty of Science, Chulalongkorn University, Bangkok, Thailand; [7]Department of Electronics, Information and Communication Engineering, Osaka Institute of Technology, Osaka, Japan; [8]PG & Research Department of Chemistry, Bishop Heber College, Tiruchirappalli, India; [9]School of Dentistry, College of Oral Medicine, Taipei Medical University, Taipei, Taiwan; [10]Department of Radiology, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan; [11]Taipei Cancer Center, Taipei Medical University, Taipei, Taiwan; [12]Department of Radiation Oncology, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan; [13]School of Health Care Administration, College of Management, Taipei Medical University, Taipei, Taiwan

Contributions: (I) Conception and design: ML Guo, ST Cheng; (II) Administrative support: HC Chuang, JF Chiou; (III) Provision of study materials or patients: LL Ting, CC Kuo, SC Jeng; (IV) Collection and assembly of data: AH Liao, HW Yu; (V) Data analysis and interpretation: S Ramanathan, H Zhou, CM Boominathan; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Prof. Ho-Chiao Chuang, PhD. Department of Mechanical Engineering, National Taipei University of Technology, No. 1, Sec. 3, Chung-Hsiao E. Rd., Taipei 10608, Taiwan. Email: hchuang@mail.ntut.edu.tw; Dr. Chia-Chun Kuo, Master. Department of Radiation Oncology, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan; School of Health Care Administration, College of Management, Taipei Medical University, Taipei, Taiwan; Department of Radiation Oncology, Taipei Medical University Hospital, No. 252, Wuxing St, Xinyi District, Taipei 110, Taiwan. Email: b8601093@tmu.edu.tw.

Background: For respiration induced tumor displacement during a radiation therapy, a common method to prevent the extra radiation is image-guided radiation therapy. Moreover, mask region-based convolutional neural networks (Mask R-CNN) is one of the state-of-the-art (SOTA) object detection frameworks capable of conducting object classification, localization, and pixel-level instance segmentation.

Methods: We developed a novel ultrasound image tracking technology based on Mask R-CNN for stable tracking of the detected diaphragm motion and applied to the respiratory motion compensation system (RMCS). For training Mask R-CNN, 1800 ultrasonic images of the human diaphragm are collected. Subsequently, an ultrasonic image tracking algorithm was developed to compute the mean pixel coordinates of the diaphragm detected by Mask R-CNN. These calculated coordinates are then utilized by the RMCS for compensation purposes. The tracking similarity verification experiment of mask ultrasonic imaging tracking algorithm (M-UITA) is performed.

^ ORCID: 0000-0001-8036-3351.

Results: The correlation between the input signal and the signal tracked by M-UITA was evaluated during the experiment. The average discrete Fréchet distance was less than 4 mm. Subsequently, a respiratory displacement compensation experiment was conducted. The proposed method was compared to UITA, and the compensation rates of three different respiratory signals were calculated and compared. The experimental results showed that the proposed method achieved a 6.22% improvement in compensation rate compared to UITA.

Conclusions: This study introduces a novel method called M-UITA, which offers high tracking precision and excellent stability for monitoring diaphragm movement. Additionally, it eliminates the need for manual parameter adjustments during operation, which is an added advantage.

Keywords: Deep learning; mask region-based convolutional neural networks (Mask R-CNN); respiratory motion; ultrasound image; tumor motion

Introduction

One of the main causes of death in the world is cancer. In 2020, nearly 20 million people suffered from cancer. The top three most common cancers are breast cancer, lung cancer, and colorectal cancer and the number of deaths from lung cancer accounts for the largest proportion of all cancers: 18% of the total number of deaths, according to Sung *et al.* (2021) (1).

Radiation therapy is a commonly used approach for cancer treatment. It delivers high-energy photons to target and destroy cancer cells in a specific area. However, this treatment can also damage surrounding healthy tissues, leading to radiation-induced side effects. According to American Cancer Society's report (2), common side effects of radiation therapy include skin inflammation, fatigue, decreased appetite, and difficulty swallowing. Severe side effects may include spondylitis and myocarditis. Additionally, during the treatment process, respiratory motion can cause organ movement (3-5), leading to excessive tumor displacement, particularly in the case of lung tumors. Consequently, these displacements contribute to the occurrence of radiation side effects.

In 2018, Dhont *et al.* (6) evaluated the 3D motion of 19 lung and 18 liver tumors recorded during stereotactic radiotherapy and compared it with motion on 4-dimensional computed tomography (4D-CT). The results showed that the tumor displacements in the superior-inferior (SI) and anterior-posterior (AP) directions were very large (>5 mm). The tumor displacements increased the planning target volume (PTV), most notably in the SI direction of lung (5–13.7 mm) and the SI direction of liver (5–8 mm). In 2009, Cerviño *et al.* (7) analyzed the correlation between the motion of lung tumor and the motion of diaphragm to research the possibility of the use of diaphragm as a surrogate for tumor motion. This study analyzed 32 cohorts of fluoroscopic image sequences from 10 lung cancer patients and developed two linear models for correlation analysis. The results suggested that the correlation between diaphragm motion and tumor motion varies by patient, but in general there is a linear relationship between diaphragm and tumor motion. In 2014, Yang *et al.* (8) evaluated the correlation factor between liver tumor motion and diaphragm displacement. The results showed that the diaphragm motion and tumor motion are highly related. In 2021, Li *et al.* (9) verified the feasibility of using the diaphragm as a tracking surrogate of lung tumor in the treatment of Cyberknife Synchrony. The results of the research showed that there was a good consistency between the lung tumor and the diaphragm in SI and right-left (RL) directions. The research pointed out that the tracking of the diaphragm helps patients with unobservable lung tumors to better perform CyberKnife Synchrony.

For tumor displacement during a radiation therapy, increasing the PTV is a conservative approach. However, enlarging the volume leads the healthy tissue around the tumor to an extra exposure to radiation beam. A common method to prevent the extra radiation is image-guided radiation therapy (10-13). Generally, linear accelerators (LINACs) for radiation therapy are equipped with cone beam computed tomography (CBCT), which can be used to scan patient's bone or metal
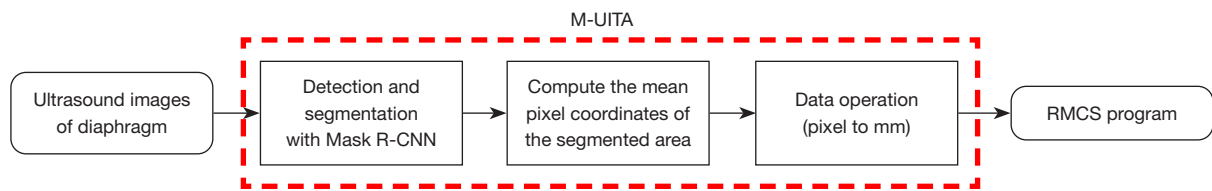
Figure 1 The workflow of M-UITA. M-UITA, mask ultrasonic imaging tracking algorithm; R-CNN, region-based convolutional neural networks; RMCS, respiratory motion compensation system.

implants. The PTV would be registered with the tumor's actual location by image fusion technology (14). However, potential effects of additional CBCT dose (15) on patients as well as the issue of low contrast of soft tissue (16) while imaging should be concerned.

Several studies on motion management were reported recently (17-19). In 2016, we proposed an ultrasound image tracking algorithm, UITA, to track diaphragm motion (20). UITA was applied to a respiratory motion compensation system (RMCS) (21) for tumor motion reduction. UITA analyzes the relative displacement and brightness difference of the region of interest, which is selected manually and used as a reference for dynamic search of the brightest point. UITA was built based on traditional computer vision. However, speckle noise is everywhere in an ultrasound image, which interferes with the stability while tracking diaphragm and compensating the respiratory induced motion. Deep learning methods offer a good object detection robustness which has a deep potential for tracking tumor motion under ultrasound imaging. In 2019, Huang *et al.* (22) combined fully convolutional neural network (FCNN) and convolutional long short-term memory (CLSTM) network to design an ultrasound-based tumor tracking technique. The results showed an average tracking error of 0.97±0.52 mm for 85 specific locations in 39 ultrasound imaging scenarios, with tracking rates ranging from 66 to 101 frames per second. For image guidance radiotherapy, this method is very effective in tracking tumor motion.

Mask R-CNN, a SOTA object detection framework introduced by He *et al.* in 2017 (23), stands out from traditional computer vision due to its ability to perform object classification, localization, and pixel-level instance segmentation. Consequently, this study aims to develop a novel ultrasound image tracking technology based on Mask R-CNN. The primary goal is to achieve superior stability in tracking the detected diaphragm motion. Furthermore, we evaluate its effectiveness by applying it to RMCS for compensation assessment.

Methods

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and approved by the Ethics Committee of Taipei Medical University Hospital (No. IRB 201902015), informed consent was taken from all the patients. To investigate the viability of implementing mask ultrasonic imaging tracking algorithm (M-UITA) in RMCS, we conducted CT image verification at Taipei Medical University Hospital. Two experiments were performed with diaphragm phantoms in this study. The first one is the verification experiment of tracking displacement accuracy of M-UITA. This experiment compares the correlation between the CBCT image and the tracking trajectory of M-UITA, to confirm whether the coordinate change of the tracked object conforms to the actual displacement; the second experiment is the respiration motion compensation experiment. The compensation rates tracked using UITA and M-UITA were compared to demonstrate the advantage of the proposed technology. *Figure 1* shows the workflow of M-UITA.

The specifications of the hardware equipment used in this work are as follows: the CBCT device model is Elekta Synergy System; the ultrasound imaging system is produced by Fukuda Denshi Company with the model UF-4000 and the ultrasonic probe model is FUT-C111A with a frequency of 3.5MHz. The CPU device model is I5-10400 and the GPU device model is NVIDIA GeForce GTX 1070ti. Python is used in this study for machine learning training and implementing M-UITA.

Experimental apparatus

The experimental setup for this study included the following apparatus: the RMCS, the respiratory motion simulation system (RMSS) (21), ultrasound equipment, and a diaphragm phantom. RMSS and diaphragm phantom were utilized to simulate human respiratory motion. The breathing waveform in this experiment was recorded using

6830

Ting et al. Mask R-CNN A for respiratory motion compensation system
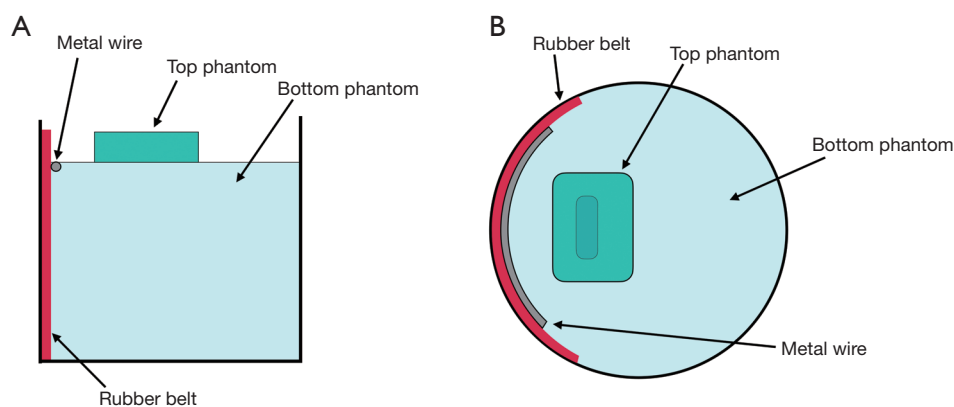
Figure 2 The diaphragm phantom. (A) Side view of the diaphragm phantom. (B) Top view of the diaphragm phantom.
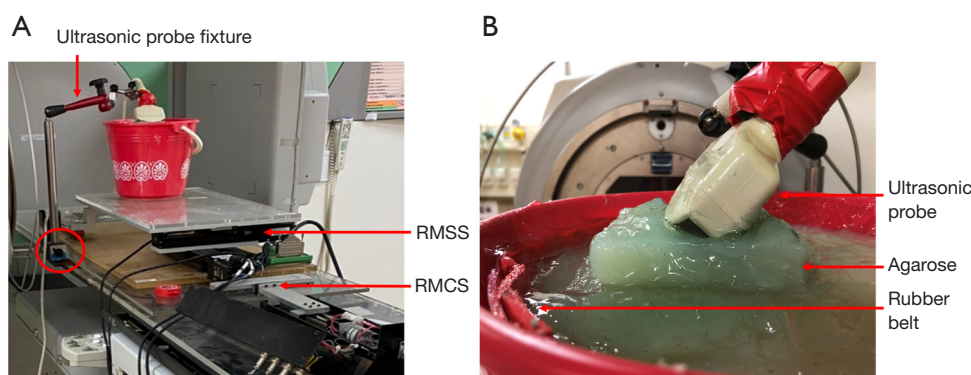

Figure 3 Experimental apparatus. (A) The fixture was set on a still surface. (B) The setting of the ultrasound probe. RMSS, respiratory motion simulation system; RMCS, respiratory motion compensation system.

ultrasound imaging of the real-time human diaphragm and capturing the diaphragmatic displacement signals with UITA. RMSS would simulate the movement of the diaphragm phantom based on the pre-recorded respiratory signals it receives.

The diaphragm phantom was designed based on the research (24) but with some improvements. The diaphragm phantom was composed of agarose, a rubber belt and a metal wire. Since the lung exists in the body, there will be a relative displacement between the surface of the human body and the diaphragm when it compresses and expands; while there is no relative displacement between the probe and the surface because the ultrasonic probe is attached on the body surface. To simulate this movement, the diaphragm phantom was designed in two layers: The top layer simulates the surface of the human body, and the bottom layer simulates the tissue in the body. *Figure 2* presents the design of the diaphragm phantom. The

ultrasonic probe was fixed by a fixture and set on a still surface where was no relative displacement with ground as shown in *Figure 3A*. When RMSS moved according to the respiratory signal, the bottom layer did the same movement; however, the top phantom wouldn't move since the probe was stuck in a notch on the upper phantom's surface. *Figure 3B* shows the setting of the ultrasound probe. A small piece of agarose on top simulates the abdominal skin, which remains stationary relative to the probe. This configuration was designed to closely replicate the actual movement of the diaphragm during human breathing. A rubber belt was affixed to the wall to simulate the diaphragm's motion, and ultrasonic waves would be reflected off the belt to generate an ultrasound image. To enhance the visibility of the rubber belt in the CT images, a metal wire was closely affixed to it, creating a distinct contour. The CT images were then analyzed using Tracker software. In each frame of the CT image series, we manually identified the feature points
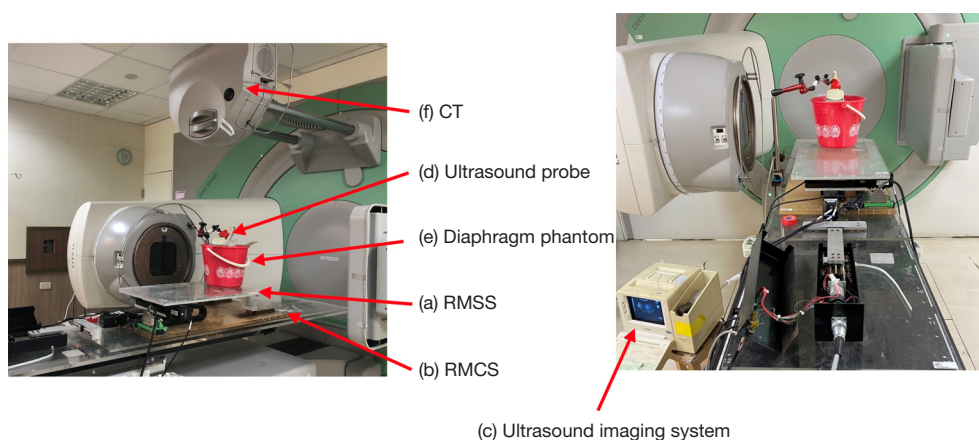
Figure 4 The overall experiment setup. CT, computed tomography; RMSS, respiratory motion simulation system; RMCS, respiratory motion compensation system.
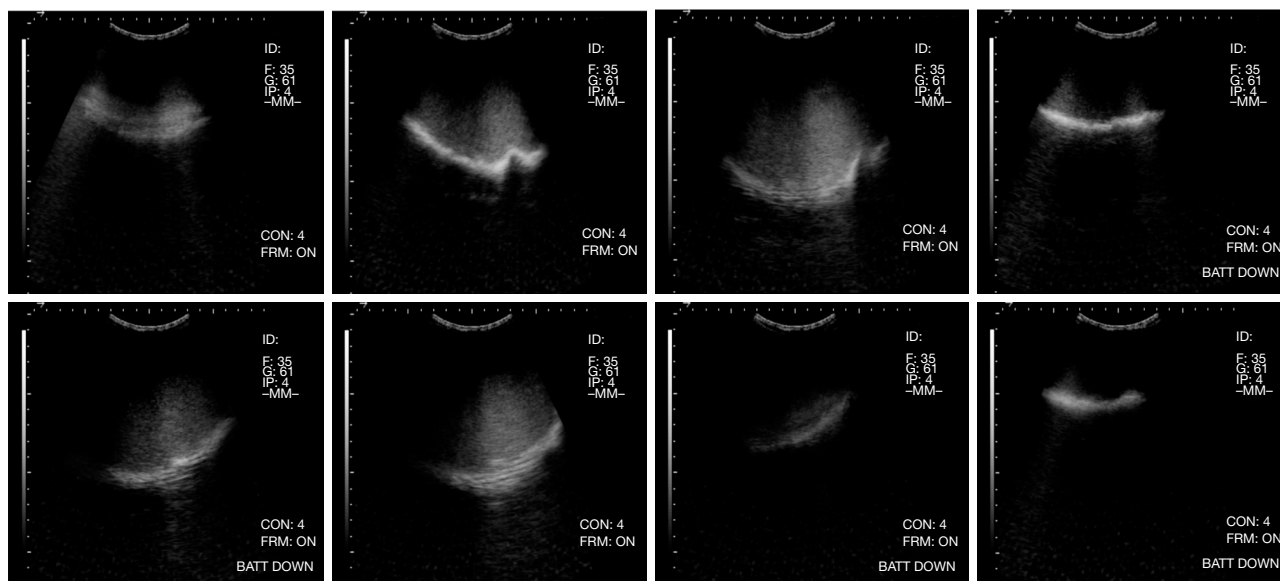
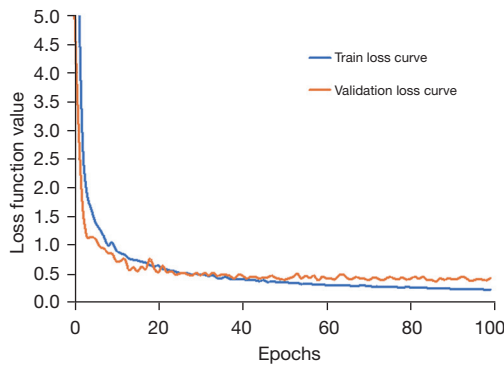**Figure 5** Examples of diaphragm images of the dataset.

of the diaphragm phantom and obtained its movement trajectories. *Figure 4* shows the complete experimental setup.

Dataset and model training

A dataset of 1,800 human diaphragm images was acquired using ultrasonic equipment. The images were 800×600 pixels in size. To increase the variety of the dataset for model training, deformed diaphragm shapes were intentionally recorded under severe breathing, with an amplitude of about ±20 mm. Examples of diaphragm images from the dataset are shown in *Figure 5*. The diaphragm region (indicated by the white line) was identified as the optimal detection point (feature point) for the entire system. The 1,800 images were divided into a training set (80%) and a validation set (20%) to prepare the dataset for training and validation. Data labeling was accomplished using VIA (VGG Image Annotator). Due to the limited number of samples in this study, a pre-trained model based on the COCO dataset (25) was utilized. The advantage of using a pre-trained model for transfer learning is that

even if the number of data samples is small, the parameters of the model can be adjusted to a good state based on the pre-trained model. In this study, anchor boxes of three scales were set according to the size of the diaphragm in ultrasound images: 64×64, 128×128, 256×256, and the aspect ratios of the anchor boxes were set to 1:1, 1:2 and 1:2, 2:1. Residual networks such as ResNet-34, ResNet-50, and ResNet-101 are often used in Mask R-CNN models as the backbone architecture for feature extraction. In this study, the pixel accuracy of diaphragm segmentation has a certain influence on the compensation effect of RMCS. According to the experimental results (26), the detection accuracy of ResNet-101 is the highest among the above three. Therefore, ResNet-101 was selected to be the

backbone of the Mask R-CNN model for training. The overall loss function of Mask R-CNN includes the loss of the RPN network and the loss of each branch: classification, bounding box regression and mask regression (23). The errors were calculated as following:

$$L_{overall} = L_{RPN} + L_{Branches} \qquad [1]$$

$$L_{RPN} = \frac{1}{N_{cls1}} \sum_i L_{cls}\left(P_i, P_i^*\right) + \lambda_1 \frac{1}{N_{reg1}} \sum_i P_i^* L_{reg}\left(t_i, t_i^*\right) \qquad [2]$$

$$L_{Branches} = \frac{1}{N_{cls2}} \sum_i L_{cls}\left(P_i, P_i^*\right)$$
$$+ \lambda_2 \frac{1}{N_{reg2}} \sum_i P_i^* L_{reg}\left(t_i, t_i^*\right) + \gamma_2 \frac{1}{N_{mask}} \sum_i L_{mask}\left(s_i, s_i^*\right) \qquad [3]$$

For training the model, the learning rate was set to 0.001, the batch size was set to 5, and the number of epochs was set to 100. *Figure 6* shows that the loss curves of the model all have a downward trend, and the errors of the training and validation datasets converge to 0.236 and 0.438.

M-UITA working process

M-UITA first detects the contour of the diaphragm in the ultrasonic image, and then obtains the coordinates of all pixels in the contour. The mean coordinate of all pixels will be calculated and converted into actual coordinates (mm) for RMCS to read. *Figure 7* displays the screenshots of

Figure 6 The loss curves of the model.
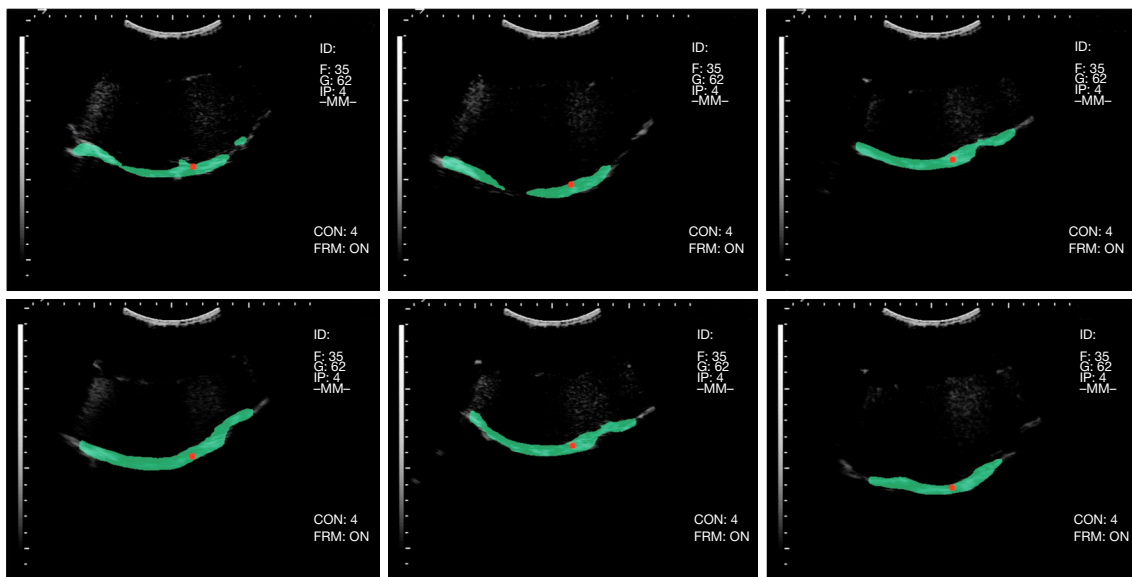
Figure 7 Screenshots of working M-UITA. M-UITA, mask ultrasonic imaging tracking algorithm.
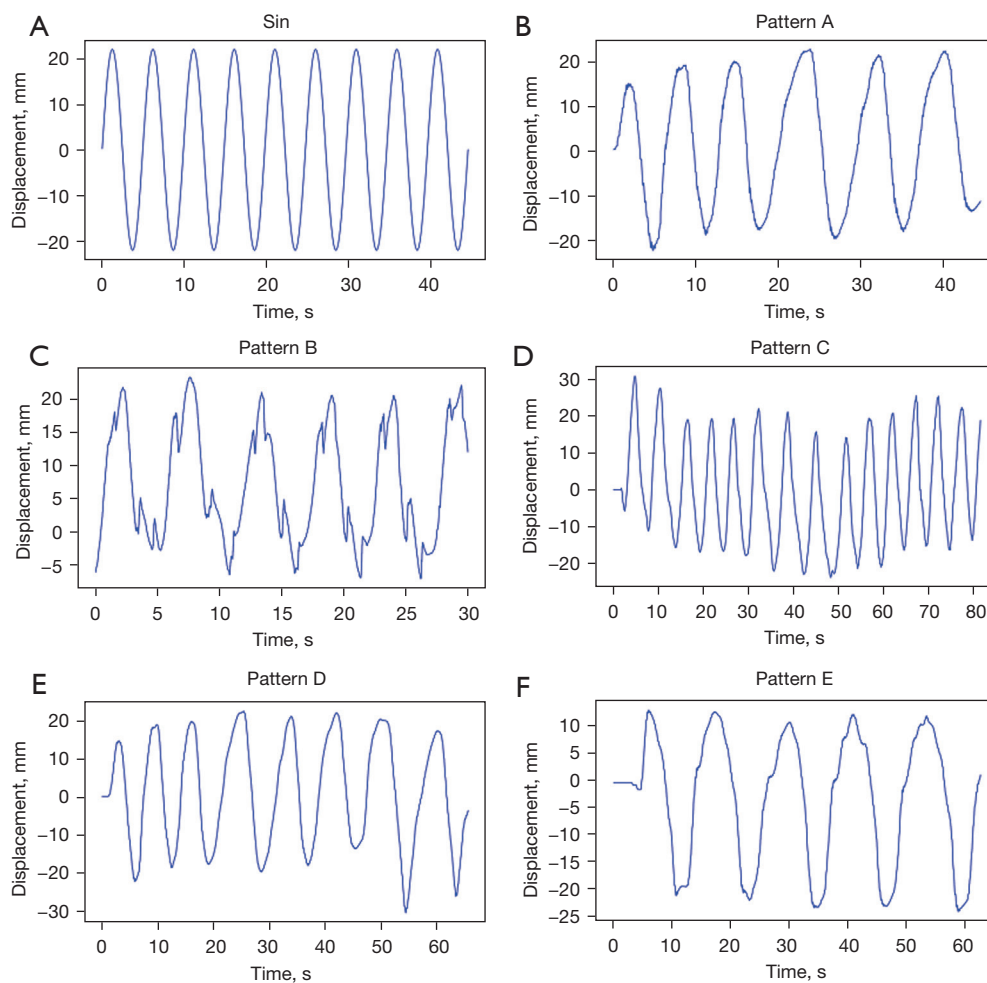
Figure 8 The pre-recorded respiratory signals. (A) Sin; (B) Pattern A; (C) Pattern B; (D) Pattern C; (E) Pattern D; (F) Pattern E.

M-UITA in action. The red dots in the figures represent the visualized mean coordinates calculated by M-UITA.

M-UITA tracking verification experiment

The purpose of this experiment is to verify whether the diaphragm tracking displacement of M-UITA is reliable. First, three pre-recorded respiratory waveforms were prepared in this experiment: sine wave, Pattern A and Pattern B. Pattern A, B were real human respiratory signals. The waveforms are shown in *Figure 8*. Diaphragm phantoms moved by based on the input respiratory waveforms to the RMSS and tracked with M-UITA. At the same time, CT was set at 90 degrees so that the high-energy rays hit the diaphragm phantom perpendicularly, which allowed us to record the pure movement of the

diaphragm phantom in S-I direction. The obtained CT image sequence were manually marked with the feature points to generate the displacement coordinates of the CT image, and the movement trajectories were used as the real values of this experiment. The detection frame rate of M-UITA is about 10 FPS, which is different from the frame rate of CT images (5.55 FPS). Therefore, it is impossible to directly calculate the point-to-point coordinate error under the same time resolution. In order to quantify the similarity between M-UITA's tracked trajectories and CT image trajectories, we calculated the discrete Fréchet distance (DFD) (27) between the two trajectories as an evaluation metric. The DFD is a measure of similarity or dissimilarity between two curves. It quantifies the minimum separation required for a continuous motion along both curves without backtracking.

6834

Ting et al. Mask R-CNN A for respiratory motion compensation system

Table 1 The result of DFDs

Pattern	Sin	A	B
DFD (mm)	3.12	3.85	3.71

DFDs, discrete Fréchet distances.

Respiration motion compensation experiment

After the verification of tracking similarity, M-UITA was applied to RMCS to conduct compensation experiments to observe the feasibility. In the experiment, three pre-recorded real human respiratory signals were prepared, Pattern C, D, and E. The respiratory signals are shown in *Figure 8*. The characteristic of Pattern C was high frequency with baseline shift; Pattern D was medium frequency with baseline shift and large fluctuations; Pattern E was low frequency and regular large fluctuations. These waveforms were input to RMSS, which made the diaphragm phantom move. M-UITA and UITA further tracked the phantom's movement, and transmitted the tracking signals to RMCS for movement compensation. In the end of the experiment, root mean square error (RMSE) and the compensation rate (CR) between the compensated residual signal and the original input signal were calculated. Each waveform was performed three times to calculate the standard deviation, and an interval of two standard deviations was used as the margin of error for the CR. RMSE and CR were calculated as following:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(RMSS_i - RMCS_i\right)^2}{n}} \qquad [4]$$

$$CR(\%) = \left(1 - \frac{RMSE_{compensated}}{RMSE_{uncompensated}}\right) \times 100\% \qquad [5]$$

Results

In the first experiment, DFD was calculated to quantify the similarity between the displacement trajectory tracked by M-UITA and the actual diaphragm displacement trajectory. Actual displacements are defined by the images obtained from CT scans. Accuracy is determined by the fact that the coordinates in the CT images are marked by a tracker with a precision of 1mm. Furthermore, M-UITA and UITA were applied for tracking the diaphragm movement with RMCS in the second experiment for evaluating the effectiveness of the proposed method.

Results of the verification experiment

The DFDs of sin, Pattern A and Pattern B were calculated to be 3.12 mm, 3.85 mm and 3.71 mm, respectively. The DFDs are presented in *Table 1*. In *Figure 9A*, the sine wave tracking signals of M-UITA almost overlapped with the actual displacement, while some defects and jitters existed at the turning point. *Figure 9B*, there was a slight translation between the M-UITA and the CT trajectories in Pattern A. It was because during the movement, the bottom phantom moved at a slow speed. When the bottom phantom overcame the friction with the top phantom, it caused inconsistent sliding. Pattern B was a waveform with many twists in a short period of time. Although there were also defects and jitters like sine wave and Pattern A at the turning point, the M-UITA and the CT trajectories were almost overlap as can be seen in *Figure 9C*.

Since the displacement of the diaphragm at the turning point of the waveform did not have a large movement like the intervals of peaks and valleys and M-UITA tracked continuously, even a slight change of detected area would cause a fluctuation of the mean coordinate, making the tracking trajectories not smooth. However, the calculated DFD of each waveform could prove that there was a high correlation between actual displacement and the displacement tracked by M-UITA. The DFD between M-UITA and the actual displacement trajectory was less than 4 mm, which is approximately equivalent to the tumor's diameter reference value, indicating that the similarity of two trajectories is very close. The radiation treatment typically sets a radiation field that includes an additional 1–2 cm margin around the tumor region.

Results of the compensation experiment

The average RMSE of the three waveforms tracked with M-UITA were: Pattern C: 7.00 mm, Pattern D: 8.20 mm, Pattern E: 6.35 mm; the average RMSE tracked with UITA were: Pattern C: 7.49 mm, Pattern D: 7.18 mm, Pattern E: 7.15 mm; while the CR of M-UITA was at most 6% higher than that of UITA. For the detailed experimental data, please refer to *Table 2* and *Table 3*. *Figure 9D* shows the CRs of the three waveforms, and the error bars were drawn with twice the standard deviation.

Deep learning-based tracking technology provide higher robustness in tracking specific objects in noisy scenes, such as ultrasound images. *Figure 10* illustrates the tracking trajectories of each input signal. In *Figure 10B*, it can be
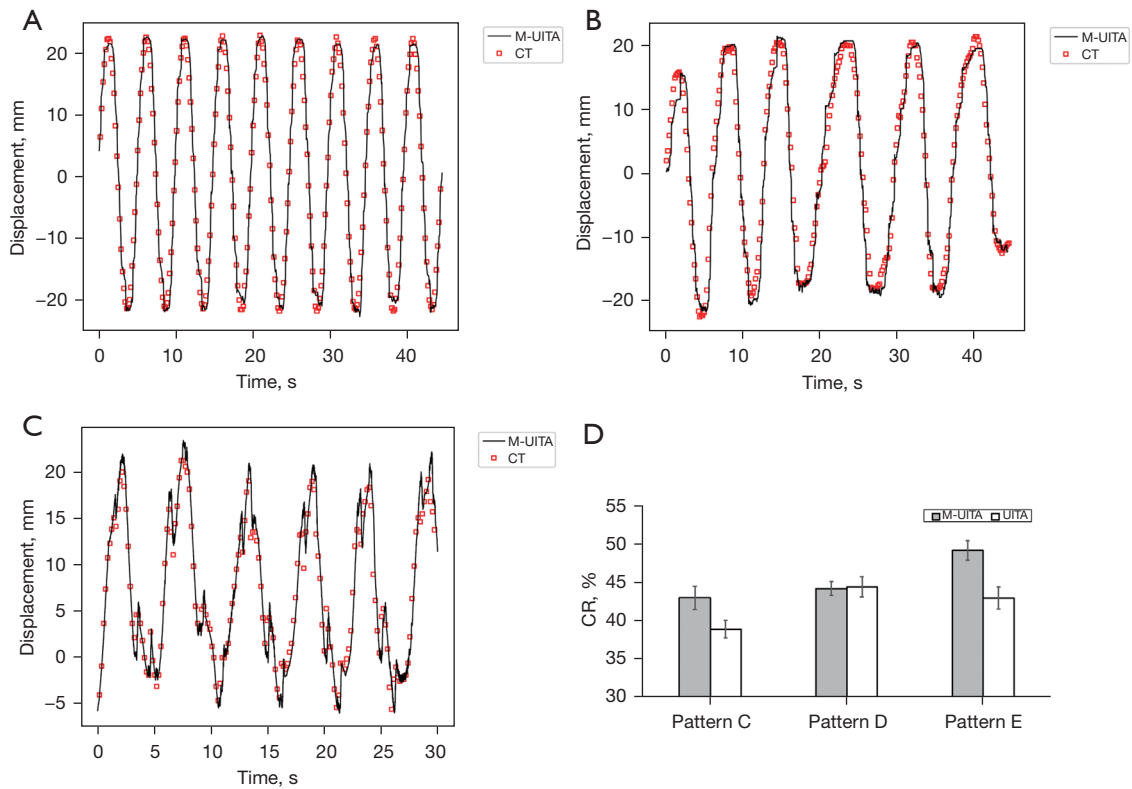
Figure 9 The tracking trajectory of (A) Sin, (B) Pattern A, (C) Pattern B and (D) CRs with error bar drawn with twice the standard deviation. M-UITA, mask ultrasonic imaging tracking algorithm; CT, computed tomography; CRs, compensation rates.

Table 2 The result of RMSEs

Pattern	M-UITA				UITA			
	#1	#2	#3	Average	#1	#2	#3	Average
C	6.89	6.99	7.12	7.00	7.59	7.48	7.41	7.49
D	8.26	8.11	8.24	8.20	7.31	7.01	7.21	7.18
E	6.35	6.44	6.25	6.35	7.20	7.19	7.07	7.15

RMSEs, root mean square errors; M-UITA, mask ultrasonic imaging tracking algorithm; UITA, ultrasonic imaging tracking algorithm.

Table 3 The result of CRs

Pattern	M-UITA				UITA			
	#1	#2	#3	Average	#1	#2	#3	Average
C	43.61	44.14	41.30	43.02	38.16	38.39	40.19	38.91
D	45.15	44.17	43.34	44.22	43.17	45.78	44.41	44.45
E	48.16	48.80	50.66	49.21	43.32	42.02	44.64	42.99

CRs, compensation rates; M-UITA, mask ultrasonic imaging tracking algorithm; UITA, ultrasonic imaging tracking algorithm.
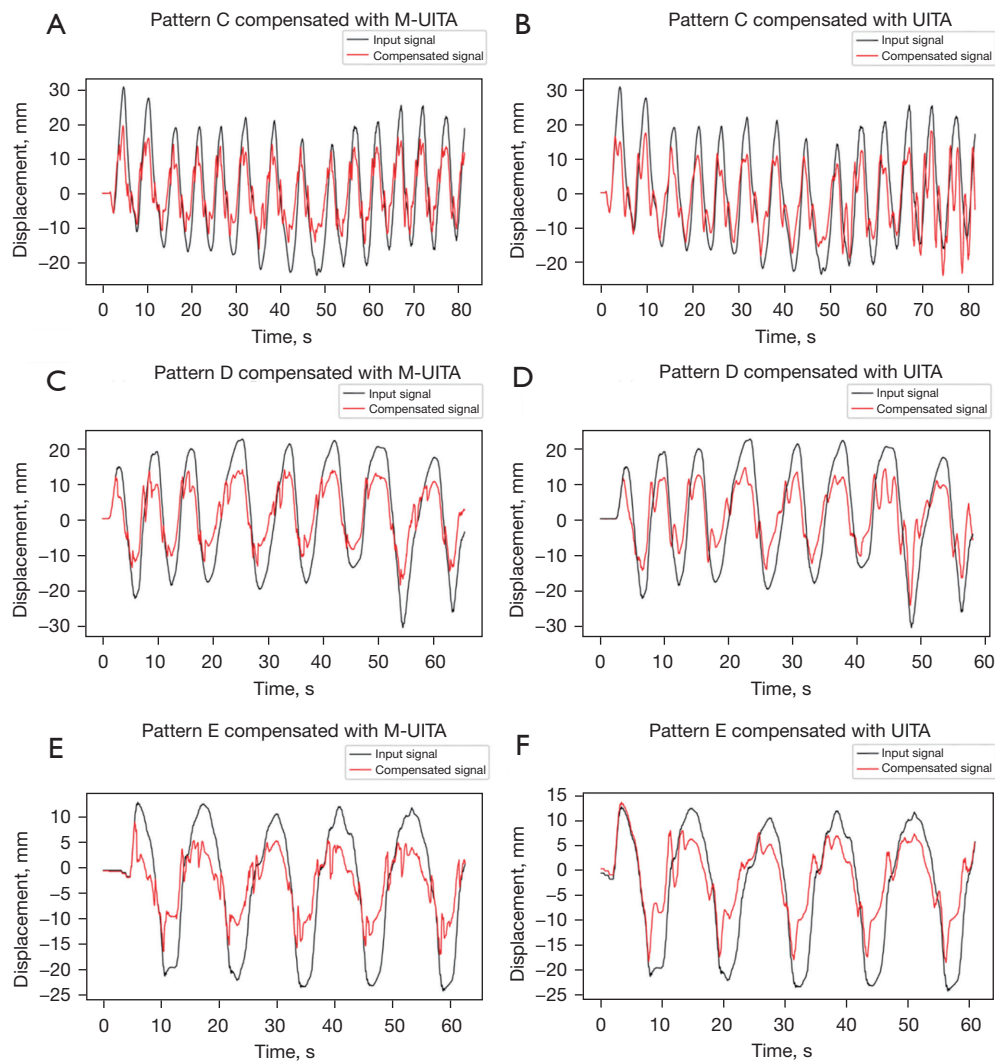
Figure 10 The tracking trajectories of: (A) Pattern C compensated with M-UITA; (B) Pattern C compensated with UITA; (C) Pattern D compensated with M-UITA; (D) Pattern D compensated with UITA; (E) Pattern E compensated with M-UITA; (F) Pattern E compensated with UITA. M-UITA, Mask Ultrasonic Imaging Tracking Algorithm; UITA, Ultrasonic Imaging Tracking Algorithm.

observed that in Pattern C, after 70 seconds, the tracking waveform of UITA exhibits a highly unstable state. This is likely due to UITA's failure to locate the feature point, resulting in an inaccurate tracking trajectory. This issue may be attributed to a slight gap between the ultrasonic probe and the upper diaphragm phantom during the compensation process, which caused the diaphragm contour to disappear. In contrast, M-UITA demonstrated a stable tracking process without encountering this problem. In Pattern D, the compensation rate using M-UITA is similar to that of UITA. As the respiratory frequency is lower, factors such as speckles and contact surface gaps in the ultrasound image

are reduced, which might not fully highlight M-UITA's advantages. However, at the turning points, the residual signal exhibits larger concaves (closer to zero), indicating that M-UITA can more accurately track the diaphragm's displacement during these critical moments.

Discussion

UITA, being a traditional computer vision method based on rules, requires manual parameter setting (e.g., binarization, erosion) before tracking. If the scene conditions do not match the settings, effective tracking may not be achieved.

This is because ensuring the purity of ultrasound images and the completeness of diaphragm shape during detection for different patients can be challenging. Moreover, during treatment, the continuous movement of RMCS leads to changes in the angle of the ultrasonic probe and its contact with the body surface, deviating from the initial position. These factors introduce uncertainties in diaphragm tracking with UITA, making the tracking process more challenging. On the other hand, M-UITA overcomes these problems. It does not require manual parameter adjustments, and it exhibits a higher tolerance for uncertainty in the image. Unlike UITA, which tracks a single feature point, M-UITA tracks the mean pixels of the diaphragm in ultrasound images. As long as the diaphragm's outline remains present, even if incomplete or distorted, M-UITA can still detect the diaphragm due to the robust generalization of the model. Tracking the mean pixel coordinates of the diaphragm area represents the overall displacement of the diaphragm more effectively than tracking a single specific point.

In fact, the ultrasound image of the human abdomen is full of noise composed of fat and other soft tissues, and the diaphragm also contains distortions when a person breathes. The diaphragm phantom used in this study is composed of double layers of agarose, a rubber belt and only translates in the SI direction. This simulation is still too ideal when compared to the real human ultrasonic images. Since M-UITA is not a rule-based traditional computer vision method, it learns the features from various samples through a deep neural network. M-UITA requires a sufficiently complex scene to highlight its excellent tracking ability. In addition, M-UITA takes about 0.35 seconds to generate the average coordinates of the detected diaphragm. The main reasons for this delay can be divided into two parts: (I) the detection of diaphragm; (II) the calculation of the mean coordinates. The former involves the multi-node parallel operation of the neural network, and the upgrade of the GPU will have the opportunity to reduce the delay; while the architecture of the ultrasonic image tracking algorithm is not a multi-node architecture, and relies more on the capability of the CPU, and the algorithm of calculation still needs to be optimized to accelerate.

In 2019, Price *et al.* (28) used a reference implant to track displacement and achieved a compensation effect as high as 98.9%. Although our experiment only achieved a compensation effect of 49.21%. However, we used a non-invasive and indirect positioning with the proposed Mask R-CNN model, to develop ultrasonic image tracking technology (M-UITA) for the compensation of the respiratory displacement. It is believed that future technical improvements to M-UITA will significantly improve the compensation effect.

Conclusions

This study developed an ultrasound image tracking technology called M-UITA, based on Mask R-CNN, and applied it to RMCS. A dataset of 1,800 real diaphragm ultrasonic images was obtained using ultrasonic equipment for model training. M-UITA calculates the mean pixel coordinates of the diaphragm segmented by Mask R-CNN and then transmits this data to RMCS. To assess the feasibility of M-UITA, tracking verification and compensation experiments were conducted. In the tracking verification, DFDs between the tracking trajectory of M-UITA and the real displacement trajectory was calculated. The calculated average DFD was less than 4 mm, which confirmed that the tracking trajectory of the proposed method had a high correlation with the actual displacement. The respiratory motion compensation experiment was conducted to assess the effectiveness of the proposed method. The compensation rate was calculated for both M-UITA and UITA, and the results showed that M-UITA had a compensation rate up to 6.22% higher than UITA. This suggests that M-UITA is more effective at compensating for respiratory motion. Additionally, M-UITA does not require additional manual parameter adjustments, which significantly reduces operational complexity.

Acknowledgments

Hospital under Contract USTP-NTUT-TMU-110-09.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
2. Kushi LH, Byers T, Doyle C, Bandera EV, McCullough M, McTiernan A, Gansler T, Andrews KS, Thun MJ; . American Cancer Society Guidelines on Nutrition and Physical Activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity. CA Cancer J Clin 2006;56:254-81; quiz 313-4.
3. Goitein M. Organ and tumor motion: an overview. Semin Radiat Oncol 2004;14:2-9.
4. von Siebenthal M, Székely G, Gamper U, Boesiger P, Lomax A, Cattin P. 4D MR imaging of respiratory organ motion and its variability. Phys Med Biol 2007;52:1547-64.
5. Pham D, Kron T, Foroudi F, Schneider M, Siva S. A

review of kidney motion under free, deep and forced-shallow breathing conditions: implications for stereotactic ablative body radiotherapy treatment. Technol Cancer Res Treat 2014;13:315-23.
6. Dhont J, Vandemeulebroucke J, Burghelea M, Poels K, Depuydt T, Van Den Begin R, Jaudet C, Collen C, Engels B, Reynders T, Boussaer M, Gevaert T, De Ridder M, Verellen D. The long- and short-term variability of breathing induced tumor motion in lung and liver over the course of a radiotherapy treatment. Radiother Oncol 2018;126:339-46.
7. Cerviño LI, Chao AK, Sandhu A, Jiang SB. The diaphragm as an anatomic surrogate for lung tumor motion. Phys Med Biol 2009;54:3529-41.
8. Yang J, Cai J, Wang H, Chang Z, Czito BG, Bashir MR, Palta M, Yin FF. Is diaphragm motion a good surrogate for liver tumor motion? Int J Radiat Oncol Biol Phys 2014;90:952-8.
9. Li GQ, Yang J, Wang Y, Qiu M, Ding Z, Zhang S, Yang SL, Peng Z. Using the Diaphragm as a Tracking Surrogate in CyberKnife Synchrony Treatment. Med Sci Monit 2021;27:e930139.
10. De Los Santos J, Popple R, Agazaryan N, Bayouth JE, Bissonnette JP, Bucci MK, Dieterich S, Dong L, Forster KM, Indelicato D, Langen K, Lehmann J, Mayr N, Parsai I, Salter W, Tomblyn M, Yuh WT, Chetty IJ. Image guided radiation therapy (IGRT) technologies for radiation therapy localization and delivery. Int J Radiat Oncol Biol Phys 2013;87:33-45.
11. Dang A, Kupelian PA, Cao M, Agazaryan N, Kishan AU. Image-guided radiotherapy for prostate cancer. Transl Androl Urol 2018;7:308-20.
12. Ren XC, Liu YE, Li J, Lin Q. Progress in image-guided radiotherapy for the treatment of non-small cell lung cancer. World J Radiol 2019;11:46-54.
13. Oktay O, Nanavati J, Schwaighofer A, Carter D, Bristow M, Tanno R, Jena R, Barnett G, Noble D, Rimmer Y, Glocker B, O'Hara K, Bishop C, Alvarez-Valle J, Nori A. Evaluation of Deep Learning to Augment Image-Guided Radiotherapy for Head and Neck and Prostate Cancers. JAMA Netw Open 2020;3:e2027426.
14. Zou W, Dong L, Kevin Teo BK. Current State of Image Guidance in Radiation Oncology: Implications for PTV Margin Expansion and Adaptive Therapy. Semin Radiat Oncol 2018;28:238-47.
15. Ibbott GS. Patient doses from image-guided radiation therapy. Phys Med 2020;72:30-1.
16. Demehri S, Muhit A, Zbijewski W, Stayman JW, Yorkston

J, Packard N, Senn R, Yang D, Foos D, Thawait GK, Fayad LM, Chhabra A, Carrino JA, Siewerdsen JH. Assessment of image quality in soft tissue and bone visualization tasks for a dedicated extremity cone-beam CT system. Eur Radiol 2015;25:1742-51.

17. Schwenke M, Strehlow J, Demedts D, Haase S, Barrios Romero D, Rothlübbers S, von Dresky C, Zidowitz S, Georgii J, Mihcin S, Bezzi M, Tanner C, Sat G, Levy Y, Jenne J, Günther M, Melzer A, Preusser T. A focused ultrasound treatment system for moving targets (part I): generic system design and in-silico first-stage evaluation. J Ther Ultrasound 2017;5:20.

18. Mihcin S, Melzer A. Principles of focused ultrasound. Minim Invasive Ther Allied Technol 2018;27:41-50.

19. Fortunato GM, Bonatti AF, Batoni E, Macaluso R, Vozzi G, De Maria C. Motion compensation system for robotic based in situ bioprinting to balance patient physiological movements. Bioprinting 2022;28:e00248.

20. Kuo CC, Chuang HC, Teng KT, Hsu HY, Tien DC, Wu CJ, Jeng SC, Chiou JF. An autotuning respiration compensation system based on ultrasound image tracking. J Xray Sci Technol 2016;24:875-92.

21. Chuang HC, Huang DY, Tien DC, Wu RH, Hsu CH. A respiratory compensating system: design and performance evaluation. J Appl Clin Med Phys 2014;15:4710.

22. Huang P, Yu G, Lu H, Liu D, Xing L, Yin Y, Kovalchuk N, Xing L, Li D. Attention-aware fully convolutional neural network with convolutional long short-term memory network for ultrasound-based motion tracking. Med Phys 2019;46:2275-85.

23. He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision. 2017.

24. Ting LL, Chuang HC, Kuo CC, Jian LA, Huang MY, Liao AH, Tien DC, Jeng SC, Chiou JF. Tracking and compensation of respiration pattern by an automatic compensation system. Med Phys 2017;44:2077-95.

25. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft coco: Common objects in context. in Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. 2014. Springer.

26. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

27. Eiter T, Mannila H. Computing discrete Fréchet distance. 1994. Available online: http://www.kr.tuwien.ac.at/staff/eiter/et-archive/cdtr9464.pdf

28. Price A, Chen J, Chao E, Schnarr E, Schreiber E, Lu L, Cox A, Chang S, Lian J. Compensation of intrafractional motion for lung stereotactic body radiotherapy (SBRT) on helical TomoTherapy. Biomedical Physics & Engineering Express 2019;5:025043.