Research Article

A Study of Applying Pix2Pix to Reduce the Cost and Damage of Radiotherapy of Cancerous Livers

Xiaochuan Liang

Horace Greeley High School, 70 Roaring Brook Rd, Chappaqua, NY 10514, USA

ABSTRACT

Radiation therapy is widely used to treat cancerous cells; however, it is costly and has significant side effects if not properly used. During treatment, some organs, such as the liver, may shift slightly due to the patient's respiration, making accurate real-time tracking and contouring essential to minimize damage to healthy tissues. Current tracking systems usually include precise real-time positioning and contouring of the treatment area, which require costly imaging technologies to run throughout the entire therapy session, increasing the overall treatment cost and limiting the availability of advanced equipment like MRIs to hospitals. This research studies the application of Pix2Pix, a Conditional Generative Adversarial Network (GAN), as an alternative to traditional contouring methods in liver cancer radiotherapy. The result shows that a well-trained Pix2Pix can help accurately track liver movement during respiration. With AI computing power becoming increasingly affordable, I anticipate that GANs like Pix2Pix can be industrialized to make cancer treatment more accessible and cost-effective.

Keywords: Cancel treatment, cost reduction, affordable radiotherapy, contouring methods, conditional GAN, Pix2Pix

INTRODUCTION

Over 50% of cancer patients undergo radiation therapy as part of their treatment (1). This method uses high-energy radiation to kill cancer cells, with each session lasting from 15 to 30 minutes, though the actual radiation exposure only occurs for a few minutes. One of the main challenges in radiation therapy is to deliver an adequate dose to the tumor while minimizing exposure to surrounding healthy tissues. This challenge becomes more difficult when the target area moves or changes shape, such as in the liver or lungs during respiration.

Respiratory motion is a major source of uncertainty in radiotherapy for abdominal tumors, leading to deviations in dose distribution and potential damage to healthy tissues. To address this issue, various techniques have been developed in the field of radiotherapy. First, expanding the treatment area by increasing the planning margins can ensure the moving target is consistently treated. However, this approach also increases the risk of damaging adjacent healthy tissues, particularly for patients with

Corresponding author: Xiaochuan Liang, E-mail: stsliang@outlook.com.

Received August 29, 2024**; Accepted** October 8, 2024
https://doi.org/10.70251/HYJR2348.23123131

larger motion ranges of up to 2-3 cm. Second, motion-management techniques, such as respiratory gating only deliver radiation during specific points of the breathing cycle, while abdominal compression uses a belt to reduce tumor movement (2). Third, fiducial markers which are tiny metal markers implanted in the treatment area, assist in tracking the tumor movement. Unfortunately, this method is incredibly invasive, and the markers may migrate away from targets and the implantation itself might lead to infections or other side effects (3). Lastly, real-time motion synchronization technology tracks the motion of the treatment area and guides the radiation beam accordingly to maintain real-time precise targeting throughout the session (4).

The real-time motion synchronization technology in radiotherapy has the potential to significantly improve the accuracy of radiotherapy by reducing treatment times and minimizing damage to healthy tissues. This technology utilizes imaging methods like CT, MRI, and PET, with fast-computing software for real-time tracking and contouring of moving treatment areas. However, continuously operating such systems is costly, making them inaccessible to some patients and hospitals that lack imaging machines. To address this issue, the radiotherapy field has been actively exploring alternative methods. A recent review highlights surface-guided radiation therapy (SGRT), which uses non-invasive optical surface imaging (OSI) to monitor patient body motion in real-time (5, 6). When combined with gating technology, SGRT is particularly effective for patients with noticeable, repetitive body movements, such as during respiration. However, this method is subject to interruptions caused by respiratory irregularities. Currently, radiation therapy is a complex process involving advanced imaging, computing, and medical expertise with simulation imaging via 3D CT or MRI to map the patient's anatomy. Oncologists will then contour the treatment area to differentiate it from healthy organs, followed by planning and implementation of the treatment (7).

With the recent clinical adoption of MR-guided linear accelerators, a large amount of continuous real-time planar MR imaging data has been captured at a rate of 4 frames per second during the radiation therapy treatment. This allows researchers to investigate the correlation between internal organ motions and external skin movements. The advanced artificial intelligence will enable us to analyze large amounts of data to generate motion models. Recently, general adversarial networks (GANs) have been developed and tested in medical imaging, showing great success, particularly in medical image classification, segmentation, and image synthesis including CT translation (8-10). Based on these successes, this research proposes a novel approach to radiation therapy by replacing costly imaging methods like CT, MRI, and PET with more accessible technologies such as cameras and ultrasound. Using body contours and diaphragm movements as inputs for a trained AI model, the goal is to accurately predict the position and boundaries of treatment areas, such as cancerous livers, allowing for precise targeting of cancer cells while minimizing damage to healthy tissue. The initial focus will be on liver cancer treatment to minimize damage and side effects of radiation therapy.

METHODS

In this article, I won't recount deeply the specifics of the machine learning model; instead, I will focus on how I prepared the inputs and desired outputs for this conditional GAN, how I adapted the model, and how I measured the success of the model.

Model of GAN

This research has explored various GANs including cycle GANs which are trained on unpaired images and conditional GANs which are trained on paired images. A comparative analysis between these two types of GANs is available in (11). I have found that Pix2Pix, a specific implementation of conditional GAN, is effective in the prediction of the centers and contours of the treatment areas based on simple and relatively affordable input data.

Data Processing

The preparation of datasets to train and test our Pix2Pix model is outlined in Figure 1 where each image is labeled with a corresponding letter.

We first extracted the images (frames) from video clips generated by real-time motion tracking and contouring systems in the radiotherapy practice of various cancerous livers. Our research yielded approximately 250 images per minute from the video clips with most videos yielding between 2000-4000 images, and the lowest producing around 1000 images. This process resulted in a few thousand images like the image shown in Figure 1A which depicts a 2-dimensional image of the thorax and abdominal area with the contour of the cancerous area for radiotherapy treatment. Figure 1B was derived from Figure 1A, representing the body contour and diaphragm, which serves as the "input" to the conditional GAN model Pix2Pix, and Figure 1C was also extracted from Figure 1A, illustrating the body contour alongside the treatment area

contour. From Figure 1C, I can determine the center of the treatment area, which is crucial for accurately targeting the radiation beam and the contour of the treatment area for guiding the radiation beam. However, using Figure 1C as the desired "target" or "output" of Pix2Pix causes a convergence issue. To address this issue, I found a filled image (Figure 1D) can greatly improve the convergence speed and accuracy of Pix2Pix training. Hence, Figure 1D was used as the target to train the Pix2Pix model.

Ultimately the training dataset and test dataset were generated, which consisted of a few thousand images with each containing a pair of a target image and an input image as illustrated in Figure 2.

Figure 3 demonstrates how the treatment areas, whether the entire liver or a portion of it, can shift, with green grid lines overlaid on the original images to visually illustrate the motion of the treatment area in each patient case.

Model Implementation and Loss Function

A few implementations of the Pix2Pix model as detailed in Isola's paper are available on the internet, utilizing either the TensorFlow framework or the PyTorch framework (12). I started the exploration with the Colab framework due to its free, though, limited TPU computing power. This led us to explore and adapt the Pix2Pix implementation published on the TensorFlow website before we moved it the calculations to a local GPU (13). In a GAN model, the **loss function** is a critical component. In the original Pix2Pix implementation, the generator loss is a weighted combination of the sigmoid cross-entropy loss of the model-predicted images compared to an array of ones, and the L1 Loss which is a mean absolute error (MAE) between the generated images and the target images (13). The discriminator loss is calculated as the sum of real loss, a sigmoid cross-entropy loss between the real photos and an array of ones, and generated loss which is the sigmoid cross-entropy loss of the generated images and an array of zeros, representing the fake images.

Figure 2. A representative image in the dataset consists of a target image (Left) and an input image (Right).

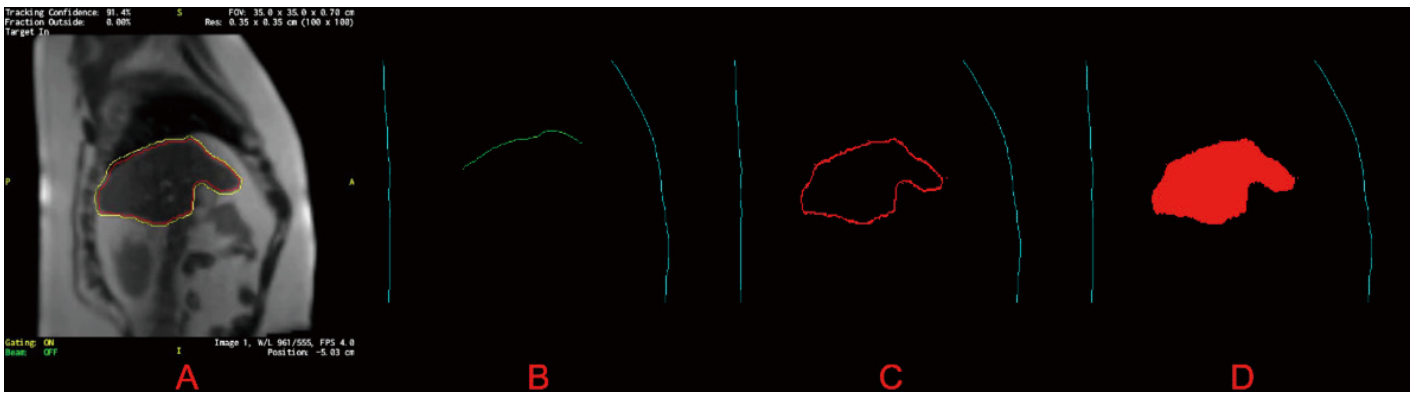

Figure 1. Extracting datasets for Pix2Pix training and testing, and A, B, C, and D images are of size 256 × 256 pixels.

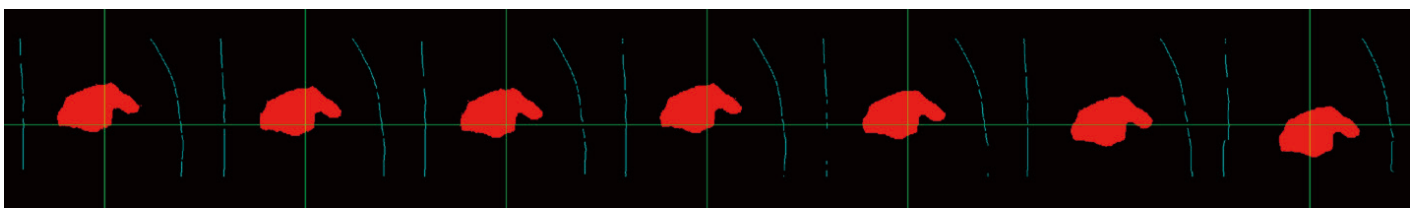

Figure 3. Sample positions and shapes of cancerous livers or parts of livers of different patients.

Measures of the Model Result

Various metrics have been developed to assess the accuracy of generated images in comparison to target images. I found that most of the regular image similarity measures as reported by Wang et al are not suited to our purpose (8). Our primary focus is on accurately identifying the center of the treatment area, which determines where the radiation beam should target, and the contour of the treatment area, which differentiates the cancerous cells from the healthy tissue.

In this research, I use these two measures:

DC: the dice coefficient, defined as (14)

$$DC = \frac{2 \mid T \cap P \mid}{\mid T \mid + \mid P \mid}$$

Where T is the set of pixels of the treatment area in the target image, while P is the set of pixels of the predicted treatment area in the model-generated image, I also use 2 variations of DC to achieve better control.

$$DC1 = \frac{\mid T \cap P \mid}{\mid T \mid}, \quad DC2 = \frac{\mid T \cap P \mid}{\mid P \mid}$$

The value of 1.0 - DC1 roughly tells us how much the target treatment area is missed, and 1.0 - DC2 roughly tells us how much of the predicted treatment area is the healthy body tissue.

CD: The Center Distance between the center of the treatment area in the target image and the center of predicted treatment is in the model-generated image. The center of an area is defined as the simple average of the coordinates of the pixels in the area.

This measure roughly tells us how far away the predicted target for the radiation beam is from the desired target point in the treatment area.

Model Training and the Exit Point

The original implementation does not implement the exit mechanism during the model training, and a notorious issue with GANs is that the training process does not always converge to the optimal state and can sometimes deviate, requiring manual checks of the training log to identify the checkpoints where the model is in the relatively best-trained state. To address this, I implemented an exit mechanism using the DC and CD measures. First, the thresholds to exit for these measures are defined for each training. For each training session, specific thresholds are set for these metrics. Training continues until a certain epoch limit is reached, typically 200,000 epochs, beyond which further training would be unproductive. During training, at every 100 epochs or 200 epochs, the trained model was tested with 12 sample points from the test dataset, with each sample as a pair of the target image and the input image. With the input, target, and model-predicted images, I can calculate the DC and CD for each sample point, and calculate the average DC and CD. I then check if they beat the threshold that I have defined for the model to exit.

Usually, the thresholds for DC, DC1, and DC2 for each sample point were set at 0.9 or 90%, and 91% or higher for the threshold for the average DC of all sample points. The thresholds for CD are mostly 2-3 pixels for each sample point with average CD across all sample points needing to be 0.5 pixels lower for the model training, the testing of the trained model was demonstrated in Figure 4, which shows a single input comprising the input image, the target image, and the conditional GAN-generated image. The fourth image in the block is the "merged" version of the target image and the model-predicted image. In this merged image, the yellow area represents the "T∩P" area, the red area represents missed cancerous cells, and the green area highlights the healthy cells mistakenly identified as cancerous by the model. The numbers on top of Figure 4 represent DC, DC1, DC2, and CD respectively.

Testing of the Trained Model

After the model exits from training, it is then run against the test dataset, with DC and CD calculated for each test data point. The testing result can be demonstrated in Figure 7 to Figure 8 at the end of this paper, with each figure representing a different data set. As a comparison reference for CD, the dimension of the treatment area of this patient is about 80 pixels. The result will be discussed in detail in the following section.

RESULTS AND DISCUSSION

We have collected more than 100 video clips from anonymous patients with various abdominal cancers, among which around 50 were cancerous liver cases. I developed the image extraction and dataset construction

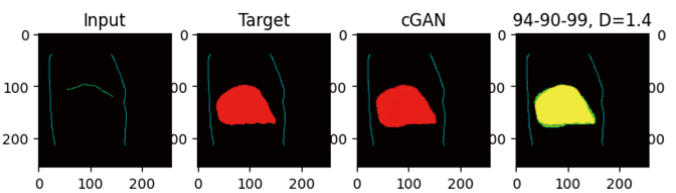

Figure 4. Intra-training checkpoint testing of the trained model in search of the exit point for a well-trained model.

process and generated the training and testing dataset for each video clip. However, given the limitations of the GPU resource I had for this research, I decided to take a few typical cases in the current phase: I divide these 50 sets of data for cancerous liver into 7 categories based on the shape of the treatment area as demonstrated by Figure 3.

In this phase of the research, our exploration, training, and fine-tuning of the Pix2Pix model were performed on one or two datasets from each of these categories. The result was presented in 3 formats with statistics of the result shown in Table 1, distribution charts of DC and CD were shown in Figure 5-6 showing the change in accuracy

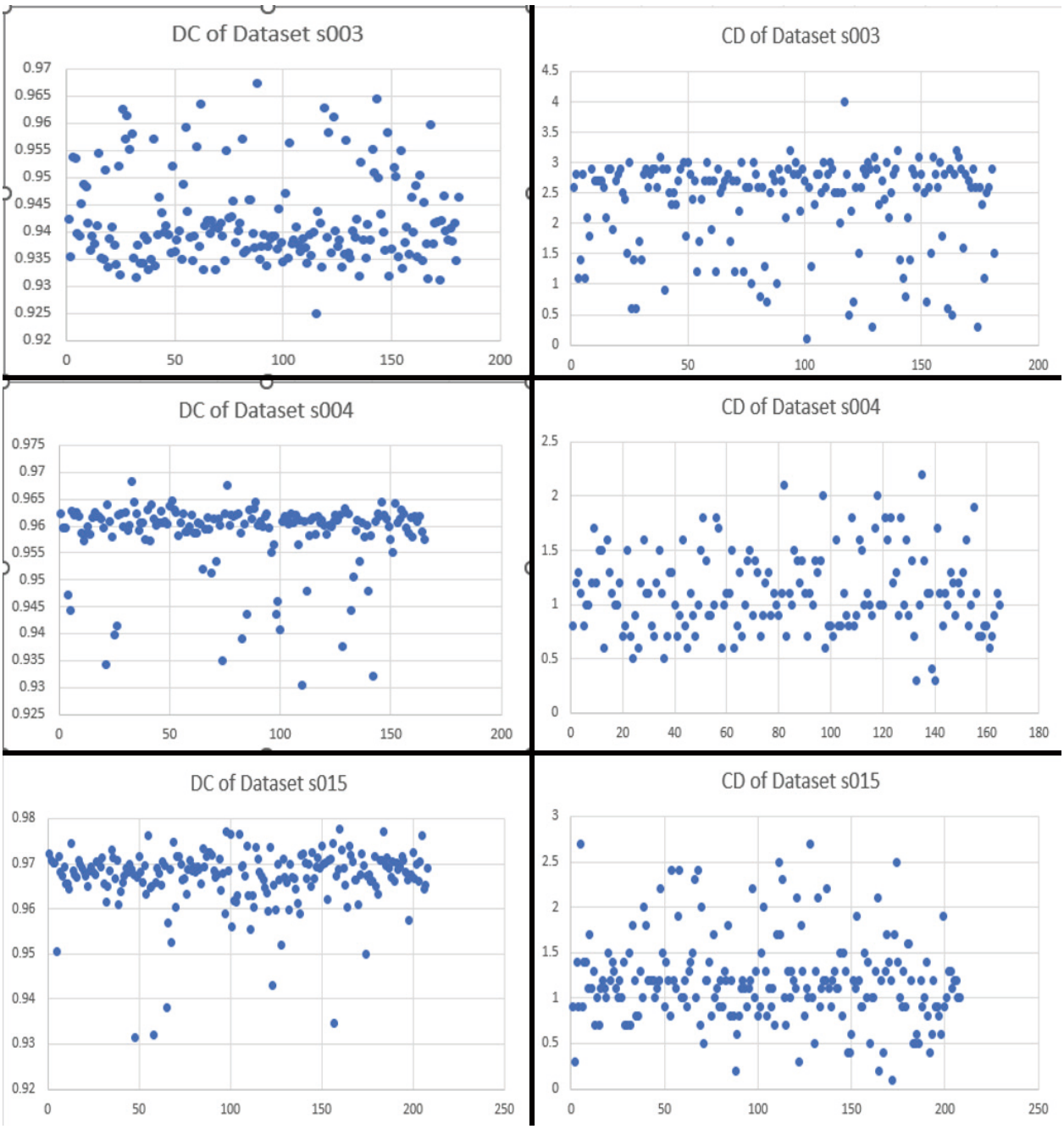

Figure 5. Distribution charts of DC and CD for datasets s003, s004, and s015, x-axis depicts the test result of the xth test image, y-axis depicts the correlating DC or CD value.

of the generated results in the form of DC and CD in relation to the number of epochs, and sample test cases as shown in Figure 7-8 demonstrate the actual accuracy of the generated model.

In Table 1, the Avg DC and StdDev DC are the mean and standard deviation of the Dice Coefficient for all test cases in the corresponding dataset. Similar terms were also given for DC1, DC2, and CD. The Avg Size value is computed as the mean of the square root of the number of pixels in the treatment area in each test case of the corresponding dataset, and this Avg Size value can be compared with the Avg CD value to decide how accurate

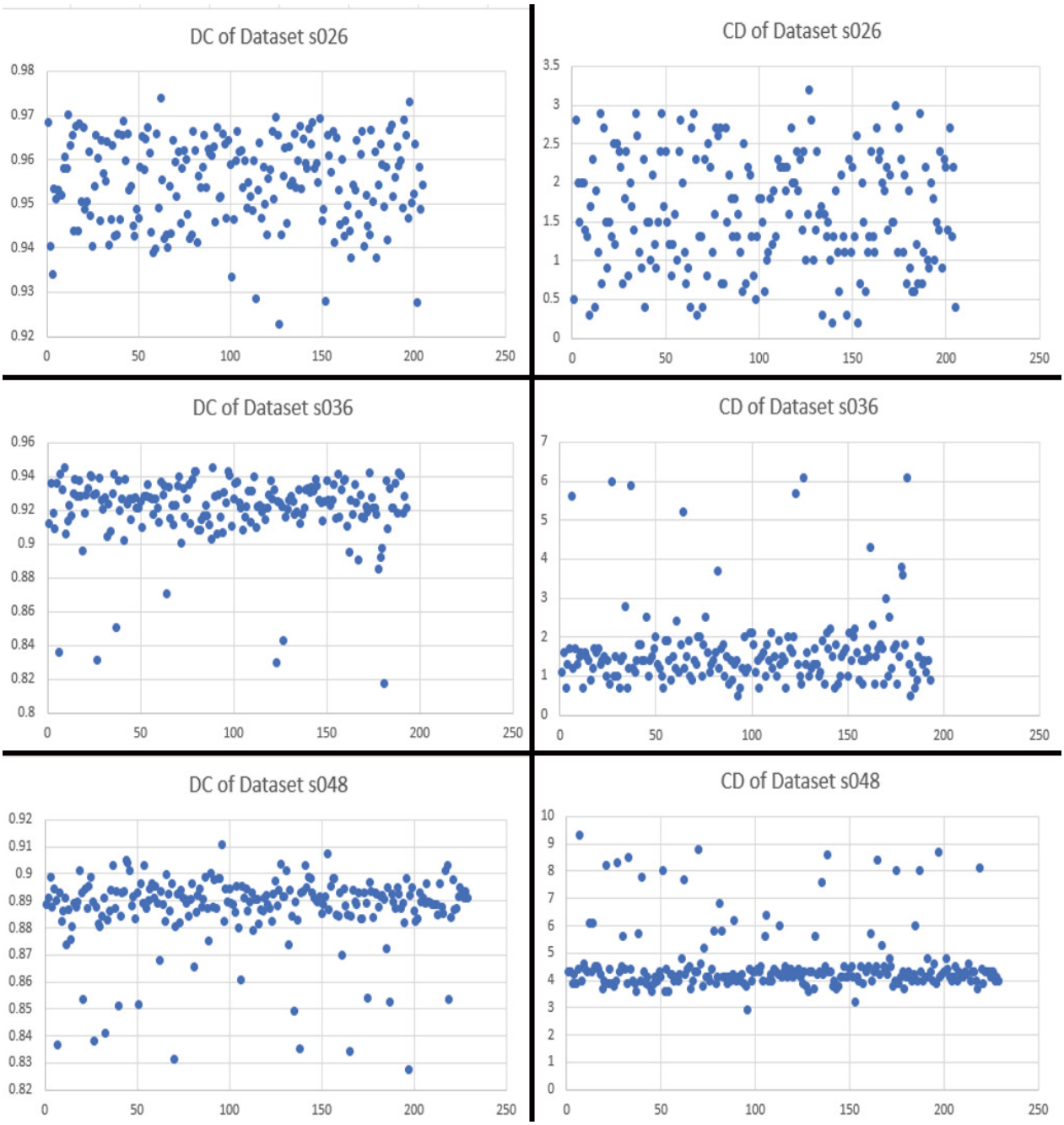

Figure 6. Distribution charts of DC and CD for datasets s026, s036, and s048, x-axis depicts the test result of the xth test image, y-axis depicts the correlating DC or CD value.

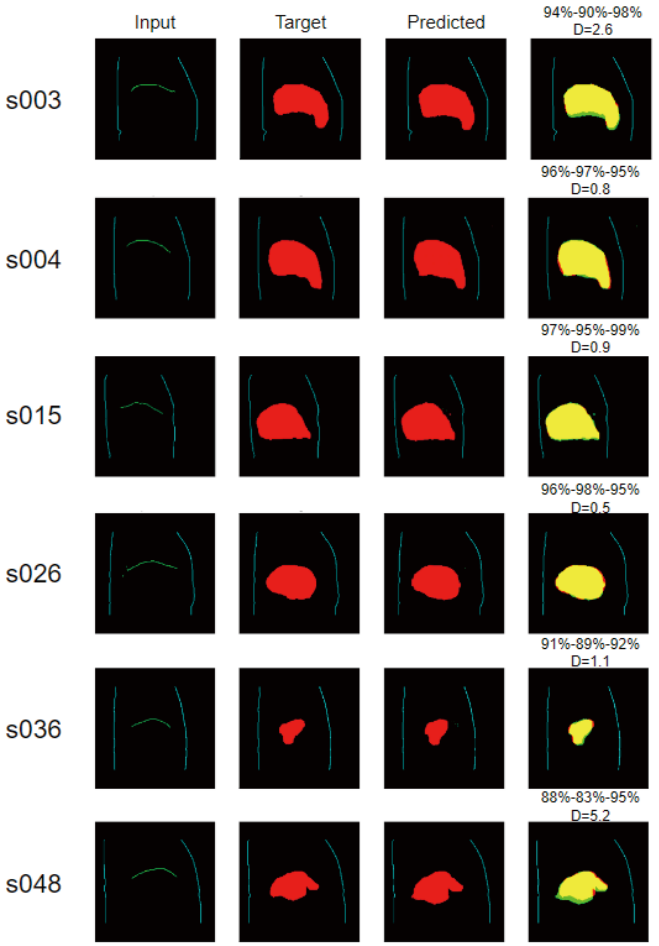

Figure 7. Sample testing result of the trained model on datasets s003, s004, s015, s026, s036, and s048.

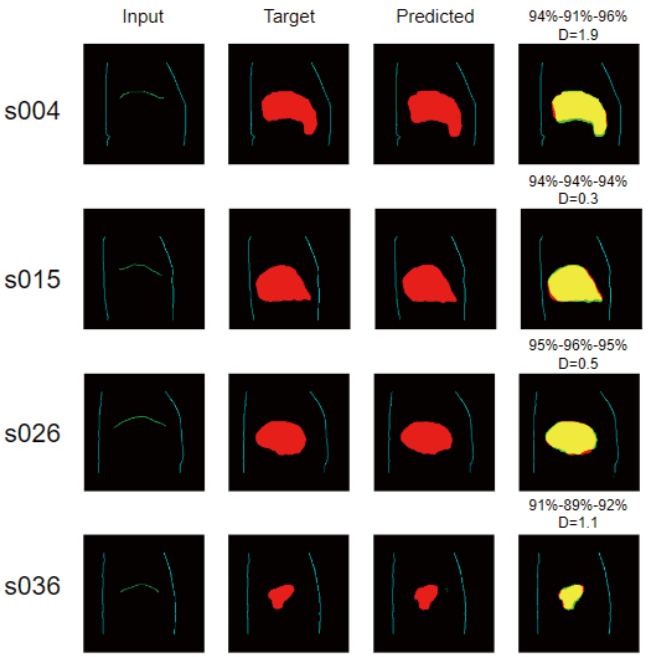

Figure 8. Sample testing result of the trained model on datasets s004, s015, s026, and s036 that are shifted horizontally or vertically to simulate the different locations a patient may lay on a machine.

Table 1. Dice Efficiency, Center Distance, and Image Size

Dataset	Avg DC	StdDev DC	Avg DC1	StdDev DC1	Avg DC2	StdDev DC2	Avg CD	StdDev CD	Avg Size	Max Move
s003	94%	0.8%	91%	2.5%	98%	1.8%	2.3	0.8	84.3	22.2
s004	96%	0.7%	97%	0.9%	95%	0.9%	1.1	0.4	85.8	10.0
s015	97%	0.7%	95%	1.1%	99%	0.7%	1.2	0.5	82.8	25.2
s026	95%	1.0%	96%	2.0%	95%	1.3%	1.6	0.7	76.9	20.4
s036	92%	2.0%	91%	2.5%	93%	3.9%	1.6	1.0	43.0	33.6
s048	89%	1.3%	84%	1.4%	94%	1.8%	4.5	1.1	68.2	39.2
-										
ss003	95%	1.1%	93%	2.3%	96%	2.0%	2.6	0.9	82.8	
ss015	95%	0.8%	95%	1.6%	95%	1.0%	1.3	1.0	81.5	
ss026	94%	1.4%	95%	2.2%	93%	2.9%	1.7	0.9	76.0	
ss036	89%	6.6%	88%	4.8%	91%	9.9%	1.9	2.3	43.1	

the predicted center of the treatment area is close to the target. I have also calculated the "Max Move" for each original test dataset, which is computed as the largest of the distances computed between the centers of the treatment areas of any two test cases in the test dataset.

In addition to the dataset built from the original video and extracted images, I also artificially shifted the entire body in the image to the bottom right, constructed a dataset with the original images, and shifted images to simulate the shift of the human body in the radiotherapy operation space or table that may occur due to human error. Such datasets are given names starting with **ss** such as ss003, ss015, etc. and they are placed in the bottom half of Table 1.

From the result, I can observe that the model can predict the moving liver's position (center) and contour very well when the treatment area is relatively larger which occurs when the majority of the liver needs to be treated. The average DC is high, the standard deviation for each of such datasets is small, and the average CD is also relatively small. These are dataset s004, s015, and s026.

The model performance decreases slightly for the cases where the shape is relatively irregular, such as a patient case or dataset s003, when the treatment area is small, such as dataset s036, or when the treatment area is relatively small and irregular, the performance decreases more, such as in dataset s048.

CONCLUSION

The result shows a very promising future for applying the Pix2Pix method in radiotherapy. The proposed approach is as follows. During a radiotherapy treatment, use some accurate real-time motion tracking methods to collect 5 minutes of data, and use this data to train the Pix2Pix model, then use the trained model to replace these expensive accurate motion tracking methods to guide the radiation rays during the rest of the treatment. Using current technologies, the abdominal diaphragm can be generated with fast ultrasound.

Such an application will greatly reduce the cost, and if the training data is produced with high accuracy, particularly if the contour of the cancerous organ is manually examined and delineated by doctors with high accuracy, then I can achieve almost the same accuracy while automating the rest of the process.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my mentor, Dr. Weihua Mao, for his unwavering support and invaluable guidance throughout this project. His insights and expertise were instrumental in shaping the direction of my research, and his patience in teaching me the fundamentals of this field is something I deeply appreciate.

DECLARATION OF CONFLICT OF INTERESTS

The author declares that there are no conflicts of interest regarding the publication of this article.

REFERENCE

1. Chaput G & Regnier L. Radiotherapy: Clinical pearls for primary care. Canadian family physician Medecin de famille canadien. 2021; 67 (10): 753-757. https://doi.org/10.46747/cfp.6710753
2. Burton A, Beveridge S, Hardcastle N, Lye J, Sanagou M & Franich R. Adoption of respiratory motion management in radiation therapy. Physics and imaging in radiation oncology. 2022; 24: 21-29. https://doi.org/10.1016/j.phro.2022.09.003
3. Mahdavi A, Mofid B & Taghizadeh-Hesary F. Intra-prostatic gold fiducial marker insertion for image-guided radiotherapy (IGRT): five-year experience on 795 patients. BMC medical imaging. 2023; 23 (1): 79. https://doi.org/10.1186/s12880-023-01036-z
4. Bertholet J, Knopf A, Eiben B, McClelland J, Grimwood A, Harris E, Menten M, Poulsen P, Nguyen DT, Keall P & Oelfke U. Real-time intrafraction motion monitoring in external beam radiotherapy. Physics in medicine and biology. 2019; 64 (15): 15TR01. https://doi.org/10.1088/1361-6560/ab2ba8
5. Chen L, Bai S, Li G, Li Z, Xiao Q, Bai L, Li C, Xian L, Hu Z, Dai G & Wang G. Accuracy of real-time respiratory motion tracking and time delay of gating radiotherapy based on optical surface imaging technique. Radiation oncology (London, England). 2020; 15 (1): 170. https://doi.org/10.1186/s13014-020-01611-6
6. Wu Y, Wang Z, Chu Y, Peng R, Peng H, Yang H, Guo K & Zhang J. Current Research Status of Respiratory Motion for Thorax and Abdominal Treatment: A Systematic Review. Biomimetics (Basel, Switzerland). 2024; 9 (3): 170. https://doi.org/10.3390/biomimetics9030170
7. Gardner SJ, Kim J & Chetty IJ. Modern Radiation Therapy Planning and Delivery. Hematology/oncology clinics of North America. 2019; 33 (6): 947-962. https://doi.org/10.1016/j.hoc.2019.08.005
8. Wang Z, Lorenzut G, Zhang Z, Dekker A & Traverso A. Applications of generative adversarial networks (GANs) in radiotherapy: Narrative review. Precision Cancer Medicine. 2022; 5(December), Article 7431. https://doi.org/10.21037/pcm-22-28

9. Yi X, Walia E & Babyn P. Generative adversarial network in medical imaging: A review. Medical Image Analysis. 2019; 58: 101552. https://doi.org/10.1016/j.media.2019.101552
10. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A & Bengio Y. Generative adversarial networks. arXiv. 2014. https://arxiv.org/abs/1406.2661
11. Saxena S & Teli MN. Comparison and analysis of image-to-image generative adversarial networks: A survey. arXiv. 2022. https://arxiv.org/abs/2112.12625
12. Isola P, Zhu J-Y, Zhou T & Efros AA. Image-to-image translation with conditional adversarial networks. arXiv. 2018. https://arxiv.org/abs/1611.07004. https://doi.org/10.1109/CVPR.2017.632
13. pix2pix: Image-to-image translation with a conditional GAN. Available from: https://www.tensorflow.org/tutorials/generative/pix2pix (accessed on 2024-8-1)
14. Gardner SJ, Mao W, Liu C, Aref I, Elshaikh M, Lee JK, Pradhan D, Movsas B, Chetty IJ, Siddiqui F. Improvements in CBCT Image Quality Using a Novel Iterative Reconstruction Algorithm: A Clinical Evaluation. Adv Radiat Oncol. 2019 Jan 10; 4 (2): 390-400. https://doi.org/10.1016/j.adro.2018.12.003