

Real-time Liver Tumor Localization via a Single X-ray Projection Using Deep Graph Network-assisted Biomechanical Modeling

Hua-Chieh Shao^a, Jing Wang^a, and You Zhang^a

^aMedical Artificial Intelligence and Automation Laboratory and Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

ABSTRACT

Real-time imaging is highly desirable in image-guided radiotherapy, as it provides instantaneous knowledge of patient's anatomy and motion during the treatment and enables online treatment adaptation to achieve the highest tumor targeting accuracy. Due to extremely limited acquisition time, only one or several X-ray projections can be acquired for real-time imaging, which poses a substantial challenge to localize the tumor from the scarce projections. For liver radiotherapy, such a challenge is further exacerbated by the diminished contrast between the tumor and the normal liver tissues. Here, we propose a framework combining graph neural network-based deep learning and biomechanical modeling to track liver tumor in real time from a single on-board X-ray projection. The liver tumor tracking is achieved in two steps. First, a deep learning network is developed to predict the liver surface deformation, using image features learned from the X-ray projection. Second, the intra-liver deformation is estimated through biomechanical modeling, using the liver surface deformation as the boundary condition to solve intra-liver tumor motion by finite element analysis. The accuracy of the proposed framework was evaluated using a dataset of 10 patients with liver cancer. The results show accurate liver surface registration from the graph-based neural network, which translates into accurate real-time, fiducial-less liver tumor localization (<1.3 mm localization error).

Keywords: Liver, real-time tumor localization, X-ray, deep learning, graph convolutional network, biomechanical modeling

1. INTRODUCTION

The introduction of conformal radiotherapy enables high-precision dose delivery to the tumor and spares surrounding normal tissues, enabling treatment margin reduction, dose escalation, and improved tumor control [1]. However, internal anatomical motion such as respiratory or cardiac motion leads to tumor location uncertainties, and may cause the radiation beams to miss the tumor and damage normal tissues. Image-guided radiation therapy widely uses X-ray based imaging to localize the tumor before and during the treatment to maintain the delivery accuracy [2]. Real-time imaging, in particular, is highly desired as it can localize the tumor instantly and allow the treatment to adapt to such real-time changes to achieve ultimate treatment accuracy. Due to the stringent temporal resolution requirement (hundreds of milliseconds) of real-time imaging, the volumetric information will be severely under-sampled via current mainstream imaging modalities including cone-beam computed tomography (CBCT). Such a degree of under-sampling makes it impossible to reconstruct high-quality CBCTs using conventional methods for tumor localization.

Due to the recent successes of deep learning (DL), several groups have proposed DL-based methods for real-time imaging. A few network architectures were proposed to reconstruct three-dimensional (3D) CBCT images from single-view or orthogonal-view X-ray projections [3-5]. Such networks, however, were built on an ill-conditioned problem, trying to estimate high-dimensional volumetric data from a single X-ray projection. Considerable reconstruction errors remain, albeit much smaller than those of the conventional reconstruction algorithms. Moreover, to track tumors in real-time, additional steps of image registration or segmentation are necessary to further localize tumors from the reconstructed CBCT images. This is particularly challenging for liver tumors due to the low contrast of liver tumors against surrounding normal liver parenchyma.

To address the above challenges toward real-time imaging, especially for liver tumor localization, we propose a mesh registration-based method combining deep neural networks with biomechanical modeling. The method directly solves

the liver tumor motion between a prior CT/CBCT image and a single X-ray projection to localize liver tumors in real-time, and effectively eliminates the need to reconstruct a high-quality, intermediate CBCT image prior to localization. Specifically, a deep graph neural network-based architecture was trained to model the correlation between patient-specific liver boundary motion and features learned on individual X-ray projections. The trained network can then predict liver boundary motion from a single real-time X-ray projection. Using the predicted liver boundary motion as the boundary condition, we further performed finite element analysis-based biomechanical modeling of liver to solve intra-liver tumor motion. The method adopts a deformation-driven approach that incorporates prior information to tackle the extreme under-sampling issue. The two-step-based registration scheme simplifies the complexity of the deep graph neural network with introduced domain knowledge (biomechanical modeling). Biomechanical modeling uses information including structure geometry, material composition and elasticity to derive physiologically and physically meaningful deformation, and complements the intensity information provided in the X-ray projection to further improve the registration and tumor localization accuracy [6].

The accuracy of the proposed technique was evaluated using 10 patients with liver cancer, and compared with the accuracy of two other techniques. The first technique uses the diaphragm as an anatomic landmark, and tracks the diaphragm motion directly from the on-board projections via template matching to represent liver tumor motion. The second technique is a principal component analysis (PCA) based method which models 3D motion into a few motion eigenvectors for dimension reduction and tumor tracking [7].

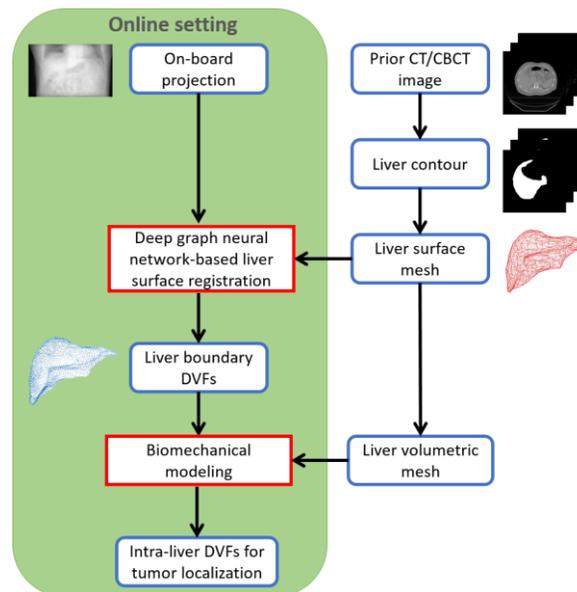


Figure 1. Workflow of the proposed method. A deep graph neural network-based model was trained to predict liver surface deformation vector field (DVF) from a single X-ray projection. Then a biomechanical model solves the intra-liver DVF using the liver surface DVF as the boundary condition for tumor localization.

2. MATERIALS AND METHODS

2.1 Method overview

In this study, liver motion and liver tumor localization were solved via deformable registration between a liver mesh (extracted offline from a prior CT/CBCT image available before the treatment) and the liver features projected on a single X-ray projection (Fig. 1). The registration was achieved via two steps: (a) liver surface motion estimation via a deep graph neural network-based structure (Fig. 2); and (b) intra-liver motion estimation via biomechanical modeling. Specifically, in step (a) a patient-specific DL model was trained to predict a liver boundary deformation vector field (DVF) that deforms the prior liver surface mesh to match with the liver shape variations encoded in the X-ray projection. In step (b), a biomechanical model of the liver was built, and an intra-liver DVF was solved through finite element analysis using the liver boundary DVF as the boundary condition.

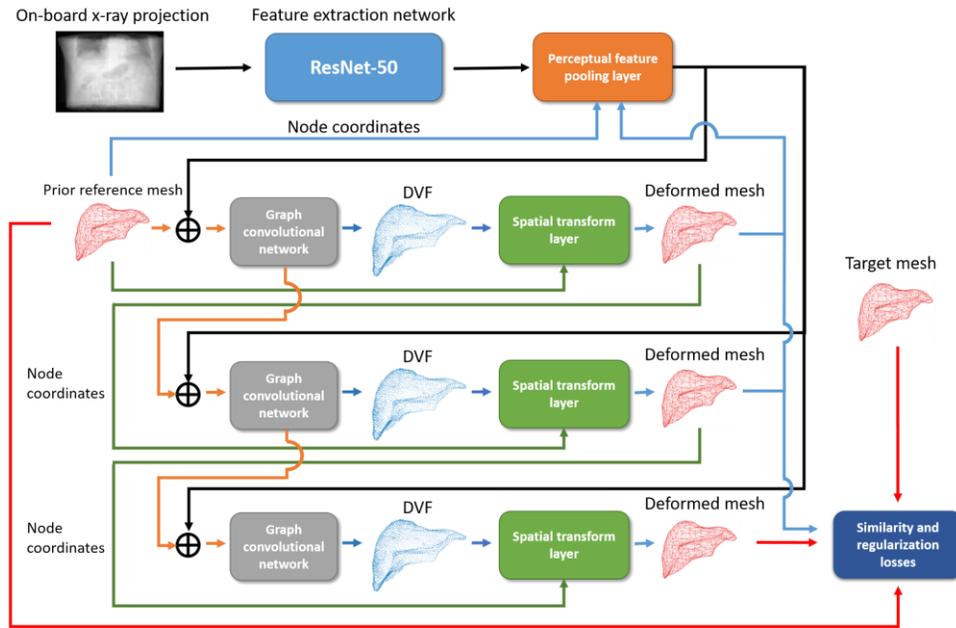


Figure 2. Overview of the deep-learning (DL) network that estimates liver boundary motion from an on-board X-ray projection. The network consists of two subnetworks performing feature extraction and liver boundary DVF prediction separately. The first subnetwork uses ResNet-50 to extract image features from an X-ray projection. The extracted feature maps were pooled for each node of a liver surface mesh by the perceptual feature pooling layer, based on the projected node coordinates on the X-ray projection. The second subnetwork, consisting of three deformation blocks, progressively estimates liver boundary DVFs. A deformation block comprises of a graph convolutional network (GCN) and a spatial transform layer. The GCN was learned to predict a liver boundary DVF based on the features extracted from the ResNet-50 subnetwork. A spatial transform layer deforms the prior reference mesh or the deformed liver surface mesh from the previous deformation block, using a GCN-predicted DVF.

2.2 The deep-learning network architecture

The network was trained to learn image features from on-board X-ray projections to predict boundary movement from the prior liver mesh to each on-board projection. The DL network architecture is illustrated in Fig. 2, which contains two subnetworks. The first subnetwork extracts image features from each on-board X-ray projection, and the extracted feature maps are pooled and fed into the second subnetwork for liver boundary DVF prediction. Here we used ResNet-50 [8] as the feature extraction network. Consisting of a series of convolutional layers stacked in a residual learning architecture, ResNet-50 extracts encoded liver shape variations, via local and global image features contained in the X-ray projection, and learns short- and long-range dependencies among these extracted features. These learned dependencies are helpful for the deformation estimation because they are shown common in respiration-induced liver motion. The perceptual feature pooling layer pools the ResNet-50 extracted feature maps by associating each 3D node of a liver surface mesh with a 2D point in the feature maps, based on the same geometry of the cone-beam projection.

The second subnetwork comprises a series of deformation blocks that progressively deforms the liver surface mesh nodes based on the extracted feature maps from the first subnetwork. Each deformation block involves a graph convolutional network (GCN, Fig. 3) and a spatial transform layer that deforms the liver surface mesh using the GCN-predicted DVF [9]. GCN performs graph-based convolutions that generalize the standard convolution operations to data structures lack of underlying Euclidean structures, such as functional networks in brain imaging. A non-Euclidean data structure can be represented by a weighted graph comprised of a set of vertices, edges connecting the vertices, and weights associated with each vertex (e.g., vertex features, DVFs, vertex-associated image features). The use of GCN is indicated for our problem, as the liver surface mesh nodes, the geometrical connectivity (edges) between the nodes, and the learned image, DVF, and vertex features associated with each node make a standard non-Euclidean data structure for inputs into the GCN. Using extracted image features that encode the liver shape, preceding DVFs, and learned vertex features from the previous block, the GCN learns to predict a liver surface DVF to further deform the surface mesh

deformed by the previous block. The inputs of the GCN in the first deformation block contain ResNet50-extracted image features and an initial DVF which was set to be zero. For each subsequent GCN, the image features were re-pooled based on the new node coordinates (Fig. 2), deformed via the spatial transform layer and the DVF predicted by the preceding block. The image features were then input into the subsequent GCN, along with the predicted surface DVF and the learned vertex features from the preceding block. We used a GCN of the same architecture as the G-ResNet [9], which is illustrated in Fig. 3. The corresponding network was modified and adapted from the Pixel2Mesh library [10]. The model training was driven by a loss function involving a mesh similarity loss and regularization losses that regularize the deformation and enforce smoothness of the boundary DVFs.

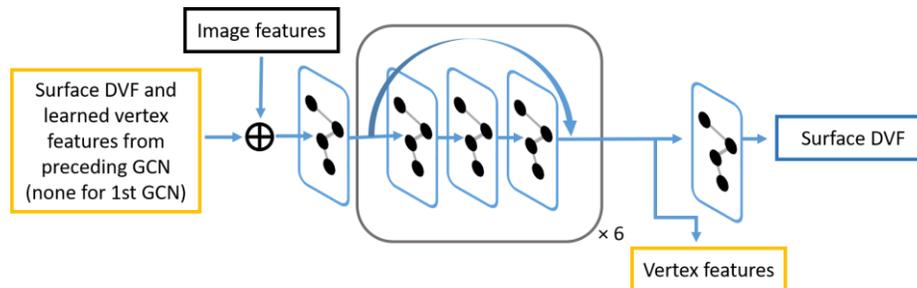


Figure 3. Graph convolutional network (GCN). The inputs contain pooled image features from the feature extraction ResNet-50 subnetwork (Fig. 2), a surface DVF, and vertex features yielded from the GCN in the previous deformation block (if any). The GCN consists of 20 graph convolution layers that, except for the entrance and exiting layers, were organized in a residual learning architecture. The GCN yields a surface DVF and vertex features to feed into the subsequent deformation block. The inputs of the first GCN in the second subnetwork contains only image features and an initial surface DVF which was set to be zero. The image features were re-pooled for each GCN based on deformed node coordinates. The rounded box in the middle represents a residual learning module containing three graph convolution layers with a shortcut connection, which iterates 6 times.

2.3 Biomechanical modeling

After the deep neural network solves a liver boundary DVF to match with the liver shape features on the X-ray projection, the intra-liver DVF was subsequently derived using a biomechanical model. Here we used the Mooney-Rivlin material model, which describes a hyperelastic (i.e., nonlinear elasticity) material that fits biological tissues well. The details of implementing the biomechanical model can be found in Ref. [6].

2.4 Dataset curation and augmentation

A dataset of 10 patients with liver cancer from our institute was used to evaluate the proposed method. The study was approved under an institutional review boards protocol. Each patient had a contrast-enhanced four-dimensional CT set from treatment planning, and the CT images were binned into 10 respiratory phases (from 0% to 90%), with 0% being the end-of-inhale phase. The CT images were resampled to a uniform size of $256 \times 256 \times 128$ with an isotropic resolution of $2 \text{ mm} \times 2 \text{ mm} \times 2 \text{ mm}$. On-board X-ray cone-beam projections were simulated from the CT images using a ray-tracing algorithm. We simulated projections from three angles: 0, 45, and 90 degrees. The 0- and 90-degree are for anterior-posterior and left-right directions, respectively.

Since each patient had only a 10-phase 4D-CT set, to generate sufficient motion variation scenarios to train the patient-specific network and avoid overfitting, we augmented the dataset of each patient by simulating realistic respiratory deformations encountered in on-board liver imaging. The augmentation was based on a PCA-based motion model of each patient [6, 11]. We first performed deformable registrations between the reference 0% phase and the other phases to attain DVFs, using the open-source software package Elastix. To improve the intra-liver DVF accuracy, we applied biomechanical modeling to derive intra-liver DVFs, using the liver surface DVFs solved by Elastix as boundary conditions. We then replaced the Elastix intra-liver DVFs with the biomechanical modeling-derived intra-liver DVFs. PCA was subsequently performed on these high-quality DVFs of each patient to obtain patient-specific principal motion components. For augmentation, the coefficients of the first three principal motion components were randomly scaled to re-generate DVFs of various magnitudes and patterns [11]. In total, for each patient we generated 1,728 augmented

samples which were partitioned into training, validation, and testing sets. The partitioning was assigned according to the original respiratory phases of the PCA coefficients prior to the random scaling. The training set includes the samples of which the original PCA coefficients were from the 10% to 40% phases; the validation set includes the samples whose original PCA coefficients were from the 60% and 70% phases; and the testing set includes the samples whose original PCA coefficients were from the 50%, 80%, and 90% phases. 50% is the end-of-exhale phase that has the largest deformation from the 0% phase.

2.5 Evaluation schemes

The deformation accuracy of liver surface meshes was evaluated using the Hausdorff distance (HD) between the deformed and the ‘ground-truth’ target liver surface meshes extracted from the augmented dataset [12]. To evaluate the performance of liver tumor tracking, we manually contoured the tumors from the prior CT images at phase 0%. The tumor contours at 0% phase were then propagated using the augmentation DVFs (I.D.) to other augmented motion states, which were used as the ‘ground-truth’ to evaluate the ones deformed by our method. The accuracy of liver tumor tracking was evaluated by the Dice similarity score (DSC), center-of-mass error (COME), and HD.

3. RESULTS

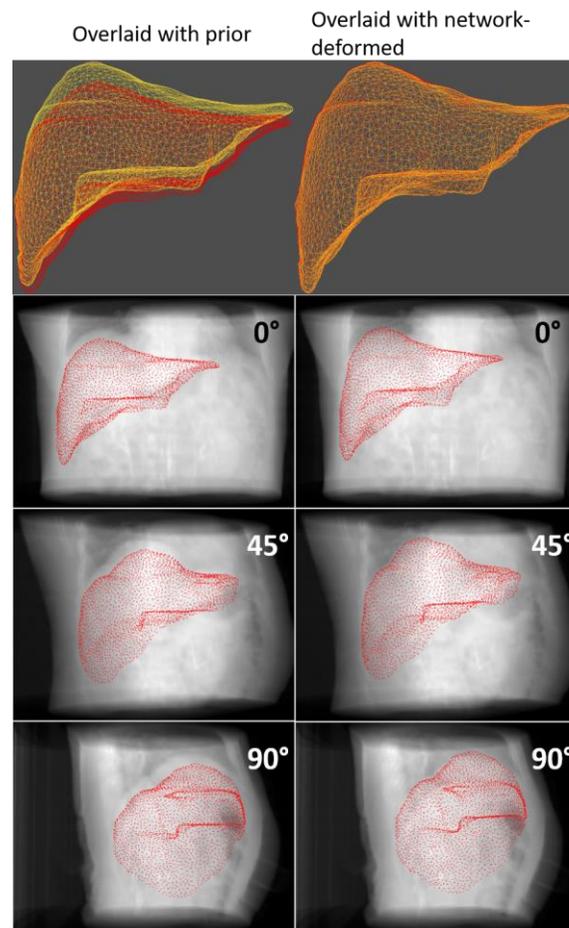


Figure 4. (First row) Liver surface overlays between the prior and ‘ground-truth’ target meshes (left) and between the graph network-deformed and target meshes (right). The yellow meshes are the target meshes corresponding to the end-of-exhale phase after augmentation, and the red meshes correspond to the prior (left panel) and deformed (right panel) meshes. (Other rows) Liver surface nodes projected on X-ray projections at three projection angles. Left and right columns show the projected nodes corresponding to the prior and deformed surface meshes, respectively.

3.1 Liver deformation accuracy

Figure 4 presents a qualitative comparison of liver surface meshes and projected nodes on X-ray projections at three projection angles (0, 45, and 90 degrees). The first row shows the surface mesh overlays between the prior and ‘ground-truth’ target meshes (left panel), and between the graph network-deformed and ‘ground-truth’ target meshes (right panel). The prior and target meshes correspond to the end-of-inhale phase and the end-of-exhale phase (with motion augmentation), respectively. The other rows show the overlay of the pre- (left panel) and post-registration (right panel) surface mesh nodes onto the corresponding X-ray projections of the end-of-exhale phase, at three different angles. Both the surface mesh overlay and the node-projection overlay demonstrate high registration accuracy.

Table 1 summarizes the mean (\pm s.d.) liver HDs of the proposed method and the PCA-based 2D-3D registration method [7]. The DL-based method results in much smaller HDs than the PCA-based 2D-3D method.

Table 1. Mean (\pm s.d.) liver Hausdorff distances.

Projection angle (degree)	Prior (mm)	Method	
		DL prediction (mm)	PCA-based 2D-3D registration (mm)
0		2.99 \pm 2.42	7.27 \pm 4.18
45	11.77 \pm 6.11	3.03 \pm 2.39	6.55 \pm 3.32
90		3.09 \pm 2.55	6.09 \pm 2.47

3.2 Liver tumor tracking accuracy

Table 2 summarizes the mean (\pm s.d.) liver tumor DSCs, COMEs, and HDs of the proposed method at three projection angles. In addition, the results of PCA-based 2D-3D registration and diaphragm tracking are also presented in the Table for comparison. The diaphragm tracking is only able to localize the diaphragm in 2D from a single X-ray projection, thus we only used it to represent liver tumor motion along the superior-inferior (SI) direction. The COME of the diaphragm-based method is thus only for SI direction, with 3D COME potentially being much larger. Table 2 clearly shows that the proposed method has the best liver tumor localization accuracy, and the performance is consistent among different angles.

Table 2. Mean (\pm s.d.) liver tumor DSC, COME, and HD. The COME for the diaphragm-based method is for the superior-inferior direction only (*).

Project-ion angle (deg.)	Metric	Prior	Method		
			DL prediction	PCA-based 2D-3D registration	Diaphragm tracking
0			0.895 \pm 0.112	0.789 \pm 0.205	
45	DSC	0.547 \pm 0.269	0.893 \pm 0.110	0.822 \pm 0.155	--
90			0.886 \pm 0.118	0.835 \pm 0.134	
0	COME (mm)		1.13 \pm 1.33	2.53 \pm 4.32	1.68 \pm 2.22*
45		6.08 \pm 4.40	1.15 \pm 1.32	1.84 \pm 1.64	2.69 \pm 2.73*
90			1.25 \pm 1.41	1.73 \pm 1.37	3.08 \pm 3.26*
0	HD (mm)		2.81 \pm 1.77	3.95 \pm 5.00	
45		7.24 \pm 4.92	2.86 \pm 1.77	3.17 \pm 1.81	--
90			2.93 \pm 1.85	3.01 \pm 1.36	

REFERENCES

- [1] Verellen, D., et al., Innovations in image-guided radiotherapy. Nature Reviews Cancer, 2007. 7(12): p. 949-960.
- [2] Dhont, J., et al., Image-guided Radiotherapy to Manage Respiratory Motion: Lung and Liver. Clinical Oncology, 2020. 32(12): p. 792-804.

- [3] Shen, L.Y., W. Zhao, and L. Xing, Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. *Nature Biomedical Engineering*, 2019. 3(11): p. 880-888.
- [4] Lei, Y., et al., Deep learning-based real-time volumetric imaging for lung stereotactic body radiation therapy: a proof of concept study. *Physics in Medicine and Biology*, 2020. 65(23).
- [5] Wei, R., et al., Real-time tumor localization with single x-ray projection at arbitrary gantry angles using a convolutional neural network (CNN). *Physics in Medicine and Biology*, 2020. 65(6).
- [6] Zhang, Y., et al., 4D liver tumor localization using cone-beam projections and a biomechanical model. *Radiother Oncol*, 2019. 133: p. 183-192.
- [7] Zhang, Y., et al., A technique for estimating 4D-CBCT using prior knowledge and limited-angle projections. *Medical Physics*, 2013. 40(12).
- [8] Paszke, A., et al., PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems 32 (Nips 2019)*, 2019. 32.
- [9] Wang, N.Y., et al., Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images. *Computer Vision - Eccv 2018, Pt Xi*, 2018. 11215: p. 55-71.
- [10] Cao, J. Pixel2Mesh. 2021 [cited 2022 January 10]; Available from: <https://github.com/noahcao/Pixel2Mesh>.
- [11] Jiang, Z., et al., Enhancing digital tomosynthesis (DTS) for lung radiotherapy guidance using patient-specific deep learning model. *Phys Med Biol*, 2021. 66(3): p. 035009.
- [12] Taha, A.A. and A. Hanbury, Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *Bmc Medical Imaging*, 2015. 15