Real-time Target Tracking in Fluoroscopy imaging using Unet with Convolutional LSTM

Tengya Peng, Zhuoran Jiang, Yushi Chang, Lei Ren

Duke University Medical Physics Graduate Program, Duke University Department of Radiation Oncology

Purpose

To investigate the feasibility of tracking targets in 2D fluoro projection images using a novel deep learning network.

Introduction

Radiation therapy, especially stereotatic body radiation therapy (SBRT) is becoming a primary modality to treat early stage non-small cell lung cancer (NSCLC). In SBRT, target localization precision is crucial for the treatment outcome due to the high fractional dose, tight PTV margin, and sharp dose fall off outside PTV. Real time 2D fluoro images are typically used to verify the tumor motion before SBRT treatments start. Physicists and physicians usually review the fluoro images to verify the target motion with reference to the PTV volume, which is a very challenging task due to the overlapping of anatomical structures in 2D projection images. As a result, the verification process is subjective and very much dependent on the experience of the clinicians. In this work, we proposed a neural network model and demonstrated its ability to automate the target localization in fluoro images. To our knowledge, this is the first study to use purely a deep learning method to track tumor motions in the 2D fluoro projection images.

Methods

Our model design, shown in Figure 1, aims to capture the consistent motion of tumors by adopting several most advanced techniques used in Video Object Segmentation (VOS). Specifically, the model is trained by generative methods, which consists of a generator and a discriminator. The generator is a coarse-to-fine architecture design, which has two Unets. Convolutional LSTM modules are introduced in our network to account for the time correlation between different frames of the fluoro images. The convolutional LSTM modules introduce so-called attention mechanism to focus more on essential temporal and spatial information in fluoro images. Apart from this, every three phases of images are combined together to track the current phase, as shown in Figure 2. The loss function contains the adversarial loss, L1 loss, SSIM loss and IOU loss, as shown in Equation 1. The generative loss is from wGAN-GP as shown in Equation 2.

$$L_{total} = \arg\min_{G}\left[\max_{D} L_{GAN}(G,D)\right] + \lambda L_1 + L_{ssim} + L_{iou} \quad (1)$$

$$V(D,G) = E_{x\sim P_{data}}[D(x)] - E_{x\sim P_G}[D(x)] - \lambda E_{\hat{x}\sim P_{\hat{x}}}\left[\left(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1\right)^2\right] \quad (2)$$

The model was trained and tested using a digital X-CAT phantom to demonstrate its feasibility since it provides the flexibility to adjust parameters such as height, width, tumor diameter, position, and respiration amplitude in X-CAT phantom. We conducted experiments from two aspects. In massive samples scenarios, 170 phantoms of different scales, tumor positions, sizes, and respiration amplitudes were generated in X-CAT. Our model was trained, validated and tested using 110, 30, and 30 phantoms respectively. In the other experiment, another 100 phantoms were generated with fixed body and tumor sizes but different respiration amplitudes to achieve the optimal performance on a specific patient and investigate the effects of motion amplitude on the tracking accuracy. In this dataset, the model was trained, validated and tested using 40, 40, and 20 phantoms, respectively. The tracking accuracy was quantitatively evaluated using intersection over union (IOU) and centroid of mass difference (COMD).

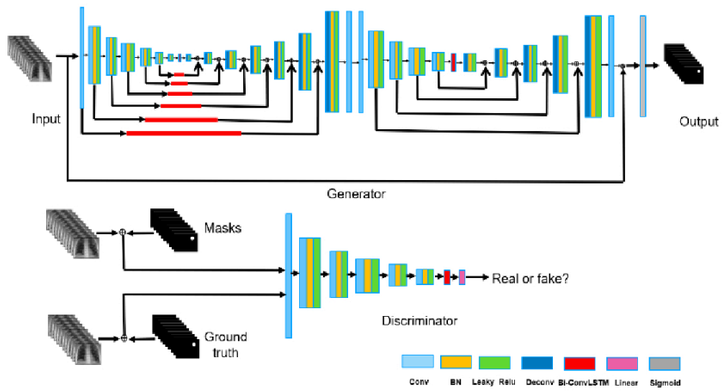

Figure 1. Overall Architecture of Our Network

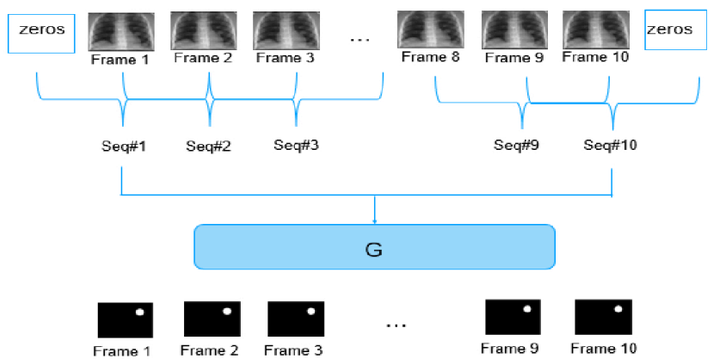

Figure 2. Structure of the Input

Results

In massive sample scenarios, the IOU achieved 0.92 while the COMD was 0.16 cm and 0.07 cm in vertical and horizontal directions on average. Three selective representative samples from our testing dataset are shown in Figure 3. In specific patient scenarios, the IOU achieved 0.98 while the COMD was 0.03 and 0.01 cm in vertical and horizontal directions. Results demonstrated the robustness of our model against breathing variations. In Figure 4, we plotted the vertical coordinate of tumor centroid in one case

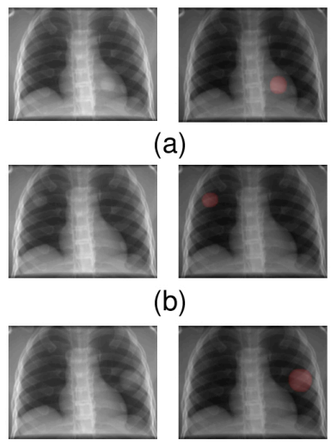

Figure 3. Results of cases in experiment of massive sample scenarios: (a) tumors overlapped by heart (b) tumors partially overlapped by ribs. (c) large tumors in the lung. Red area indicates the tumor location identified by the algorithm.

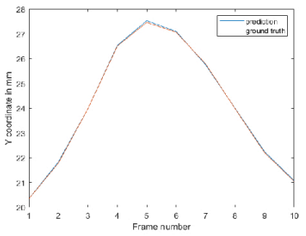

Figure 4. Results for a specific patient scenario: plot of vertical coordinate of tumor centroid

Conclusion

Our study showed the feasibility to use deep learning to track targets in x-ray fluoro projection images without aid of markers. The technique can be valuable for both pre- and during-treatment real time target verification using fluoro imaging in lung SBRT treatments.