HHS Public Access

Author manuscript

*Rep U S*. Author manuscript; available in PMC 2024 June 29.

Published in final edited form as:

*Rep U S*. 2023 October ; 2023: 6593–6600. doi:10.1109/iros55552.2023.10342115.

# Landmark Based Bronchoscope Localization for Needle Insertion Under Respiratory Deformation

**Inbar Fried**[1,2], **Janine Hoelscher**[1], **Jason A. Akulian**[3], **Stephen Pizer**[1], **Ron Alterovitz**[1]

[1]I. Fried, J. Hoelscher, S. Pizer, and R. Alterovitz are with the Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

[2]I. Fried is also with the Medical Scientist Training Program, University of North Carolina School of Medicine, Chapel Hill, NC, 27599, USA.

[3]J. A. Akulian is with the Division of Pulmonary Diseases and Critical Care Medicine at the University of North Carolina at Chapel Hill, NC 27599, USA.

## Abstract

Bronchoscopy is currently the least invasive method for definitively diagnosing lung cancer, which kills more people in the United States than any other form of cancer. Successfully diagnosing suspicious lung nodules requires accurate localization of the bronchoscope relative to a planned biopsy site in the airways. This task is challenging because the lung deforms intraoperatively due to respiratory motion, the airways lack photometric features, and the anatomy's appearance is repetitive. In this paper, we introduce a real-time camera-based method for accurately localizing a bronchoscope with respect to a planned needle insertion pose. Our approach uses deep learning and accounts for deformations and overcomes limitations of global pose estimation by estimating pose relative to anatomical landmarks. Specifically, our learned model considers airway bifurcations along the airway wall as landmarks because they are distinct geometric features that do not vary significantly with respiratory motion. We evaluate our method in a simulated dataset of lungs undergoing respiratory motion. The results show that our method generalizes across patients and localizes the bronchoscope with accuracy sufficient to access the smallest clinically-relevant nodules across all levels of respiratory deformation, even in challenging distal airways. Our method could enable physicians to perform more accurate biopsies and serve as a key building block toward accurate autonomous robotic bronchoscopy.

## I. Introduction

Lung cancer is the leading cause of cancer related deaths in the United States, contributing to over 120,000 deaths annually. Early diagnosis is critical in improving outcomes for patients, however, just 25% of patients get diagnosed at an early and favorable stage [1]. Currently, the least invasive method for definitively diagnosing suspicious lung nodules as cancer is bronchoscopy, a procedure during which a physician inserts a bronchoscope through the patient's mouth, navigates through the airways, and inserts a needle into the lung to biopsy a nodule. Since clinically relevant nodules can be as small as 8mm in diameter

ifried01@cs.unc.edu .

[2], accurate localization of the bronchoscope with respect to a planned needle insertion site in the airways is critical for the procedure's success. Despite the availability of commercial systems that aid with peripheral lung navigation, a large proportion of bronchoscopies fail to result in a diagnosis in part due to poor intraoperative localization [3], [4]. Given the importance of bronchoscopy as a tool for early definitive diagnosis of lung nodules, improving bronchoscope localization is critical.

Localizing a bronchoscope during a procedure is challenging for several reasons, including the repetitive appearance of the airways, the airway lumen's lack of distinguishing features, and the deformable nature of the environment. Traditionally, physicians navigate the bronchoscope by mentally registering the 2D video frames from the bronchoscope's built-in camera to a virtual 3D reconstruction of the patient's airways from a preoperative CT scan (see Fig. 1). To help the physician with this challenging registration task, several commercial systems have been developed that use additional sensors such as electromagnetic (EM) sensors or fiber Bragg gratings. However, these approaches have difficulty tracking when there is a discrepancy between the anatomical map created from a preoperative CT scan and the intraoperative patient anatomy. This problem, referred to as CT-to-body divergence, is prevalent in bronchoscopy because of the constant respiratory motion of the lungs, and it often worsens throughout the procedure due to physiologic changes in the tissue resulting in any initial registration becoming outdated [5]. Using the built-in camera for localization can overcome this problem from global respiratory motion since the camera's internal view of the airways remains unchanged (i.e., the bronchoscope is moving along with the lungs). However, the repetitive appearance of the airways and the airway lumen's lack of texture make visual odometry or global pose estimation challenging. Addressing these challenges in intraoperative bronchoscope localization in the vicinity of a planned needle insertion would improve biopsy accuracy and enable accurate autonomous robotic bronchoscopy.

To overcome these challenges, we introduce a camera-based method for localizing a bronchoscope with respect to a planned needle insertion pose using anatomical landmarks in the airways and deep learning. We use the bronchial tree bifurcations on the airway wall as our landmarks as they do not vary significantly with both global respiratory motion and local deformation, and they are distinct geometric features in an otherwise feature-sparse environment. By localizing the bronchoscope in the bifurcation's local coordinate frame, our method is not affected by the ambiguity of the anatomy that makes global pose estimation challenging or from accumulating errors that burden relative pose estimation. Furthermore, by not being directly tied to a specific CT scan's global coordinate frame, our method can generalize across patients, which is important in the medical domain as data is often limited. To the best of our knowledge, this is the first method for bronchoscope localization that has shown such interpatient generalizability.

To localize the bronchoscope, we trained a deep neural network to estimate the bronchoscope's pose in reference to bifurcations on the airway wall whose relative transformation to the planned needle insertion site is known from the preoperative CT scan. The network takes as input a single image frame from the bronchoscope's camera and outputs the relative transformation to the nearest visible bifurcation in real-time.

We show that our method is tolerant to physiologic respiratory motion by evaluating our method in simulations of human lungs deforming based on patient-derived respiratory data. We also show that our method performs well in distal airways (see Fig. 2) which have been notoriously challenging for existing approaches because of the increased deformation in these regions. Our method achieves average translation and orientation errors under 3mm and 10 °, respectively, across all airway generations. By accurately estimating the bronchoscope pose at the time of needle insertion, our method could enable physicians to perform more accurate biopsies using current bronchoscopes and could also be integrated with a robotic system to enable accurate autonomous bronchoscopy.

## II. Related Work

### A. Bronchoscope Localization

Bronchoscope localization methods can be broadly classified into three categories.

**1) Sensor Based Methods:** Sensor based localization methods leverage a sensor other than the bronchoscope's built-in camera to localize the bronchoscope in the patient's anatomy. A common approach is electromagnetic (EM) sensor-based localization, which is used by several commercial systems [6]. Another approach is fiber optic shape sensing technology, which has shown improved localization performance [7]. While fiber optic sensors are not affected by noise from metal objects like EM sensors are, both of these approaches have poor tracking accuracy when there is a discrepancy between the intraoperative patient anatomy and the preoperative CT scan (i.e., CT-to-body divergence). To overcome this issue, the use of cone beam CT for intraoperative imaging during needle deployment has been proposed [8], but at the expense of additional radiation exposure to the medical team and patient.

**2) Vision Based Methods:** The bronchoscope's camera moves with the airways when the lung deforms during respiration, which has motivated research into overcoming the challenge of CT-to-body divergence by leveraging the optical information from the camera for localization. A common approach has been to localize the bronchoscope by finding the best match for the real bronchoscopic view in a large collection of virtual bronchoscopic views generated using a patient's preoperative CT scan [9], [10]. Variations on this approach have included performing comparisons in eigenspace [11], identifying regions of interest in the image [12], and using geometric properties of the environment such as depth and surface normals [13], [14]. Recently, several groups have leveraged deep neural networks to extract higher quality depth maps [15] and also to perform style transfer to make the real image share appearance properties with the virtual domain [16], [17]. The challenge with this general approach has been that the repetitive appearance of the airways creates a global ambiguity whereby multiple distinct regions in the airways look nearly identical. To address some of these challenges, several groups have proposed methods that navigate the bronchoscope by tracking lumen centers [18], counting branching points [19], and extracting structural lumen characteristics [20]. An alternative approach has been to use visual odometry (VO), both in the context of incremental image alignment [21], established SLAM algorithms [22], incremental point-cloud registration [23], and using deep neural

networks for direct relative pose regression [24], [25]. The lack of texture in the airways and variable speed of the bronchoscope limits the success of classic feature-tracking approaches like SLAM and VO, while direct relative pose prediction also suffers from accumulating errors. A recent work proposed localizing the bronchoscope by first predicting which airway is visible and then regressing the camera's pose relative to the furthest point in the airway [26], but the work only reports results when the model correctly identified the airway, ignoring the effect of global ambiguity, and excluded evaluation cases where there was visible CT-to-body divergence. While many of these works have shown great localization performance in proximal airways, very few evaluated their methods in distal airways and those that did have not shown similarly good performance. Distal airways represent the greatest challenge in clinical bronchoscope localization because of the increased respiratory deformation in these regions.

**3) Hybrid Methods:** Hybrid methods combine information from both sensor-based and vision-based techniques to improve the overall localization of the bronchoscope. These approaches typically use the EM sensor to get a rough estimate of the bronchoscope's location and then refine the prediction by optimizing the similarity between the bronchoscope's intraoperative view with a dataset of virtual bronchoscopy images [27], [28], [29], [30], [31]. These works have shown promising results, but their vision-based solutions still suffer from the same drawbacks as the vision-only methods described above.

The method we propose can be classified as either a hybrid solution or a vision-only based solution in that it can be integrated with an existing sensor-based system for navigation or be used independently with the physician providing the relevant navigation input prior to needle insertion.

### B. Toward Autonomous Robotic Bronchoscopy

While accurate localization in the airways is challenging, it is a core component for enabling autonomous robotic bronchoscopy. Several works have developed such systems using various types of localization techniques mentioned above [26], [32], [33]. Additionally, the feasibility of leveraging autonomous robots for automating nodule targeting with a needle has been shown in animal models [34], [35], for which localization prior to needle insertion is critical [36].

## III. Problem Definition

We consider the entire bronchoscopic procedure in two stages when describing our method: a preoperative stage and an intraoperative stage. The workflow of our method according to these stages is shown in Fig. 3.

In the preoperative stage prior to a bronchoscopic procedure, a patient undergoes a CT scan which generates a static 3D image of the anatomy. Physicians use this image to extract 3D segmentations of the airways and lung nodule, and to virtually plan their procedure. Let S represent the volumetric CT scan image that is defined in the coordinate frame of the CT scanner, $C^{CT}$. Let $A$ be the 3D segmented model of the airways. Let $\mathbf{p} \in SE(3)$, also defined in $C^{CT}$, represent the planned needle insertion pose (position and orientation) along

the airway wall to reach the nodule. This pose can be determined manually by a physician or using a motion planning algorithm [37], [36]. Although $\mathbf{p}$ is defined in $SE(3)$, since the bronchoscope and needle are coaxial, rotations about the axial axis of the bronchoscope with respect to $\mathbf{p}$ do not affect the needle's final insertion position. Finally, let $L$ be a set of poses, $\mathbf{l}_i \in SE(3)$, that correspond to the bifurcations on the airway wall. This set of landmark poses is either provided by the physician or automatically extracted from $S$.

In the intraoperative stage, which occurs hours to days after the preoperative stage, the physician uses the built-in camera and any additional available sensor information (e.g., EM) to navigate the bronchoscope through the airways to a planned needle insertion pose. During the procedure, the lungs are undergoing physiologic respiratory deformation. Let $\mathbf{b} \in SE(3)$ represent the pose of the bronchoscope in world coordinate frame, $C^W$, and $I$ represent a 2D RGB image from the bronchoscope's built-in camera. We assume that a coarse mapping exists from $C^W$ to $C^{CT}$ such that the physician is able to identify which bifurcation, $\mathbf{l}_i$, they see in $I$. This mapping is determined either from an additional sensor (e.g., EM) or directly by the physician. To get an accurate localization of the bronchoscope with respect to $\mathbf{p}$, we need to refine this localization while overcoming the effect of respiratory motion. To localize the bronchoscope in the real-world deformed lung, we need to estimate the bronchoscope's pose relative to $\mathbf{p}$ given $I$. In order to achieve real-time performance, we need to perform this estimation in under 33 milliseconds to provide localization at the bronchoscope camera's typical frame rate of thirty frames per second. Let ${}_{\mathbf{b}}\widehat{\mathbf{T}}_{\mathbf{p}} \in SE(3)$ represent our prediction for the true transformation, ${}_{\mathbf{b}}\mathbf{T}_{\mathbf{p}} \in SE(3)$, from $\mathbf{b}$ to $\mathbf{p}$.

We evaluate our prediction's accuracy using the equation:

$$\mathbf{E} = {}_{\mathbf{b}}\widehat{\mathbf{T}}_{\mathbf{p}}^{-1} \cdot {}_{\mathbf{b}}\mathbf{T}_{\mathbf{p}} \tag{1}$$

where $\mathbf{E}$ is the relative transformation between the true and predicted transformations [38]. When ${}_{\mathbf{b}}\widehat{\mathbf{T}}_{\mathbf{p}}$ and ${}_{\mathbf{b}}\mathbf{T}_{\mathbf{p}}$ are equal, $\mathbf{E}$ is the identity matrix. We calculate the translational and rotational error using the following two equations [39]:

$$\mathbf{E}_{\text{trans}} = \|\text{trans}(\mathbf{E})\|_2 \tag{2}$$

$$\mathbf{E}_{\text{rot}} = \arccos\left(\frac{\text{trace}\left(\text{rot}(\mathbf{E})^{-1} \cdot \text{rot}(\mathbf{E})\right)}{2}\right), \tag{3}$$

where trans() and rot() are functions that extract the translational and rotational components of **E**, respectively. Eq. 3 is equivalent to calculating the angle from the axis-angle representation of rot(**E**).

We formalize the acceptable translation and orientation error between the planned and achieved needle insertion as done in prior work [23]. Specifically, we consider a lung nodule with radius $r$ and a straight needle with maximum insertion length $h$. Fig. 4 illustrates the different error scenarios in a 2D representation of a needle biopsy. When there is only translation error, the translation error tolerance, $e_t$, is equal to the radius of the nodule, $r$ (Fig. 4b). When there is only orientation error, the orientation error tolerance, $e_r$, is $\arctan(r/h)$ (Fig. 4c). When there is both translation and orientation error, the error tolerance values depend on one another and on their relative directions (Fig. 4d). Therefore, we can estimate the ranges of acceptable translation and orientation errors as $[-r, r]$ and $[-\arctan(2*r/h), \arctan(2*r/h)]$, respectively. In this paper, we consider $r = 4$ mm and $h = 25$ mm, which represents the smallest clinically-relevant nodules and the capability of current clinical tools [2]. Based on these values, we calculate $e_t = 4$ mm and $e_r = 17$ °.

## IV. Method

In this section, we describe our proposed method and how it integrates in a typical clinical workflow for a patient, starting with preoperative CT image acquisition and ending in bronchoscope localization relative to a planned needle insertion pose.

### A. Data Acquisition

Using $S$, we manually segmented $A$ and manually determined the 6 DoF pose, $\mathbf{l}_i \in L$, of every bifurcation on the airway wall. Using a simulator we developed in 3D Slicer [40], we generated virtual bronchoscopies in $A$. For each camera frame throughout the bronchoscopy, we recorded the camera's pose with respect to the closest visible bifurcation in its view. The result of the simulation is a dataset of $(I, {}_{\mathbf{b}}\mathbf{T}_{\mathbf{l}_i})$ pairs. Fig. 5 shows the simulation environment including a sample airway segmentation, its tagged landmark poses, and the simulated bronchoscopy view.

To increase the breadth of our dataset, we augmented the data in several ways. To increase the diversity of camera views, for each camera frame, we generated additional $(I, {}_{\mathbf{b}}\mathbf{T}_{\mathbf{l}_i})$ pairs by transforming the bronchoscope's pose by some translational and rotational noise, $\mathbf{b}' = \mathbf{b} \cdot \mathbf{T}_{\text{noise}}$. The Euler angles for the rotational noise were sampled from a uniform distribution over [−20,20] degrees, and the translational noise was sampled from a Gaussian distribution $\mathcal{N}$ (0, 2) mm for the camera's x and y axes and from a uniform distribution over [−25,0] mm for the z axis. Since every bronchoscopy we simulate starts at the trachea, the frequency with which landmarks are seen decreases with the depth of the landmark in the airways. To increase the representation of distal bifurcations in our dataset, we created additional $(I, {}_{\mathbf{b}}\mathbf{T}_{\mathbf{l}_i})$ pairs by transforming the virtual camera randomly with respect to the pose of each bifurcation. To encourage the model to be invariant to color features

and instead focus on geometric properties of the bifurcations, we augmented our data by randomly coloring triangles in the mesh representation of $A$ within our simulator.

We perform several filtering steps to ensure that the data we generate is informative. We filter images where 30% or more of the image is purely white. To avoid images where the nearest landmark is on the border of the camera's view and not in its main area of focus, as can happen when the bronchoscope travels past a landmark, we filter all images where the landmark appears in the outer 90% of the image. Additionally, since rotations about the axial axis of the bronchoscope do not affect the needle's position after its insertion, we remove ambiguity about this axis from the dataset by arbitrarily setting all landmark poses to one hemisphere. To avoid ambiguity on the edges of the hemisphere, we remove all images within ten degrees of the borders. We consider a landmark as the true label of a given camera frame when it is the nearest landmark in the image and within 25mm from the camera's position. The camera is always within 25mm of a landmark other than when it is at the top of the trachea or the top of the left main bronchus. Since physicians do not pierce the airways at these proximal generations, excluding these images from our dataset does not affect the clinical application our method is addressing.

## B. Learning Landmark Based Localization

Since $\mathbf{p}$ is defined in $C^{CT}$ and is not visible in $I$, which is defined in $C^{W}$, we cannot predict ${}_{\mathbf{b}}\mathbf{T}_{\mathbf{p}}$ directly from $I$. However, the bifurcations, $\mathbf{l}_i$, are visible in $I$. Given $I$ with $\mathbf{l}_i$ as the nearest visible landmark, we need to estimate ${}_{\mathbf{b}}\mathbf{T}_{\mathbf{l_i}}$, the relative transformation from $\mathbf{b}$ to $\mathbf{l}_i$ in $\mathbf{b}$'s frame.

To predict ${}_{\mathbf{b}}\mathbf{T}_{\mathbf{l_i}}$, we learn a function $\mathscr{F}(I) = [\hat{\mathbf{t}}, \hat{\mathbf{q}}]$, where $\hat{\mathbf{t}} \in \mathbb{R}^3$ represents a prediction for the true relative translation, $\mathbf{t}$, and $\hat{\mathbf{q}} \in \mathbb{S}^2$ represents a prediction for the true relative orientation, $\mathbf{q}$. $\mathbf{q}$ is composed of the roll and pitch angles, $\beta$ and $\gamma$, respectively. We do not learn $\alpha$, the relative angle about the axial axis of the bronchoscope, since changes about this axis do not affect the final position of the needle. This reduction in the dimensionality of the output space reduces the number of variables that our method needs to regress compared to other orientation representations such as quaternions $\left(\mathbb{R}^4\right)$, log quaternions $\left(\mathbb{R}^3\right)$, or axis-angle $\left(\mathbb{R}^3\right)$. We represent $\mathscr{F}$ as a deep neural network. We use the following loss function to train our network:

$$L = \|\mathbf{t} - \hat{\mathbf{t}}\|_1 + \mu\|\mathbf{q} - \hat{\mathbf{q}}\|_1 \tag{4}$$

where $\mu$ is a hyperparameter that scales the orientation loss to help jointly learn translation and orientation. We train our model in a supervised fashion using the ground-truth $\left(I, {}_{\mathbf{b}}\mathbf{T}_{\mathbf{l_i}}\right)$ pairs we generated. The result of training is a learned model that, given $I$, predicts ${}_{\mathbf{b}}\mathbf{T}_{\mathbf{l_i}}$ for the nearest visible landmark, $\mathbf{l}_i$, in $I$.

Our model is structured as an encoder-decoder network. We use a ResNet [41] model with 50 layers that was pretrained on ImageNet [42] as our encoder and a series of four convolutional layers as our decoder, similar to the pose decoder presented in [43]. We implemented our method in Pytorch and trained the model for 20 epochs using Adam [44] with a weight decay of $1e^{-3}$. We set our learning rate to $1e^{-4}$ for the first 6 epochs and then decay it by $1e^{-1}$ every 6 epochs. We used a batch size of 32 and an input image resolution of 480 × 544. We set the loss scaling term, $\mu$, to 10. Training was done using an NVIDIA Tesla V100 GPU and takes approximately 24 hours.

### C. Intraoperative Localization

Using our trained model, $\mathscr{F}$, we can recover the pose of the bronchoscope relative to the planned needle insertion pose. Given $I$, we use $\mathscr{F}$ to predict the transformation ${}_{\mathbf{b}}\mathbf{T}_{\mathbf{l}_i}$. Since both $\mathbf{p}$ and the landmarks, $\mathbf{l}_i \in L$, are defined in $C^{CT}$, we can calculate the relative transformation from $\mathbf{l}_i$ to $\mathbf{p}$ in $\mathbf{l}_i$'s frame, ${}_{\mathbf{b}}\mathbf{T}_{\mathbf{l}_i}$, using $S$. Since the bifurcations do not vary significantly under respiratory motion, the transformation ${}_{\mathbf{b}}\mathbf{T}_{\mathbf{l}_i}$ does not vary between $C^{CT}$ and $C^{W}$. We then calculate $\mathbf{b}$ relative to $\mathbf{p}$ as ${}_{\mathbf{b}}\mathbf{T}_{\mathbf{p}} = {}_{\mathbf{b}}\widehat{\mathbf{T}}_{\mathbf{l}_i} \cdot {}_{\mathbf{b}}\mathbf{T}_{\mathbf{l}_i}$ (see Fig. 6). Since our prediction of ${}_{\mathbf{b}}\mathbf{T}_{\mathbf{p}}$ is done with respect to the local coordinate frame of $\mathbf{l}_i$, our method is tolerant to deformation from respiratory motion.

## V. Evaluation

We evaluated our approach on 3 human patient cases for which chest CT scans were publicly available in the EXACT'09 dataset [45]. For each patient, we segmented $A$ from the expiratory CT scan using 3D Slicer and manually labeled $L$. We use the airways at expiration since it has been shown that the lung closely resembles this state intraoperatively [46], [47]. Table I shows characteristics of the data. We applied our method and trained a model $\mathscr{F}$ on each patient individually from data spanning all airway generations.

### Respiratory Deformation Evaluation Data

During a bronchoscopy procedure, patients are mechanically ventilated and the lungs deform similarly to a normal respiratory cycle. This constant motion of the anatomy causes CT-to-body divergence and presents a challenge for localization. To evaluate the performance of our model under physiologic respiratory motion, we created an evaluation dataset by deforming $A$ from the expiratory state over varying levels of patient-derived respiratory deformation characteristics. To estimate a patient's respiratory deformation, we performed deformable image registration [48] between the CT scan at expiration with a CT scan for the same patient at inspiration. These CT scan pairs are available in EXACT'09. The result of this deformable registration is a 3D deformation field that we interpolated to generate a series of three deformed segmentations corresponding to increasing levels of deformed airways states between expiration and inspiration. We interpolated the deformation field to match reported displacement values of the lungs throughout breathing [8], [49], [50], producing Low, Medium, and High deformation states. We then generated our evaluation data using our simulator on the airways in the Low, Medium, and High deformation states.

Fig. 7 shows the respiratory motion for one patient in our dataset along with the deformation displacement by lung region across all patients.

We report results for our method across the three levels of deformation, which the models never saw or used during training. To show that our method can accurately localize the bronchoscope in peripheral airways, we report results for all distal airway generations. In all experiments, we use Eq. 2 and Eq. 3 to evaluate our model. Since we do not regress $\alpha$, we ignore its effect when reporting the orientation error. Our method localizes the bronchoscope in under 20 milliseconds per image, achieving real-time performance.

## B. Tolerance to Respiratory Motion

Fig. 8 shows our model's results by airway generation and level of deformation averaged across the three patients. Figs. 8(a) and (b) show the model's error in translation, $\mathbf{E}_{\text{trans}}$, and orientation, $\mathbf{E}_{\text{rot}}$, respectively. These results show that the model is able to accurately localize the bronchoscope with respect to the nearest landmark in the anatomy, achieving overall average (± standard deviation) translational errors of 2.97mm ± 2.20mm (median 2.36mm) and average rotational errors of 9.57 ° ± 8.60 ° (median 7.59 °), which are below the strictest clinical requirements we calculated for $e_t$ and $e_r$ in Sec. III. These results also show that the model's performance is invariant to depth in the airways, accurately localizing the camera in distal airway generations that are notoriously challenging for accurate localization because of increased levels of deformation. We also see that the method's absolute performance does not vary significantly with increasing levels of deformation, indicating that the model is able to overcome the effects of respiratory deformation when localizing the bronchoscope.

## C. Generalization Across Patients

While in a typical clinical setting every patient will have a preoperative CT scan, being able to generalize across patients is beneficial in terms of leveraging limited data in the medical domain and in developing robust learning-based algorithms. To evaluate the ability of our method to generalize across patients, we evaluated our trained models on simulated bronchoscopies of a separate patient from a different dataset. We used a segmentation of a human CT scan from [51], a publicly available dataset of medical environments for robot motion planning, and we manually labeled the bifurcations. Fig. 9 shows the average performance of the three models that were trained on the three patients from EXACT'09 and evaluated on this new patient. The model trained on EXACT09–18 generalizes slightly less well than the other two models likely as a result of the fewer bifurcations in that patient which limits the variety of landmarks the model is exposed to during training. Overall, the results show that our method generalizes across patients, achieving similar performance to the performance in the experiments above under respiratory deformation. This suggests that the models learn generalizable features for accurate localization.

## D. Tracking a Bronchoscope Over Time

To evaluate our method as it would be used in a bronchoscopy procedure, we apply our method to simulated bronchoscopy video sequences where the camera approaches a planned needle insertion pose. We evaluate our model's ability to accurately localize the bronchoscope throughout the sequence of frames. We arbitrarily selected four needle

insertion poses in deformed airways from our evaluation dataset and recorded videos of the bronchoscope traveling towards the planned needle insertion pose. The insertion poses span airway generations four through nine. The bronchoscope starts 15 to 20mm from the planned insertion pose and the videos range from 209 to 421 frames. We randomly added translational and rotational noise with 25% probability per frame to the camera throughout the sequence to simulate unexpected motion intraoperatively. The average translation and orientation errors across the 4 videos was 1.8 mm ± 0.9 mm and 3.7° ± 1.4°, respectively. Fig. 10 shows the localization results of our method on two sample sequences. These results show that our method is able to accurately localize the bronchoscope along a bronchoscopic sequence. Furthermore, since our method considers every image independently, an error from a poor prediction on one challenging image does not propagate to future predictions.

## VI. Conclusion

In this paper, we developed a real-time camera-based method for accurately localizing a bronchoscope intraoperatively with respect to a planned needle insertion pose using the airway bifurcations along the airway wall as landmarks and leveraging deep learning. We show that since the bifurcations do not vary significantly with respiratory motion, our method overcomes localization challenges from global and local respiratory deformation. We show that by localizing the bronchoscope in the local coordinate frame of the bifurcations, our method is unaffected by the repetitive appearance of the airways that afflicts global pose estimation and that the geometric features of the bifurcations enable our method to learn features that generalize across varying levels of intra-patient deformation and inter-patient variation despite the lack of texture on the airway wall. Our method is able to localize the bronchoscope in distal airways with levels of accuracy required for the smallest clinically-relevant lung nodules. In the future, we hope to evaluate our method on real bronchoscopy images and address challenges from local deformations caused by the bronchoscope. We also hope to analyze the relationship between the number of tagged landmarks and our method's performance. Our hope is that our method can be integrated with existing bronchoscope navigation systems to increase physicians' nodule targeting accuracy and also serve as a core component in an autonomous robotic bronchoscopic needle insertion system.

## Acknowledgments

This research was supported by the U.S. National Institutes of Health (NIH) under awards F30CA265234 and R01EB024864 and the U.S. National Science Foundation (NSF) under award 2008475.

## References

[1]. Siegel RL, Miller KD, Wagle NS, and Jemal A, "Cancer statistics, 2023," CA: a cancer journal for clinicians, vol. 73, no. 1, pp. 17–48, 2023, publisher: Wiley Online Library. [PubMed: 36633525]

[2]. Bueno J, Landeras L, and Chung JH, "Updated Fleischner Society guidelines for managing incidental pulmonary nodules: Common questions and challenging scenarios," Radiographics, vol. 38, no. 5, pp. 1337–1350, 2018, publisher: Radiological Society of North America. [PubMed: 30207935]

[3]. Silvestri GA, et al. , "An evaluation of diagnostic yield from bronchoscopy," Chest, vol. 157, no. 6, pp. 1656–1664, June 2020. [PubMed: 31978428]

[4]. Yarmus L, et al. , “A prospective randomized comparative study of three guided bronchoscopic approaches for investigating pulmonary nodules,” Chest, vol. 157, no. 3, pp. 694–701, Mar. 2020. [PubMed: 31678307]

[5]. Pritchett MA, Bhadra K, Calcutt M, and Folch E, “Virtual or reality: divergence between preprocedural computed tomography scans and lung anatomy during guided bronchoscopy,” Journal of Thoracic Disease, vol. 12, no. 4, pp. 1595–1611, Apr. 2020. [PubMed: 32395297]

[6]. Burks AC and Akulian J, “Bronchoscopic diagnostic procedures available to the pulmonologist,” Clinics in Chest Medicine, vol. 41, no. 1, pp. 129–144, Mar. 2020. [PubMed: 32008625]

[7]. Fielding D, et al. , “First human use of a new robotic-assisted fiber optic sensing navigation system for small peripheral pulmonary nodules,” Respiration, vol. 98, no. 2, pp. 142–150, 2019. [PubMed: 31352444]

[8]. Reisenauer J, Duke JD, Kern R, Fernandez-Bussy S, and Edell E, “Combining shape-sensing robotic bronchoscopy with mobile three-dimensional imaging to verify tool-in-lesion and overcome divergence: A pilot study,” Mayo Clinic Proceedings: Innovations, Quality & Outcomes, vol. 6, no. 3, pp. 177–185, June 2022.

[9]. Bricault I, Ferretti G, and Cinquin P, “Registration of real and CT-derived virtual bronchoscopic images to assist transbronchial biopsy,” IEEE Trans. Med. Imaging, vol. 17, no. 5, pp. 703–714, Oct. 1998. [PubMed: 9874294]

[10]. Helferty J, Sherbondy A, Kiraly A, and Higgins W, “Computer-based system for the virtual-endoscopic guidance of bronchoscopy,” Computer Vision and Image Understanding, vol. 108, no. 1–2, pp. 171–187, Oct. 2007. [PubMed: 18978928]

[11]. Shinohara R, et al., “Branch identification method for CT-guided bronchoscopy based on eigenspace image matching between real and virtual bronchoscopic images,” in Medical Imaging, vol. 6143. SPIE, 2006, pp. 385–396.

[12]. Deguchi D, et al. , “Selective image similarity measure for bronchoscope tracking based on image registration,” Medical Image Analysis, vol. 13, no. 4, pp. 621–633, Aug. 2009. [PubMed: 19592291]

[13]. Deligianni F, Chung A, and Yang G-Z, “Patient-specific bronchoscope simulation with pq-space-based 2D/3D registration,” Computer Aided Surgery, vol. 9, no. 5, pp. 215–226, 2004. [PubMed: 16192063]

[14]. Shen M, Giannarou S, and Yang G-Z, “Robust camera localisation with depth reconstruction for bronchoscopic navigation,” Int. J. CARS, vol. 10, no. 6, pp. 801–813, June 2015.

[15]. Shen M, Gu Y, Liu N, and Yang G-Z, “Context-aware depth and pose estimation for bronchoscopic navigation,” IEEE Robot. Autom. Lett, vol. 4, no. 2, pp. 732–739, Apr. 2019.

[16]. Zhao C, Shen M, Sun L, and Yang G-Z, “Generative Localization With Uncertainty Estimation Through Video-CT Data for Bronchoscopic Biopsy,” IEEE Robot. Autom. Lett, vol. 5, no. 1, pp. 258–265, Jan. 2020.

[17]. Wang C, et al. , “Anatomy aware-based 2.5D bronchoscope tracking for image-guided bronchoscopic navigation,” Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, pp. 1–8, 2022.

[18]. Sanchez C, Esteban Lansaque A, Borŕ as A, Diez-Ferrer M,` Rosell A, and Gil D, “Towards a videobronchoscopy localization system from airway centre tracking,” in Proc. Int. Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP), 2017, pp. 352–359.

[19]. Wang C, et al. , “Depth-based branching level estimation for bronchoscopic navigation,” Int. J. CARS, vol. 16, no. 10, pp. 1795–1804, Oct. 2021.

[20]. Shen M, Giannarou S, Shah PL, and Yang G-Z, “BRANCH: Bifurcation recognition for airway navigation based on structural characteristics,” ” in MICCAI 2017.

[21]. “Interactive CT-video registration for the continuous.”

[22]. Wang C, et al. , “A visual SLAM-based bronchoscope tracking scheme for bronchoscopic navigation,” Int. Journal CARS, vol. 15, no. 10, pp. 1619–1630, Oct. 2020.

[23]. Banach A, King F, Masaki F, Tsukada H, and Hata N, “Visually navigated bronchoscopy using three cycle-consistent generative adversarial network for depth estimation,” Medical Image Analysis, vol. 73, p. 102164, Oct. 2021.

[24]. Sganga J, Eng D, Graetzel C, and Camarillo D, “Offsetnet: Deep learning for localization in the lung using rendered images,” in Proc. Int. Conf. on Robotics and Automation (ICRA). IEEE, 2019, pp. 5046–5052.

[25]. Borrego-Carazo J, Sanchez C, Castells-Rufas D, Carrabina J, and Gil D, “BronchoPose: an analysis of data and model configuration for vision-based bronchoscopy pose estimation,” Computer Methods and Programs in Biomedicine, vol. 228, p. 107241, 2023.

[26]. Sganga J, Eng D, Graetzel C, and Camarillo DB, “Autonomous driving in the lung using deep learning for localization,” arXiv preprint arXiv:1907.08136, 2019.

[27]. Mori K, et al. , “Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration,” in MICCAI. Springer, 2005, pp. 543–550.

[28]. Soper T, Haynor D, Glenny R, and Seibel E, “*In Vivo* validation of a hybrid tracking system for navigation of an ultrathin bronchoscope within peripheral airways,” IEEE Trans. Biomed. Eng, vol. 57, no. 3, pp. 736–745, Mar. 2010. [PubMed: 19846362]

[29]. Reichl T, Luo X, Menzel M, Hautmann H, Mori K, and Navab N, “Deformable registration of bronchoscopic video sequences to ct volumes with guaranteed smooth output,” in MICCAI. Springer, 2011, pp. 17–24.

[30]. Luo X,´ M. Feuerstein, T. Kitasaka, and K. Mori, “Robust bronchoscope motion tracking using sequential Monte Carlo methods in navigated bronchoscopy: dynamic phantom and patient validation,” Int. J. CARS, vol. 7, no. 3, pp. 371–387, May 2012.

[31]. Deligianni F, Chung AJ, and Yang G-Z, “Nonrigid 2D/3D registration for patient specific bronchoscopy simulation with statistical shape modeling: Phantom validation,” IEEE Trans. Med. Imag, vol. 25, no. 11, pp. 1462–1471, 2006.

[32]. Zou Y, Guan B, Zhao J, Wang S, Sun X, and Li J, “Robotic-assisted automatic orientation and insertion for bronchoscopy based on image guidance,” IEEE Trans. Med. Robot. Bionics, vol. 4, no. 3, pp. 588–598, Aug. 2022.

[33]. Jaeger HA, et al. , “Automated catheter navigation with electromagnetic image guidance,” IEEE Trans. Biomed. Eng, vol. 64, no. 8, pp. 1972–1979, Aug. 2017. [PubMed: 28362578]

[34]. Fried I, et al. , “Design considerations for a steerable needle robot to maximize reachable lung volume,” Proc. Int. Conf. on Robotics and Automation (ICRA), pp. 1418–1425, 2021.

[35]. Kuntz A, et al. , “Autonomous Medical Needle Steering In Vivo,” arXiv preprint arXiv:2211.02597, 2022.

[36]. Hoelscher J, et al. , “A metric for finding robust start positions for medical steerable needle automation,” in Proc. Int. Conf. on Intelligent Robots and Systems (IROS), Sept. 2022.

[37]. Hoelscher J, “Backward planning for a multi-stage steerable needle lung robot,” IEEE Robot. Autom. Lett, vol. 6, no. 2, pp. 3987–3994, 2021. [PubMed: 33937523]

[38]. Sturm J, Engelhard N, Endres F, Burgard W, and Cremers D, “A benchmark for the evaluation of RGB-D SLAM systems,” in Proc. Int. Conf. on Intelligent Robots and Systems (IROS) IEEE, Oct. 2012, pp. 573–580.

[39]. Shavit Y and Ferens R, “Introduction to camera pose estimation with deep learning,” arXiv preprint arXiv:1907.05272, 2019.

[40]. Kikinis R, Pieper SD, and Vosburgh KG, “3D Slicer: a platform for subject-specific image analysis, visualization, and clinical support,” Intraoperative imaging and image-guided therapy, pp. 277–289, 2014.

[41]. He K, Zhang X, Ren S, and Sun J, “Deep residual learning for image recognition,” in Proc. Conf. on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[42]. Russakovsky O, et al., “Imagenet large scale visual recognition challenge,” Int. J. of Computer Vision, vol. 115, pp. 211–252, 2015, publisher: Springer.

[43]. Godard C, Mac Aodha O, Firman M, and Brostow GJ, “Digging into self-supervised monocular depth estimation,” in Proc. Int. Conf. on Computer Vision, 2019, pp. 3828–3838.

[44]. Kingma DP and Ba J, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

[45]. Lo P, et al. , “Extraction of airways from ct (EXACT’09),” IEEE Trans. Med. Imaging, vol. 31, no. 11, pp. 2093–2107, Nov. 2012. [PubMed: 22855226]

[46]. Chen A, Pastis N, Furukawa B, and Silvestri GA, “The effect of respiratory motion on pulmonary nodule location during electromagnetic navigation bronchoscopy,” Chest, vol. 147, no. 5, pp. 1275–1281, May 2015. [PubMed: 25357229]

[47]. Furukawa BS, Pastis NJ, Tanner NT, Chen A, and Silvestri GA, “Comparing pulmonary nodule location during electromagnetic bronchoscopy with predicted location on the basis of two virtual airway maps at different phases of respiration,” Chest, vol. 153, no. 1, pp. 181–186, Jan. 2018. [PubMed: 28629919]

[48]. Klein S, Staring M, Murphy K, Viergever MA, and Pluim JP, “Elastix: a toolbox for intensity-based medical image registration,” IEEE Trans. Med. Imag, vol. 29, no. 1, pp. 196–205, 2009.

[49]. Castillo R, et al. , “A reference dataset for deformable image registration spatial accuracy evaluation using the COPDgene study archive,” Phys. Med. Biol, vol. 58, no. 9, pp. 2861–2877, May 2013. [PubMed: 23571679]

[50]. Hofstad EF, et al. , “Intraoperative localized constrained registration in navigated bronchoscopy,” Medical Physics, vol. 44, no. 8, pp. 4204–4212, Aug. 2017. [PubMed: 28543091]

[51]. Fried I, Akulian JA, and Alterovitz R, “A clinical dataset for the evaluation of motion planners in medical applications,” arXiv preprint arXiv:2210.10834, 2022.

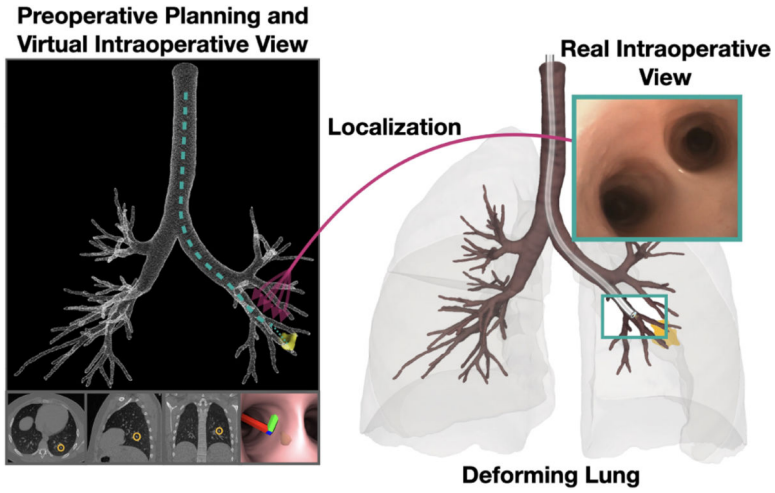


**Fig. 1.**

During bronchoscopy, a physician deploys a bronchoscope through the airways to target a suspicious lung nodule (yellow) for biopsy or treatment. The physician needs to localize the bronchoscope's tip intraoperatively (right) in reference to the planned needle insertion pose generated from the reconstruction of a preoperative CT scan (left). The airways' repetitive appearance, the airway lumen's limited distinguishable features, and respiratory deformation make this a challenging task.

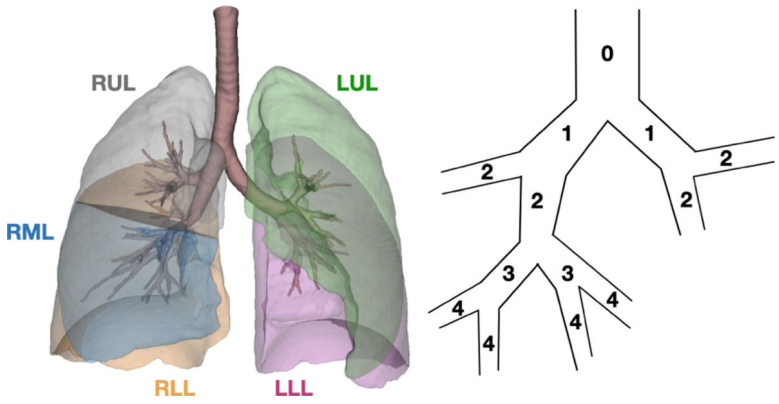


**Fig. 2.**

Human lungs consist of a left lung and a right lung. The right lung has three lobes: right upper lobe (RUL, gray), right middle lobe (RML, blue), and right lower lobe (RLL, orange). The left lung has two lobes: left upper lobe (LUL, green), and left lower lobe (LLL, pink). The bronchial tree (tan) begins at generation 0 and the generation count is incremented at each bifurcation (right). Localization in distal airways (generations 4+) is clinically challenging because of increased respiratory deformation in these peripheral lung regions.

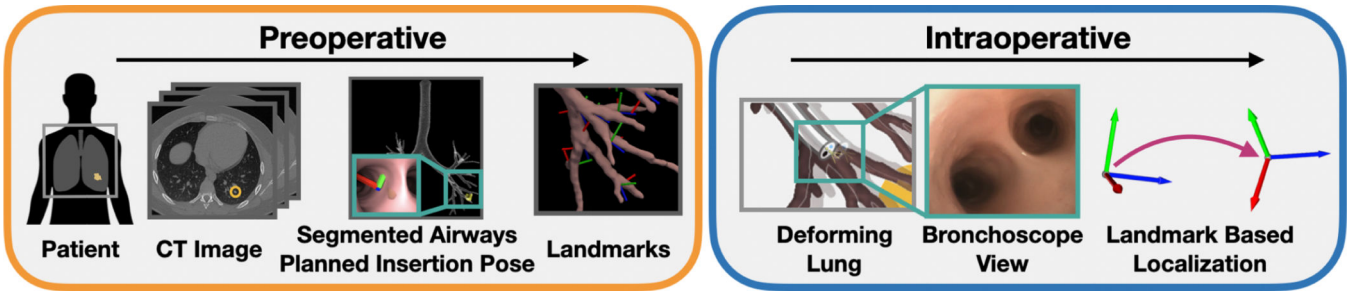


**Fig. 3.**

Workflow of our proposed method. In the preoperative stage, a patient undergoes a CT scan that generates a static 3D image of the anatomy which is used to segment the airways, plan the procedure, and identify landmarks. In the intraoperative stage, bronchoscope localization is performed using the bronchoscope camera's view in a lung undergoing respiratory motion.

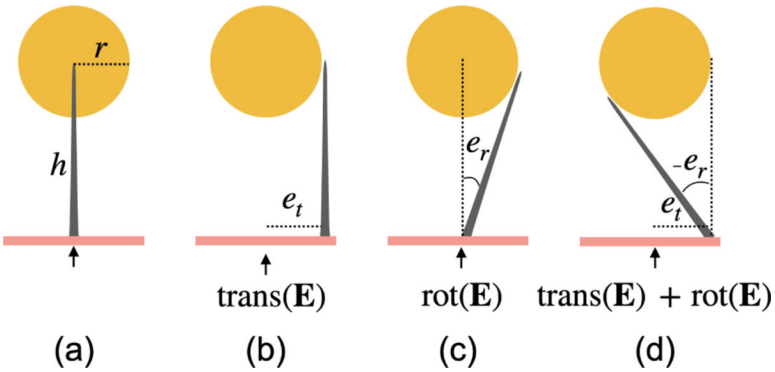


**Fig. 4.**

Assuming a nodule (yellow) with radius *r* and a needle with maximum insertion length *h*, we can calculate the error tolerance from a planned insertion site (arrows) on the surface of the airway wall (tan).

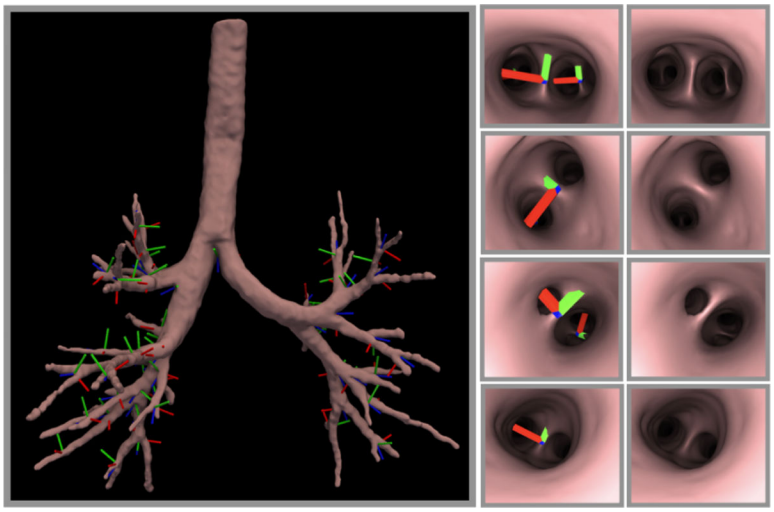

**Fig. 5.**

A 3D model of the airways segmented from a preoperative CT scan where each bifurcation is tagged with a 6 DoF pose (visualized using the coordinate frames) (left). Simulated bronchoscopy views from inside the airways with the bifurcation tags (middle), and the same views as they exist in the dataset without explicit tags (right).

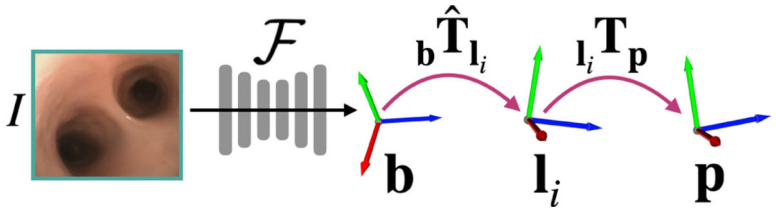


**Fig. 6.**

We localize the bronchoscope (**b**) with respect to the planned needle insertion pose (**p**) using the bifurcations on the airway wall ($\mathbf{l}_i$).

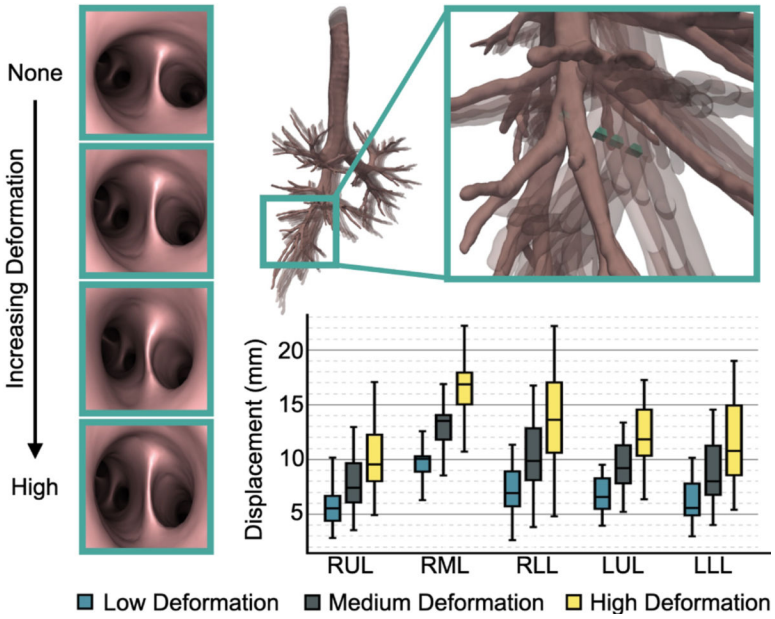


**Fig. 7.**

Intraoperative respiratory motion results in CT-to-body divergence. The bronchoscope (teal frustum, magnification) is displaced with the airways, which affects localization methods like EM, but the camera's internal view is only affected by local deformations, which are less significant at bifurcations (left). A boxplot showing the displacement of bifurcations in our evaluation data across the three levels of deformation by lung lobes averaged across the three patients.

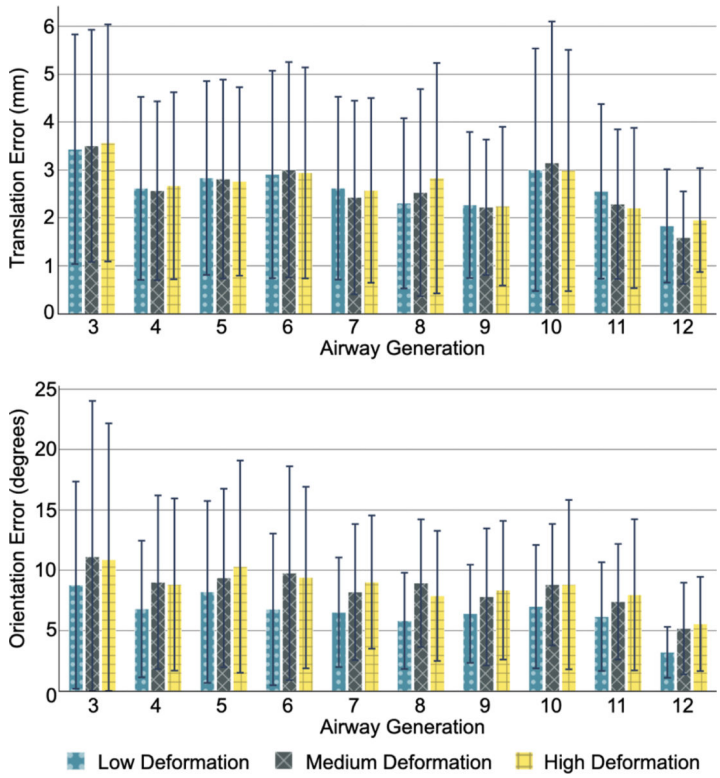


**Fig. 8.**

Average translation (top) and orientation (bottom) regression performance averaged across the three patients at each airway generation, grouped by level of deformation.

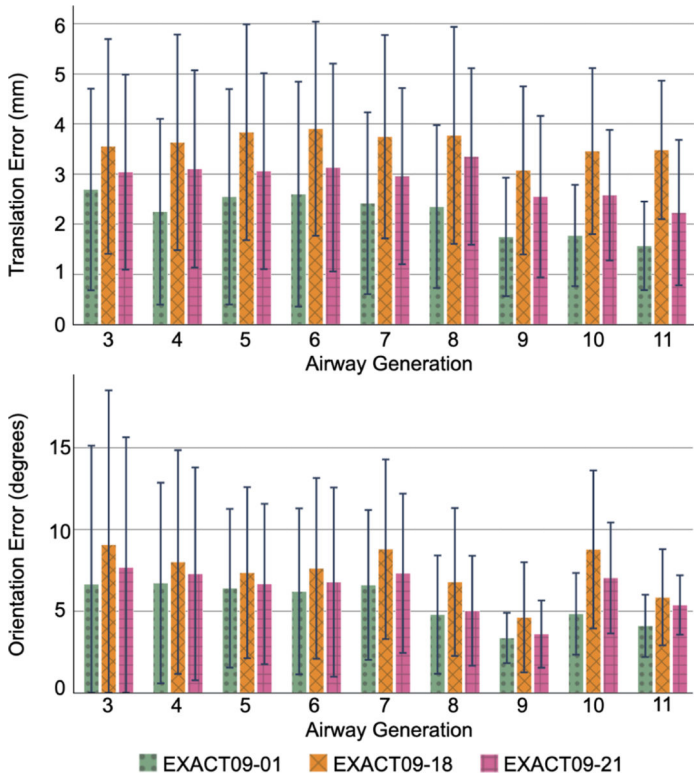


**Fig. 9.**

Using the three models trained on patients in the training set, we display average translation (top) and orientation (bottom) regression performance on 9,875 images from a new, separate patient.

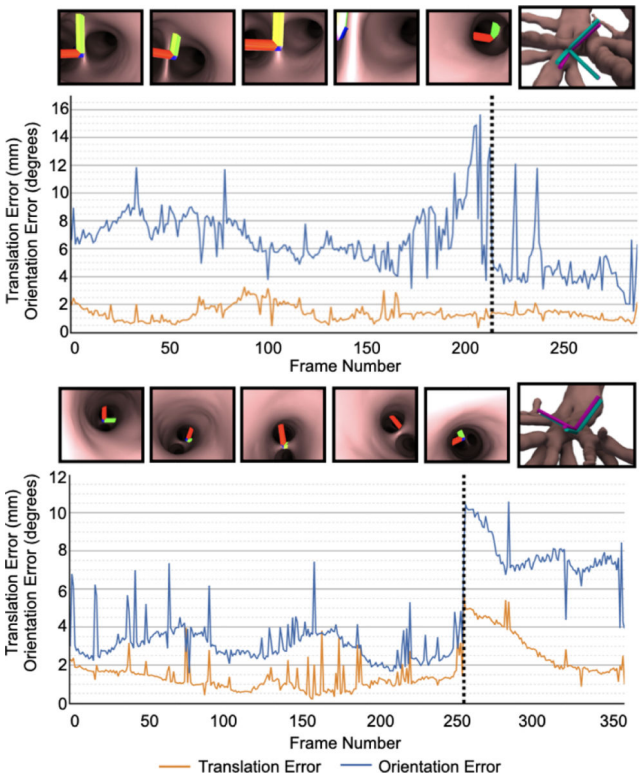


**Fig. 10.**

Tracking performance of our method approaching a piercing pose on two virtual bronchoscopy sequences. Black dotted lines indicate a change in the visible bifurcation. Coordinate frames in the images inside the airways show the estimated pose of the bronchoscope projected onto the bifurcation using ${}_{\mathbf{b}}\widehat{\mathbf{T}}_{\mathbf{l}_i}$. Teal and pink coordinate frames are the planned and estimated needle insertion poses, respectively.

**TABLE I**

DATASET STATISTICS.

| Patient ID | CT Dimensions | Voxel Dimensions | Number of Landmarks | Maximum Airway Generation | Number of Training Images | Number of Testing Images By Deformation Level |
|---|---|---|---|---|---|---|
| EXACT09–01 | (512, 512, 587) | 0.58 × 0.58 × 0.60 | 77 | 12 | 192,102 | 21,208 / 22,423 / 23,568 |
| EXACT09–18 | (512, 512, 423) | 0.69 × 0.69 × 0.70 | 32 | 8 | 91,572 | 9,936 / 8,970 / 9,643 |
| EXACT09–21 | (512, 512, 536) | 0.60 × 0.60 × 0.60 | 58 | 11 | 161,598 | 15,967 / 15,954 / 16,772 |