Real-Time Patient-Specific Lung Radiotherapy Targeting using Deep Learning

Markus D. Foote*
Scientific Computing and Imaging Institute
Department of Bioengineering
University of Utah
Salt Lake City, UT, USA
foote@sci.utah.edu

Blake Zimmerman
Scientific Computing and Imaging Institute
Department of Bioengineering
University of Utah
Salt Lake City, UT, USA

Amit Sawant
Department of Radiation Oncology
The University of Maryland, School of Medicine
Baltimore, MD, USA

Sarang Joshi
Scientific Computing and Imaging Institute
Department of Bioengineering
University of Utah
Salt Lake City, UT, USA

Abstract

Radiation therapy has presented a need for dynamic tracking of a target tumor volume. Fiducial markers such as implanted gold seeds have been used to gate radiation delivery but the markers are invasive and gating significantly increases treatment time. Pretreatment acquisition of a 4DCT allows for the development of accurate motion estimation for treatment planning. A deep convolutional neural network and subspace motion tracking is used to recover anatomical positions from a single radiograph projection in real-time. We approximate the nonlinear inverse of a diffeomorphic transformation composed with radiographic projection as a deep network that produces subspace coordinates to define the patient-specific deformation of the lungs from a baseline anatomic position. The geometric accuracy of the subspace projections on real patient data is similar to accuracy attained by original image registration between individual respiratory-phase image volumes.

1 Introduction

According to the Centers for Disease Control and Prevention, lung cancer is the leading cause of cancer death, accounting for 27% of all cancer deaths in the United States [1]. Computed tomography (CT) is a widespread screening and diagnosis tool for lung tumors; however, respiratory motion causes uncertainty in tumor localization, leading to complications in treatment procedures [2]. Radiotherapy is often a successful treatment option for lung cancer [3]. The combination of radiotherapy and uncertain tumor location results in decreased dose delivered to the tumor and radiation dose being delivered to the healthy surrounding tissue and nearby vital organs.

Accurate estimation of organ movement and deformations plays a crucial role in dose calculations and treatment decisions in radiation therapy of lung cancer [4]. State-of-the-art radiation treatment planning uses 4D (3D + time) computed tomography (4DCT) scans as a planning tool to deliver radiation dose to the tumor during treatment [5]. Radiation dose delivered to surrounding healthy tissue and vital organs causes undesired side effects of radiotherapy. Understanding respiratory motion allows for more accurate treatment planning, resulting in more targeted dose delivery to only the tumor.

*Corresponding author.

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

While general motion patterns in radiotherapy patients are relatively well understood, cycle-to-cycle variations remain a significant challenge and may account for observed discrepancies between predictive treatment plan indicators and clinical patient outcomes, especially due to an increased irradiated volume which limits adequate dose to the target [6]. Gating methods have developed into adaptive radiotherapy but require implanted markers and lengthen treatment time. Long treatment sessions and implanted fiducial markers remain problematic and invasive methods for rigid motion tracking in conformal stereotactic body radiation therapy [7]. Accurate and fast noninvasive target tracking that accounts for real-time cycle-to-cycle variations in respiratory motion is a recognized need in radiotherapy [8].

We propose a real-time framework to recover the anatomic position of the tumor target and surrounding anatomy using radiographic images of the treatment area. Using pre-treatment 4DCT datasets, the patient-specific respiratory motion patterns can be identified through our framework in real-time. The framework and motion results can be incorporated as the position monitoring subsystem in a conformal radiotherapy system [8]. With deep learning the combination of these motion patterns observed in real-time radiographs can be recovered to determine the shift in target position from the targeted baseline position. This method eliminates the need for invasive fiducial markers while still producing targeted radiation delivery during variable respiration patterns. While the proposed framework requires training of the model on a per-patient basis, only inference is required during treatment time. Inference of the deep convolutional network and subsequent linear combination of patient-specific motion components can be calculated faster than real-time and drive motion-tracking of conformal radiotherapy treatment.

2 Related Work

2.1 PCA Breathing Models

The respiratory cycle exhibits low-rank behavior and hysteresis, leading to application of principal component analysis to study the deformation of lungs and other organs through time [9]. Previous work by Sabouri *et al.* has used such PCA approximations of the respiratory cycle to correlate the lung's principal deformations to respiratory markers directly observable during treatment such as body surface tracking during conformal radiation therapy [10].

2.2 Rank Constrained Density Motion Estimation

Our targeting procedure builds upon the rank-constrained diffeomorphic density matching problem, summarized here for completeness [11–14]. The non-smooth rank of a set of deformations is measured by a smooth surrogate using the nuclear norm of the deformation set matrix. The deformation set matrix is constructed with each vectorized deformation field as a row,

$$X = \begin{bmatrix} \varphi_1^{-1}(x) - x \\ \varphi_2^{-1}(x) - x \\ \vdots \\ \varphi_{N-1}^{-1}(x) - x \end{bmatrix} = \{\varphi_i^{-1}(x) - x\} \tag{1}$$

where φ_i^{-1} is the inverse of the deformation from the i-th image in the image series to a selected reference image. We can thus define the nuclear norm for deformations between N images as

$$\|X\|_* = \sum_i^{N-1} \sigma_i(X) \tag{2}$$

as there are $N-1$ deformations between N images, giving only $N-1$ singular values σ.

As CT measures the linear attenuation coefficient which is proportional to the true density, 4DCT volumes are interpreted as densities, thus values change with compression or expansion. A density $I\,dx$ is acted upon by a diffeomorphism φ to compensate for changes of the density by the deformation:

$$(\varphi, I\,dx) \mapsto \varphi_*(I\,dx) = \left(\varphi^{-1}\right)^*(I\,dx) = \left(|D\varphi^{-1}| I \circ \varphi^{-1}\right) dx \tag{3}$$

where $|D\varphi^{-1}|$ denotes the Jacobian determinant of φ^{-1}. The Riemannian geometry of the group of diffeomorphisms with a suitable Sobolev H^1 metric is linked to the Riemannian geometry of

densities with the Fisher-Rao metric [12, 15, 16]. The Fisher-Rao metric is used as the measure for image similarity due to the property that it is invariant to the action of diffeomorphisms:

$$d_F^2\left(I_0\,dx, I_1\,dx\right) = \int_\Omega \left(\sqrt{I_0} - \sqrt{I_1}\right)^2 dx\ . \tag{4}$$

The linkage between a suitable Sobolev H^1 metric and the Fisher-Rao metric allows for evaluation of the distance in the space of diffeomorphisms in closed form. The Fisher-Rao metric, an incompressibility measure, and surrogate rank measure can then be used to match an image set by minimizing the energy functional:

$$E(\{\varphi_i\}) = \sum_i^{N-1} \left[\int_\Omega \left(\sqrt{\left|D\varphi_i^{-1}\right| I_i \circ \varphi_i^{-1}} - \sqrt{I_0} \right)^2 dx \right. \\ \left. + \int_\Omega \left(\sqrt{\left|D\varphi_i^{-1}\right|} - 1 \right)^2 f\,dx \right] + \alpha \sum_i^{N-1} \sigma_i\left(\{\varphi_i^{-1}(x) - x\}\right) \tag{5}$$

where I_0 is a chosen base or reference density and I_i are the other $N-1$ densities in the series. The first term here penalizes dissimilarity between the two densities. The second term penalizes deviations from a volume-preserving deformation. The penalty function f acts as weighting of the volume-preserving measure. The final term penalizes the surrogate rank of the subspace in which the set of deformations exist. This optimization produces a related set of rank-optimized deformations with improved geometric accuracy over other registration methods due to the increased physiologic relevance of the low-rank deformations which match well with the general reversal process of an inhale-exhale cycle, along with hysteresis in other components.

2.3 Digitally Reconstructed Radiographs

For a given 4DCT dataset, it is sufficient for training to simulate the 2D projection images at different respiratory states from the 3D volumes acquired at different states. These digitally reconstructed radiographs (DRR) are commonly used for radiation therapy planning. This allows for simulation of the 2D projections that would be acquired at the different phases of breathing throughout the radiotherapy treatment [17]. A point $(u, v) \in \mathbb{R}^2$ in the DRR projection image is defined by the line integral of linear attenuation coefficients (values in the CT volume) over distance l from the x-ray source to the projection image. The path can be written as $p(s) = (s(u-u_0), s(v-v_0), sl)$ where $s \in [0, 1]$ and (u_0, v_0) is the piercing point, or the unique point in the projection image that is closest to the x-ray source. The DRR operation generates a 2D projection of the 3D CT volume that can be used for training in place of actual radiography recordings. These projections closely approximate the radiographic imagery that would be acquired during a state-of-the-art radiotherapy treatment session.

3 Methodology

3.1 Subspace Learning

An existing 4DCT with 10 phase volumes of a lung cancer patient for treatment planning from the University of Maryland was provided in anonymized form. As the original volumes contain significant empty space around the patient, the volumes were cropped to contain only the patient's thorax. Resulting images are $304 \times 233 \times 133$ with resolution $1.17 \times 1.17 \times 2$ mm. These 10 3D CT volumes, each representing a distinct respiratory phase, are used for rank constrained density registration as described by Foote *et al.*to produce 9 deformation fields that map to the full-exhale phase volume from each of the other phase volumes [11]. The low-rank subspace of the space of diffeomorphisms in which the rank-optimized deformation components exist is determined via principal component analysis (PCA). By viewing each deformation as a point in a high dimensional diffeomorphism space, each deformation forms a row of the data matrix $X \in \mathbb{R}^{(304\times233\times133\times3)\times9}$. The Gram matrix XX^T is calculated for PCA rather than the covariance matrix X^TX, as the eigenvalues are equivalent and the memory requirement for storing a small (9×9) matrix is manageable over a large $\left(\sim 256^3 \times \sim 256^3\right)$ matrix. The eigenvalues and vectors of the gram matrix are used to define weights for the principal components. Rank-optimized motion component fields are calculated from these weights as linear combinations of the original deformation fields.

Table 1: Data explanation power of cumulative rank-optimized components of deformation from rank-constrained density deformation optimization. These values are calculated from the normalized cumulative sum of eigenvalues.

Number of Included Components	Percentage of Dataset Explained
1	0.8587115
2	0.9939205
3	0.9978112
4	0.9992142
5	0.9996793
6	0.9999000
7	0.9999791
8	1.0

3.2 Synthetic Dataset Generation

A dataset for training is generated by projecting through the full-exhale CT volume with applied deformations according to weights of the rank-optimized components. The weights are generated as a dense grid of 201 evenly spaced points in each of the first two rank-optimized component directions (40401 total samples). Only the first two rank-optimized components are used as they explain 99.39% of the deformation data (Table 1). The range of the weight points is set as the maximum magnitude of the corresponding rank-optimized component for that dimension. These weights serve as the target labels for training and are then used to calculate a linear combination of the rank-optimized component deformations. The resulting grid of deformations are calculated and applied to the full-exhale 4DCT volume using PyCA [18]. DRR images are calculated by ray casting and integrating along a cone-beam geometry through the deformed full-exhale 4DCT volume. The projection image was positioned on the back face of the CT volume with twice the spatial resolution as the CT, resulting in a 454×608 image with resolution 0.58×0.58 mm. The x-ray source was positioned three times the thickness of the CT volume away from the projection image plane (Fig. 1). Projection images were preprocessed before training by normalizing the image intensities to [0,1] and subsequent histogram equalization.

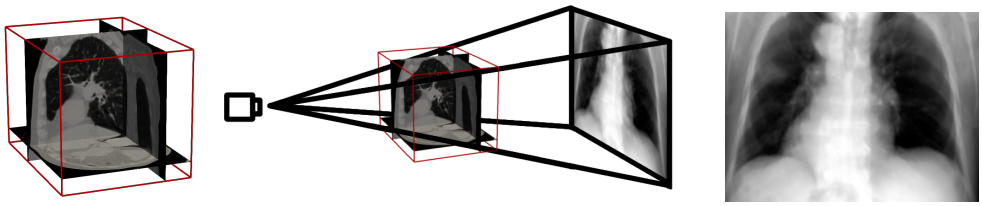

Figure 1: Left: Three coordinate-aligned slices of the deformed full-exhale 4D CT volume. Center: Representative geometry of digitally reconstructed radiograph process including the x-ray source position, CT volume, and radiograph image on the detector plane. Right: Representative DRR from the synthetic dataset used for training.

3.3 Network Architechture

The mapping from rank-optimized component weights to a projection image is highly nonlinear, including diffeomorphic transformation and radiographic projection. We aim to approximate the inverse of this mapping with a deep convolutional neural network as a regression problem. A promising network architecture for this application is DenseNet [19]. DenseNet has achieved state-of-the-art performance in classification on several datasets while using fewer parameters and decreasing over-fitting. The direct connections between constant spatial layers allow for reusing features learned from previous layers and improve the flow of gradients throughout the network. Additionally, DenseNet is ideal for real-time applications due to reduced model complexity and depth.

We use an efficient DenseNet model in PyTorch slightly tailored to our application [20, 21]. The input to the network is a single gray scale projection image. The rank-optimized component weights used to produce the projection image are the regression points (Fig. 2). Summarized here for completeness, the DenseNet architecture convolves each input with a kernel size of 3 and filter number of $2 \times k$ before being input to the first dense block, where k is the growth rate. Each subsequent dense block consists of 8 convolutional layers all with filter size of 3 and a growth rate of 12. The 8 convolutional layers within a dense block are each preceded by a batch normalization layer and a rectified linear unit (ReLU) layer. The output from a dense block enters a transition block where batch normalization and non-linear activation are followed by a convolutional feature map reduction layer (compression=0.5) and spatial reduction of 2 in each dimension using max-pooling. A total of 5 dense and transition blocks are used to reduce the spatial dimensions to final feature maps. The final layer of the network was a linear layer that regresses the feature maps to the rank-optimized component weights.

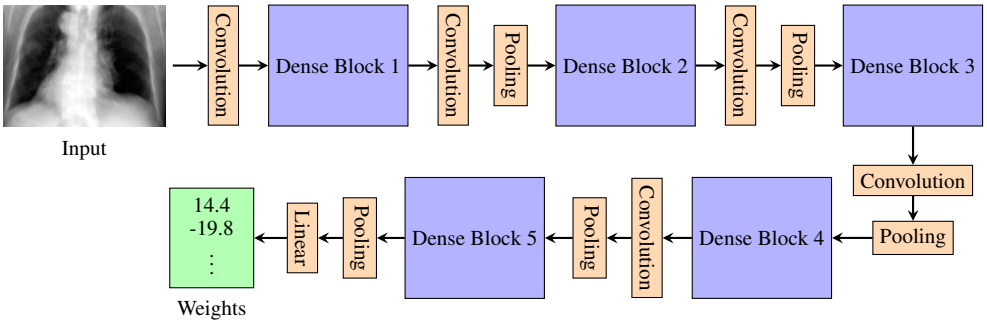

Figure 2: DenseNet architecture used to learn the inverse deformation-projection estimate. Each dense block represents a group of 8 feature layers and growth rate 12 with an included transition layer. A final linear layer recovers the weights for deformation components.

3.4 Training

The network is trained for 200 epochs using a smooth-L_1 loss function with a learning rate of 0.001 and batch size of 32. The learning rate is decreased by a factor of 10 when the training loss plateaus within a relative threshold of 0.005 for the last 10 epochs. Half of the generated dataset is randomly selected to be the training set; the other half was used as the testing set to monitor for model over-fitting during the training process.

4 Experiments

4.1 Training and Test Performance

Training performance was tracked by the smooth-L_1 loss on the testing dataset inference (Fig. 3). Learning rate scheduling resulted in learning rate decreases at epochs 67 and 168. Training for 200 epochs takes approximately 10 hours on a single Nvidia Tesla V100 GPU.

4.2 Spline Model Deformation Validation

A breathing model was made using a spline interpolation of the first two rank-optimized component weights from the 9 deformations (Fig. 4). At any point along the model curve, the coordinate weights provide a linear combination of the rank-optimized component deformations to apply to the full-exhale volume to obtain a model volume for the corresponding breathing phase. Projections through these model volumes in the same manner as the training dataset generation produced model radiograph for evaluation. As with the training images, these projection images were normalized and histogram equalized. Inference of the trained network was used to recover the weights of the rank-optimized components. Inference on a single Nvidia Titan V GPU has a throughput of 1051 images/second with the same PyTorch implementation – significantly faster than real-time. The relation of these weights to the original model is shown in Figure 4.

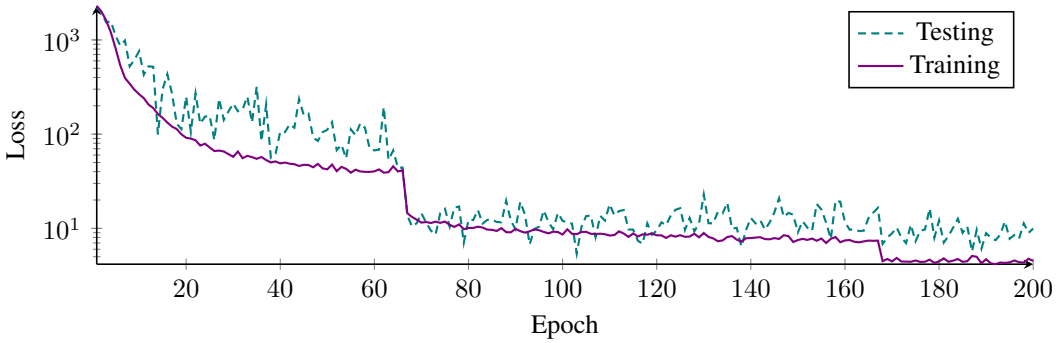

Figure 3: Training and test loss during training. Note the logarithmic vertical axis.

Accuracy of the deformation recovered from the inferred weights is measured by maximum deformation distance error compared to the reconstruction of these deformations with the two rank-optimized deformation component weights directly from the model. From the 40 model points, the maximum error between the applied deformation and the recovered deformation in the tumor region was 0.057 mm.

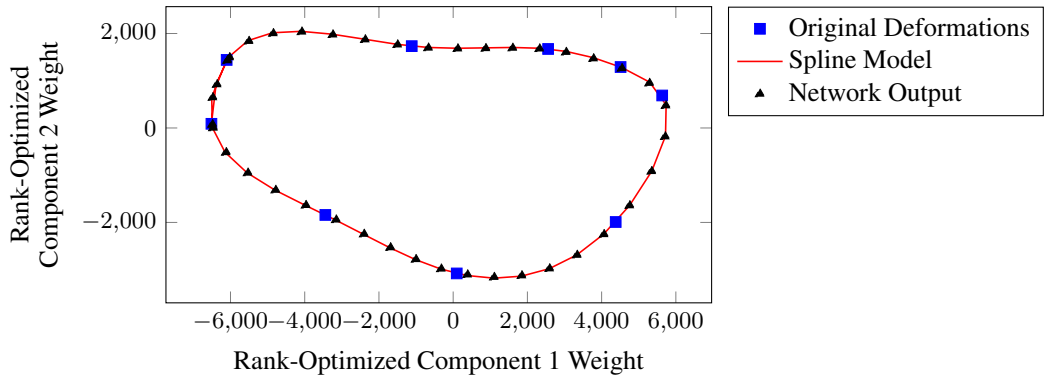

Figure 4: Deformation component weights for each of the 9 original components (blue squares) are control points for the spline-based breathing model (black dashed line). The respiratory cycle proceeds clockwise around the loop. DRR images derived from the weights along this spline model are input to the network. Resulting inferred weights (red line) closely align with the model.

4.3 RCCT Phase Patient Data and Geometric Validation

Until this point, DRR images derived from 9 of the 10 original 4DCT volumes have not been used as input for the network for either training or previous testing. Rather than synthetic models of respiratory phases, we now use the original CT volumes captured at 9 stages of the respiratory cycle. Similar to training dataset generation, DRR images through these (undeformed) volumes are calculated, normalized, and histogram equalized. No deformation is applied to these CT volumes as each volume represents an intrinsic respiratory deformation relative to the full-exhale anatomical state. Evaluation of the network gives rank-optimized deformation component weights describing this intrinsic deformation that can be compared in 2D component weight space with the weights directly recovered from rank-constrained density matching weights (Fig. 5).

The weights produced by the network are used to reconstruct a deformation field as a linear combination of the rank-optimized deformation component fields. These reconstructed fields are directly comparable to the original deformation fields calculated in [11]. The geometric accuracy of the deformations is measured by the DICE coefficient to quantify the overlap between a deformed source mask region (such as tumor or full-lung regions) and the same mask region in the target volume (Fig. 6). Both source and target mask regions are from the manual expert segmentation of each respiratory phase volume used for radiotherapy treatment planning.

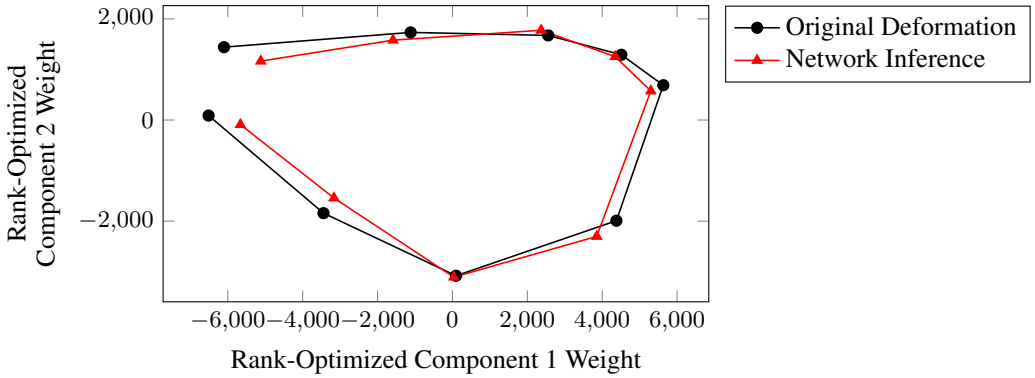

Figure 5: Weights of rank-optimized deformation components recovered by network on real patient data (red triangles) align well with original deformation weights for rank-optimized components (black circles). The respiratory cycle proceeds clockwise around the loop.

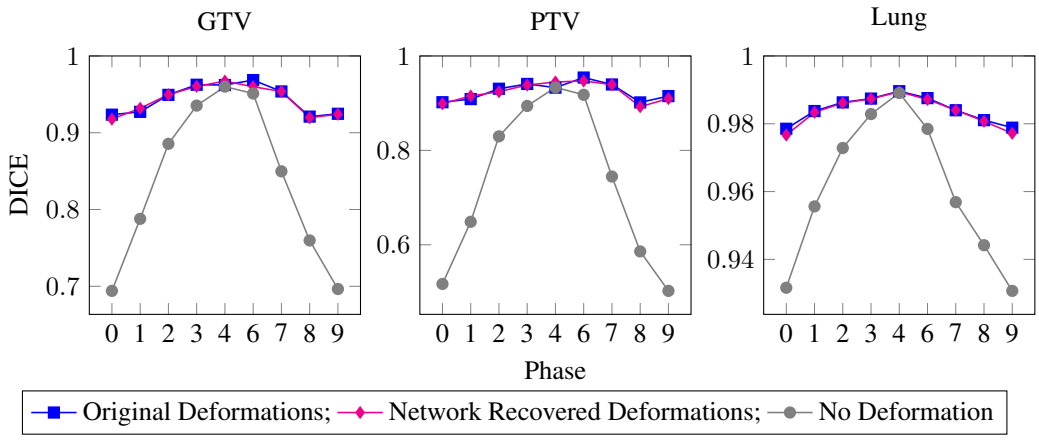

Figure 6: DICE coefficients for original rank-optimized deformation, deep-learned deformation from dynamic DRR images, and no deformation. Phase 5 is the full-exhale volume used as a reference for all other phases and therefore not directly reported. GTV - gross tumor volume; PTV - planned treatment volume; Lung - full lung volume.

5 Discussion

In this paper, we have shown that estimation of anatomic position in the presence of breathing variability is possible with the combination of (1) rank-constrained density motion estimation for determination of motion components and (2) deep learning for subsequent identification of the weights of motion components in real-time. While this framework is extensible to dimensions higher than 2, the rank-constrained nature of the deformation fields produces accurate results with only two dimensions. The level of accuracy is a trade-off against the curse of dimensionality in both the dataset size and computational cost; however, using rank-constrained deformation components increases the accuracy of resulting deformations with the same number of dimensions.

The speed and accuracy attained by this framework is suitable for inclusion in a tumor position monitoring component of a conformal radiation therapy system. Determination of 3D tumor location from noninvasive 2D radiographic projections via deep learning instead of variational optimization approaches for deformation determination provides real-time results that can be applied to conformal radiation therapy for lung tumors.

Acknowledgments

This work was partially supported through research funding from the National Institute of Health (R01CA169102).

References

[1] U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999-2014 Incidence and Mortality Web-based Report. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute, 2017.

[2] Yelin Suh, Walter Murray, and Paul J. Keall. IMRT Treatment Planning on 4D Geometries for the Era of Dynamic MLC Tracking. *Technology in Cancer Research & Treatment*, 13(6):505–515, dec 2014.

[3] Julian R. Molina, Ping Yang, Stephen D. Cassivi, Steven E. Schild, and Alex A. Adjei. Non-Small Cell Lung Cancer: Epidemiology, Risk Factors, Treatment, and Survivorship. *Mayo Clinic Proceedings*, 83(5):584–594, may 2008.

[4] S E Geneser, J D Hinkle, R M Kirby, B Wang, B Salter, and S Joshi. Quantifying variability in radiation dose due to respiratory-induced tumor motion. *Medical image analysis*, 15(4):640–9, aug 2011.

[5] Paul J. Keall, Sarang Joshi, S. Sastry Vedam, Jeffrey V. Siebers, Vijaykumar R. Kini, and Radhe Mohan. Four-dimensional radiotherapy planning for DMLC-based respiratory motion tracking. *Medical Physics*, 32(4):942–951, mar 2005.

[6] Amit Sawant, Paul Keall, Kim Butts Pauly, Marcus Alley, Shreyas Vasanawala, Billy W Loo, Jacob Hinkle, and Sarang Joshi. Investigating the feasibility of rapid MRI for image-guided motion management in lung cancer radiotherapy. *BioMed research international*, 2014:485067, jan 2014.

[7] P J Keall, V R Kini, S S Vedam, and R Mohan. Motion adaptive x-ray therapy: a feasibility study. *Physics in Medicine and Biology*, 46(1):1–10, jan 2001.

[8] Amit Sawant, Raghu Venkat, Vikram Srivastava, David Carlson, Sergey Povzner, Herb Cattell, and Paul Keall. Management of three-dimensional intrafraction motion through real-time DMLC tracking. *Medical Physics*, 35(5):2050–2061, 2008.

[9] Ruijiang Li, John H Lewis, Xun Jia, Tianyu Zhao, Weifeng Liu, Sara Wuenschel, James Lamb, Deshan Yang, Daniel A Low, and Steve B Jiang. On a PCA-based lung motion model. *Physics in Medicine and Biology*, 56(18):6009–6030, sep 2011.

[10] Pouya Sabouri, Markus Foote, Maida Ranjbar, Mohammad Tajdini, Sina Mossahebi, Sarang Joshi, and Amit Sawant. A Novel Method Using Surface Monitoring to Capture Breathing-Induced Cycle-To-Cycle Variations with 4DCT. In *59th Annual Meeting of The American Association of Physicists in Medicine*, Denver, CO, 2017.

[11] Markus Foote, Pouya Sabouri, Amit Sawant, and Sarang Joshi. Rank Constrained Diffeomorphic Density Motion Estimation for Respiratory Correlated Computed Tomography. In M.J. Cardoso and T Arbel, editors, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10551, pages 177–185. Springer Verlag, 2017.

[12] Martin Bauer, Sarang Joshi, and Klas Modin. Diffeomorphic Density Matching by Optimal Information Transport. *SIAM Journal on Imaging Sciences*, 8(3):1718–1751, jan 2015.

[13] Caleb Rottman, Martin Bauer, Klas Modin, and Sarang C. Joshi. Weighted Diffeomorphic Density Matching with Applications to Thoracic Image Registration. *5th MICCAI Workshop on Mathematical Foundations of Computational Anatomy (MFCA 2015)*, pages 1–12, 2015.

[14] Caleb Rottman, Ben Larson, Pouya Sabouri, Amit Sawant, and Sarang Joshi. Diffeomorphic Density Registration in Thoracic Computed Tomography. In Sebastien Ourselin, Leo Joskowicz, Mert R. Sabuncu, Gozde Unal, and William Wells, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part III*, volume 9902 of *Lecture Notes in Computer Science*, pages 46–53. Springer International Publishing, 2016.

[15] B. Khesin, J. Lenells, G. Misiołek, and S. C. Preston. Geometry of Diffeomorphism Groups, Complete integrability and Geometric statistics. *Geometric and Functional Analysis*, 23(1):334–366, feb 2013.

[16] Klas Modin. Generalized Hunter–Saxton Equations, Optimal Information Transport, and Factorization of Diffeomorphisms. *Journal of Geometric Analysis*, 25(2):1306–1334, apr 2015.

[17] Caleb Rottman, Lance McBride, Arvidas Cheryauka, Ross Whitaker, and Sarang Joshi. Mobile C-arm 3D Reconstruction in the Presence of Uncertain Geometry. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9350, pages 692–699. 2015.

[18] J. Preston, J. Hinkle, N. Singh, C. Rottman, M. Foote, and S. Joshi. PyCA: Python for Computational Anatomy. https://bitbucket.org/scicompanat/pyca.

[19] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269. IEEE, jul 2017.

[20] Geoff Pleiss, Danlu Chen, Gao Huang, Tongcheng Li, Laurens van der Maaten, and Kilian Q. Weinberger. Memory-Efficient Implementation of DenseNets. jul 2017.

[21] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in PyTorch. In *NIPS-W*, 2017.