



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

## PANS: Probabilistic Airway Navigation System for Real-time Robust Bronchoscope Localization

Qingyao Tian<sup>1,2</sup>, Zhen Chen<sup>5</sup>, Huai Liao<sup>3</sup>, Xinyan Huang<sup>3</sup>, Bingyu Yang<sup>1,2</sup>,  
Lujie Li<sup>4</sup>, and Hongbin Liu<sup>1,5✉</sup>

<sup>1</sup> Institute of Automation, Chinese Academy of Sciences, Beijing, China  
liuhongbin@ia.ac.cn

<sup>2</sup> School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China

<sup>3</sup> Dept. of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

<sup>4</sup> Dept. of Radiology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China

<sup>5</sup> Centre for Artificial Intelligence and Robotics, Chinese Academy of Sciences, Hong Kong, China

**Abstract.** Accurate bronchoscope localization is essential for pulmonary interventions, by providing six degrees of freedom (DOF) in airway navigation. However, the robustness of current vision-based methods is often compromised in clinical practice, and they struggle to perform in real-time and to generalize across cases unseen during training. To overcome these challenges, we propose a novel Probabilistic Airway Navigation System (PANS), leveraging Monte-Carlo method with pose hypotheses and likelihoods to achieve robust and real-time bronchoscope localization. Specifically, our PANS incorporates diverse visual representations (*e.g.*, odometry and landmarks) by leveraging two key modules, including the Depth-based Motion Inference (DMI) and the Bronchial Semantic Analysis (BSA). To generate the pose hypotheses of bronchoscope for PANS, we devise the DMI to accurately propagate the estimation of pose hypotheses over time. Moreover, to estimate the accurate pose likelihood, we devise the BSA module by effectively distinguishing between similar bronchial regions in endoscopic images, along with a novel metric to assess the congruence between estimated depth maps and the segmented airway structure. Under this probabilistic formulation, our PANS is capable of achieving the 6-DOF bronchoscope localization with superior accuracy and robustness. Extensive experiments on the collected pulmonary intervention dataset comprising 10 clinical cases confirm the advantage of our PANS over state-of-the-arts, in terms of both robustness and generalization in localizing deeper airway branches and the efficiency of real-time inference. The proposed PANS reveals its potential to be a reliable tool in the operating room, promising to enhance the quality and safety of pulmonary interventions.

**Keywords:** Surgical navigation· 6-DOF bronchoscope localization· Probabilistic formulation

## 1 Introduction

Bronchoscopy has been a fundamental tool for examining and diagnosing airway lesions, as well as performing biopsies for lung disorders [1]. During bronchoscopy, surgeons utilize a camera-equipped flexible endoscope to perform a thorough inspection of the accessible bronchial branches and to navigate to lung peripheries and nodules as seen in pre-operative CT scans. However, the bronchoscope's limited field of view demands significant clinical experience for accurate localization within the airways. Therefore, there is an urgent demand for automatic methods to localize the bronchoscope with 6 DOF to assist airway navigation.

Various technologies have emerged to address this challenge, including electromagnetic (EM) navigation [7], 3D shape sensing [6], and visually navigated bronchoscopy (VNB). Particularly, VNB emerges as a promising area of study for its potential of accurate localization with cost-effectiveness. Existing VNB approaches predominantly utilize singular visual cues for bronchoscope tracking, which are achieved by registration [15,4,19,2,17], retrieval [27,18], or visual odometry [5,3]. Despite these efforts, current VNB studies are still in a developmental phase, facing several barriers to clinical implementation. Validation relies on biased and limited datasets that mainly cover proximal branches, such as insertion to a target site [18,2,22], casting doubts on their applicability in diverse clinical scenarios. Additionally, the absence of real-time operation in many VNB solutions [19,24,2,8,13] and the necessity for extensive, case-specific training [18,27] hinder their integration into routine clinical workflows. Consequently, VNB techniques are constrained by their limited robustness, absence of real-time performance, and lack of adaptability.

The challenge of limited robustness in VNB suggests the adoption of strategies used by experienced surgeons, who employ a comprehensive visual approach, including monitoring bronchoscopic motion and identifying anatomical landmarks, to pinpoint the bronchoscope's location. In response, we present the Probabilistic Airway Navigation System (PANS), enhancing localization robustness through a Monte Carlo framework. Replicating the surgeons' approach, PANS integrates diverse visual representations (*e.g.*, odometry and landmarks) for bronchoscope tracking. Specifically, PANS incorporates the Depth-based Motion Inference (DMI) module to propagate endoscopic pose hypotheses overtime, and the Bronchial Semantic Analysis (BSA) module that enhances pose likelihood measurement by distinguishing airway landmarks in bronchoscopic frames. The DMI module enhances case generalization by initially estimating the depth of incoming endoscopic frames prior to inferring camera motion. This transformation from RGB frames to the depth domain mitigates the issue posed by patient-specific textures and variable endoscopic illumination conditions, enabling a motion inference network trained on synthetic data to adapt to real endoscopic images effectively. By distinguishing airway landmarks, the BSA module adeptly clarifies ambiguities in similar airway areas, thereby significantly enhancing the overall precision of localization. It employs a novel metric for evaluating the congruence between depth estimations and segmented airway models, enhancing pose estimation's robustness. Furthermore, for real-time application across

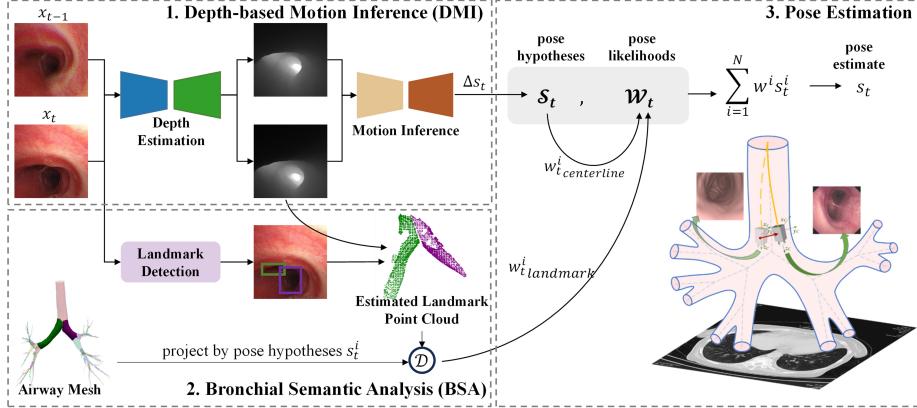


Fig. 1: Overview of the proposed PANS for real-time bronchoscope localization.

varied cases, PANS incorporates optimized lightweight neural networks for depth estimation, motion inference, and landmark detection. We perform an extensive comparative evaluation against SOTA methods, showcasing the superior performance of our PANS on real clinical datasets. Additionally, thorough ablation studies highlight the effectiveness of leveraging multiple visual representations and underscore the significance of the proposed BSA module in enhancing localization accuracy.

## 2 Method

### 2.1 Overview

Our PANS is illustrated in Fig. 1. We consider the problem of estimating the 6 DOF pose of the endoscope, which encompasses its position and orientation, as inferring a latent state  $s$  from camera image observations  $x$ . Due to the challenges in parameterizing the inherent uncertainties in the motion and operating conditions of the endoscope, we adopt a non-parametric way of representing probability distribution for the current endoscope pose state  $s_t$ , or commonly referred to as belief. This is achieved by utilizing a particle filter to approximate the belief distribution through a set of discrete samples or particles  $S_t = \{s_t^{[1]}, s_t^{[2]}, \dots, s_t^{[N]}\}$ . Each particle  $s_t^{[i]}$  ( $1 \leq i \leq N$ ) is a hypothesis of the endoscope state  $s_t$  at time  $t$ , and  $N$  is the number of particles. The likelihood of each particle's represented state is denoted in the set  $W_t = \{w_t^{[1]}, w_t^{[2]}, \dots, w_t^{[N]}\}$ . The essence of endoscopic tracking within our framework lies in generating  $S_t$  that more closely approximates the actual endoscopic state and in accurately determining  $W_t$ .

The current belief  $S_t$  is defined through a recursive Bayesian update. Initially, we estimate the endoscope's odometry by deriving the relative motion from  $x_{t-1}$  and  $x_t$ , represented by  $\Delta s$ . We propose a Depth-based Motion Inference (DMI)

module to calculate  $\Delta s$  (§ 2.2). This estimate is then used to propagate the particle set  $S_t$  by sampling from the state transition distribution:

$$s_t^{[i]} \sim p(s_t | s_{t-1}^{[i]}, \Delta s). \quad (1)$$

Subsequently, the weights  $W_t$  are updated based on the current observation  $x_t$ , using the proposed Bronchial Semantic Analysis (BSA) module (§ 2.3) as well as the centerline prior (§ 2.3), represented as:

$$w_t^{[i]} = p(s_t^{[i]} | x_t). \quad (2)$$

The bronchoscopic tracking pipeline is summarized in Algorithm 1.

## 2.2 Depth-based Motion Inference for Pose Hypotheses

In our PANS, we devise the DMI to accomplish motion prediction, by sampling endoscopic state particles by a noisy state transition:

$$\Delta s = f_\theta(x_{t-1}, x_t), \quad (3)$$

$$s_t^{[i]} = s_{t-1}^{[i]} \oplus \Delta s + \epsilon^{[i]} \sim \mathcal{N}, \quad (4)$$

where  $f_\theta$  is DMI module,  $\theta$  are the feedforward network parameters,  $\oplus$  denotes pose concatenation, and  $\epsilon^{[i]}$  represents Gaussian noise. Specifically, the DMI enhances case generalization by employing a two-step motion inference process, following our previous study [22]. It first estimates the depth of endoscopic frames and then deduces the camera motion. This approach reduces issues from variable textures and lighting, allowing a motion inference network trained on virtual depth to adapt to real endoscopic footage.

The depth estimation network is based on a cycle generative architecture [28]. Given an endoscopic frame  $x_t \in X$ , the depth estimation network  $G_{depth} : X \rightarrow Z$  maps the frame into its depth space  $Z$ , represented by  $z_t = G_{depth}(x_t)$ . Adversarial loss  $L_{adv}$ , cycle consistency loss  $L_{cyc}$  and identity loss  $L_{iden}$  are adopted to guide the network in learning domain transfer and content preservation. The LS-GAN loss [14] is implemented as adversarial loss  $L_{adv}$  and L1 losses are used for  $L_{cyc}$  and  $L_{iden}$ . To counteract scale instability and potential structural alterations in objects across different frames [11], view consistency loss  $L_{rec}$  and geometry consistency loss  $L_{gc}$  in [22] are adopted to enhance the depth estimation network's scale perception. Concurrently, the motion inference network is trained by minimizing the L2 norm of the discrepancy between predicted transformation and ground truth transformation, utilizing depth and endoscopic pose from virtual bronchoscopy as training data.

## 2.3 Bronchial Semantic Analysis for Pose Likelihoods

During measurement updates, particle weights are computed based on the similarity between the current observation and the supposed observation of each

**Algorithm 1** The Pipeline of PANS for Bronchoscope Localization.

---

```

1: Input: Airway segments, initial bronchoscope pose, bronchoscopic frames  $\{x_t\}_{t=1}^T$ 
2: Output: Estimated pose  $\hat{s}_t$  for each frame  $x_t$ 
3: Initialize particle set  $S_0$  around initial pose
4: for each time step  $t = 1$  to  $T$  do
5:   Estimate odometry  $\Delta s$  using DMI (Eq. 3) based on  $x_{t-1}$  and  $x_t$ 
6:   Update  $S_t$  by applying  $\Delta s$  to  $S_{t-1}$  (Eq. 4)
7:   Calculate likelihoods  $W_t$  using BSA and centerline constraint (Eq. 5-8)
8:   Compute estimated pose  $\hat{s}_t = \sum_{i=1}^N w_t^i s_t^i$  (Eq. 9)
9:   Resample particles  $S_t$  based on  $W_t$  to focus on higher probability states
10: end for

```

---

particle state. Previous studies in bronchoscope localization have predominantly relied on image intensity or structural similarity for matching virtual and real endoscopic frames [12,19,2,8,13]. However, the high degree of similarity in endoscopic frames across various airway sections would lead these methods to encounter local minima, hindering their tracking performance.

To address the limitation of existing methods [12,19,2,8,13], we propose the BSA to calculate pose likelihood based on endoscopic frame semantic analysis. Specifically, we first identify the anatomical branches (*e.g.*, the right upper lobe) visible in the endoscopic view as landmarks. Subsequently, the BSA assesses the likelihood of pose hypotheses by examining the alignment between the depth estimation of each identified branch and the segmented airway model.

To overcome the similarity in different areas of the airway, BSA’s landmark detection encompasses a three-stage pipeline: lumen detection, tracking, and airway association, as detailed in our prior work, BronchoTrack [23]. Initially, BSA detects airway lumens without recognizing their anatomical branch label, using the high-performance YOLOv7 [25] detector. These lumens are then tracked across subsequent frames using motion modeling and deep appearance descriptors, which allow for re-identification even after temporary disappearance. Finally, a training-free airway association module matches these tracked lumens to their corresponding anatomical branches by establishing a hierarchy of lumens, leveraging CT segmentation for lumen recognition and utilizing tracking data for label propagation. This approach maps each observed lumen to its anatomical branch, and propagates their anatomical branch label temporally by tracking them in successive frames.

Following the identification of airway anatomical branches, the BSA module evaluates the likelihood of each pose hypothesis by measuring the discrepancy between the depth estimation of detected branches and the projected point cloud of these branches from the particle’s state, as follows:

$$w_{t,l}^{[i]} = \sum_k D(h(o_{t_k}), \text{proj}(m_k, s_t^{[i]})), \quad (5)$$

where  $k$  is the anatomical label of detected airway branches,  $D$  is a distance metric. The depth estimation  $h(o_{t_k})$  of each detected branch  $k$  is derived from estimated depth within detected bounding box and is compared with  $\text{proj}(m_k, s_t^{[i]})$ ,

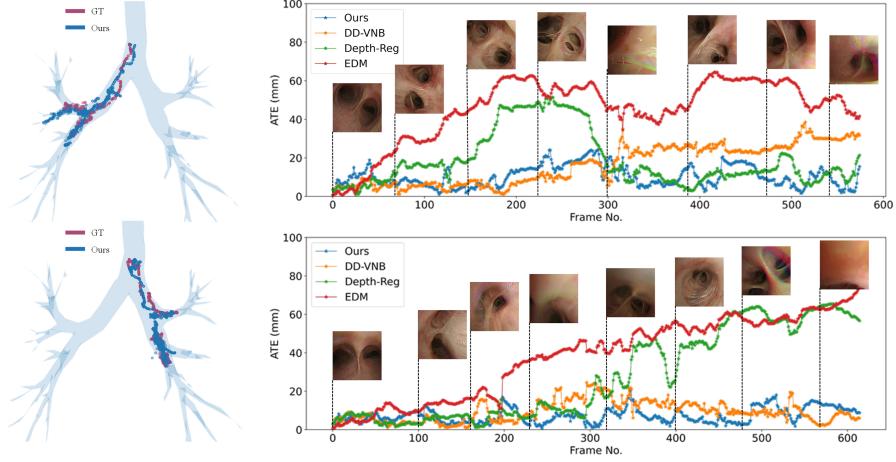


Fig. 2: Localization trajectory and error on an example patient case. Two inspection trajectories into the left and right lung are presented separately.

which signifies the point cloud  $m_k$  of segmented airway branch  $k$  projecting from world coordinate to camera coordinate.

We adapt the Chamfer Distance into a binary count metric to evaluate the overlap between two point clouds, accounting for outliers in the depth-estimated cloud. This metric  $D$  is defined as:

$$D(A, B) = \frac{1}{|A|} \left| \left\{ a \in A \mid \min_{b \in B} \{d(a, b)\} < \rho \right\} \right| \quad (6)$$

where  $d$  is the Euclidean distance,  $\rho = 3\text{mm}$  is threshold for matching, and  $|\cdot|$  signifies counting of points.

Moreover, as bronchoscopes are typically near and oriented along the tubular's centerline, we introduce a centerline constraint as prior knowledge, as:

$$w_{t,c}^{[i]} = \mathcal{N}(e|0, \sigma_1) \cdot \mathcal{N}(\phi|0, \sigma_2), \quad (7)$$

where  $e$  is the distance from state  $s_t^{[i]}$  to the closest centerline, and  $\phi$  represents the angle between state  $s_t^{[i]}$  and the nearest centerline.  $\sigma_1$  and  $\sigma_2$  are set to  $\frac{r}{2}$  and  $\frac{\pi}{6}$ , with  $r$  being the radius of the nearest branch.

Combining  $w_{t,l}^{[i]}$  and  $w_{t,c}^{[i]}$ , the particle weights are updated by:

$$w_t^{[i]} = \frac{w_{t,l}^{[i]} \cdot w_{t,c}^{[i]}}{\|w_{t,l}^{[i]} \cdot w_{t,c}^{[i]}\|}. \quad (8)$$

By calculating the likelihoods of endoscopic state proposals, or particles, the estimated endoscopic pose is calculated by the weighted sum of particles as:

$$\hat{s}_t = \sum_{i=1}^N w_t^i s_t^i. \quad (9)$$

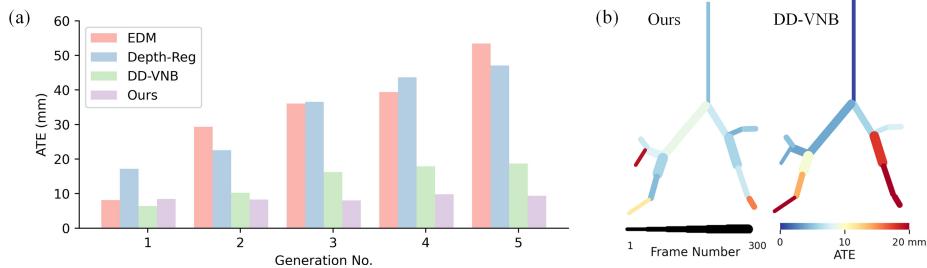


Fig. 3: (a) Localization ATE on different airway generations. (b) Localization ATE for the top two methods across various branches in an example case.

### 3 Experiment

#### 3.1 Dataset and Implementation Details

We conducted our experiments on 31 bronchoscopic intervention cases, each from a different patient. For experiments, we allocated 20 for training, 1 for validation, and 10 for testing. Each test case covers both lungs through two sequences. Bronchoscopic videos are filmed with the frame rate around 15fps using Olympus bronchoscope during regular inspection procedure, where the bronchoscope is inserted into every possible airway for thorough inspection. Patient airway meshes are semantically segmented from CT scans [20,26], categorized by branch anatomy. Our testing dataset spans shallow to deep airway branches up to the 5th generation. It features video frames with challenges like poor visibility, motion blur, and bubbles, presenting a wider range of difficulties than typically reported. Fig. 2 displays the trajectory and select frames from a representative case. Ground truth bronchoscope poses are manually labeled by experts through registering virtual and real bronchoscopic views. More details are provided in the supplementary material.

We utilized the Pytorch framework on an NVIDIA RTX 3090 GPU for training. All networks process frames resized to  $256 \times 256$  pixels. The depth estimation network, based on ResNet-50 [9], was trained with a learning rate of 0.0001 over 100 epochs and a batch size of 1, adhering to the weights of losses specified in [22]. The motion inference network, utilizing a FlownetC [10] encoder and MLP decoder, was trained for 300 epochs with a learning rate of 1e-5 and a batch size of 64. The YOLOv7 [25] lumen detector underwent training for 300 epochs with a batch size of 64. The appearance descriptor for lumen tracking adopts a ResNet-50 architecture trained by image retrieval follows [23].

#### 3.2 Comparison Results

Our PANS is evaluated against SOTA monocular endoscopic localization techniques. This comparison encompasses depth registration (Depth-Reg) approaches [19,2], the DD-VNB [22], and EDM [16]. Evaluation metrics adhere to VNB research standards [19,8,22]. Table 1 shows that our method outperforms SOTA

Table 1: Localization across the 10-patient cases encompassing 20 sequences and 10,004 frames.

Method	ATE (mm)↓	SR5↑	SR10↑
EDM	$35.7 \pm 23.2$	3.90%	8.8%
Depth-Reg	$35.1 \pm 28.4$	13.9%	25.7%
DD-VNB	$15.0 \pm 11.7$	22.9%	44.3%
PANS <i>w/o DVR</i>	$18.9 \pm 18.6$	15.1%	40.1%
PANS <i>w/o BSA</i>	$15.1 \pm 13.1$	21.8%	49.2%
<b>PANS</b>	<b><math>8.7 \pm 6.0</math></b>	<b>28.4%</b>	<b>70.0%</b>

by a significant margin across all evaluated metrics, achieving a localization ATE of  $8.7 \pm 6.0$  mm and SR10 of 70.0%. When compared to methods that rely solely on registration [19,2] or visual odometry [16], PANS demonstrates superior performance, benefiting from diverse visual representations. Moreover, PANS outperforms hybrid approach [22], underscoring its effective probabilistic framework and semantic analysis. Fig. 2 illustrates a test case visualization; additional results are in the supplementary materials.

The comparative localization accuracy of our method against SOTA across various airway generations is depicted in Fig. 3. Our technique demonstrates consistent accuracy even as it navigates deeper into the airway generations. In contrast, DD-VNB, Depth-Reg, and EDM exhibit a decline in performance with increasing airway depth. Specifically, EDM, which relies on incremental camera motion estimation, experiences a progressive deviation from the ground truth over time. Both DD-VNB and Depth-Reg struggle in the deeper airways, potentially due to the structural homogeneity present in the smaller bronchi.

### 3.3 Ablation Studies

We assess the impact of Diverse Visual Representations (DVR) and BSA in Table 1. The PANS *w/o DVR* variant, which excludes the landmark weight and localize depending solely on motion inference with centerline guidance, demonstrates a notable drop in performance, highlighting the critical contribution of multiple visual representations to accuracy and robustness. Furthermore, PANS *w/o BSA* substitutes landmark-based matching with global NCC depth matching, following the approach in [22,19]. The increase in SR10 indicates enhanced tracking success across more frames, suggesting greater localization robustness.

We also conduct a particle number sensitivity test on the validation set to assess the trade-off of accuracy and computational efficiency [21], by normalizing the lowest ATE to 100% accuracy, as illustrated in Fig. 4. For comparison and ablation studies, we set the particle number to 216, aligning PANS’s speed with the video capture rate of approximately 15Hz for real-time processing. Our PANS surpasses other methods in speed, especially those independent of case-specific training. Depth-Reg methods such as Shen et al. [19] operate at 0.35 fps, and

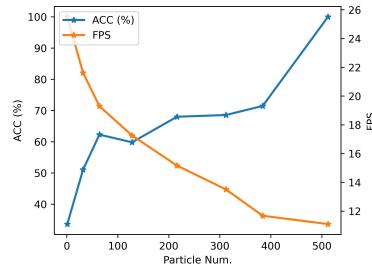


Fig. 4: Performance of PANS with different particle numbers.

Banach et al. [2] take roughly 2s per camera pose. Other approaches include Gu et al. [8] at 5 fps, Luo et al. [13] at 0.41s per frame, and Wang et al. [24] at 0.08s per frame. Although DD-VNB [22] demonstrates real-time speeds, it does not match PANS’s accuracy and robustness in deeper airway generations.

## 4 Conclusion

In this work, we propose an innovative bronchoscope localization framework, named PANS, leveraging comprehensive deep representations into a Monte-Carlo framework to enhance robustness. Specifically, we propose the Depth-based Motion Inference (DMI) module to propagate endoscopic pose hypotheses over time, and the Bronchial Semantic Analysis (BSA) module that enhances pose likelihood measurement by distinguishing airway landmarks in bronchoscopic frames. Tested on a 10-patient dataset from real interventions, our PANS excelled in tracking within complex airway branches, achieving real-time performance and generalization beyond the training data.

**Acknowledgments.** This work was supported by the Centre of AI and Robotics, Hong Kong Institute of Science and Innovation, Chinese Academy of Sciences, sponsored by InnoHK Funding, HKSAR, partially supported by Institute of Automation, Chinese Academy of Sciences, and partially supported by Sichuan Science and Technology Program (Grant number: 2023YFH0093).

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Andolfi, M., Potenza, R., Capozzi, R., Liparulo, V., Puma, F., Yasufuku, K.: The role of bronchoscopy in the diagnosis of early lung cancer: a review. *Journal of thoracic disease* **8**(11), 3329 (2016)
2. Banach, A., King, F., Masaki, F., Tsukada, H., Hata, N.: Visually navigated bronchoscopy using three cycle-consistent generative adversarial network for depth estimation. *Medical image analysis* **73**, 102164 (2021)
3. Borrego-Carazo, J., Sanchez, C., Castells-Rufas, D., Carrabina, J., Gil, D.: Bronchopose: an analysis of data and model configuration for vision-based bronchoscopy pose estimation. *Computer Methods and Programs in Biomedicine* **228**, 107241 (2023)
4. Deguchi, D., Mori, K., Feuerstein, M., Kitasaka, T., Maurer Jr, C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Selective image similarity measure for bronchoscope tracking based on image registration. *Medical Image Analysis* **13**(4), 621–633 (2009)
5. Deng, J., Li, P., Dhaliwal, K., Lu, C.X., Khadem, M.: Feature-based visual odometry for bronchoscopy: A dataset and benchmark. In: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 6557–6564. IEEE (2023)
6. Floris, I., Adam, J.M., Calderón, P.A., Sales, S.: Fiber optic shape sensors: A comprehensive review. *Optics and Lasers in Engineering* **139**, 106508 (2021)

7. Folch, E.E., Labarca, G., Ospina-Delgado, D., Kheir, F., Majid, A., Khandhar, S.J., Mehta, H.J., Jantz, M.A., Fernandez-Bussy, S.: Sensitivity and safety of electromagnetic navigation bronchoscopy for lung cancer diagnosis: systematic review and meta-analysis. *Chest* **158**(4), 1753–1769 (2020)
8. Gu, Y., Gu, C., Yang, J., Sun, J., Yang, G.Z.: Vision–kinematics interaction for robotic-assisted bronchoscopy navigation. *IEEE Transactions on Medical Imaging* **41**(12), 3600–3610 (2022)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
10. Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T.: Flownet 2.0: Evolution of optical flow estimation with deep networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2462–2470 (2017)
11. Karaoglu, M.A., Brasch, N., Stollenga, M., Wein, W., Navab, N., Tombari, F., Ladikos, A.: Adversarial domain feature adaptation for bronchoscopic depth estimation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24. pp. 300–310. Springer (2021)
12. Luo, X., Mori, K.: A discriminative structural similarity measure and its application to video-volume registration for endoscope three-dimensional motion tracking. *IEEE transactions on medical imaging* **33**(6), 1248–1261 (2014)
13. Luo, X., Xie, L., Zeng, H.Q., Wang, X., Li, S.: Monocular endoscope 6-dof tracking with constrained evolutionary stochastic filtering. *Medical Image Analysis* **89**, 102928 (2023)
14. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2794–2802 (2017)
15. Mori, K., Deguchi, D., Sugiyama, J., Suenaga, Y., Toriwaki, J.i., Maurer Jr, C.R., Takabatake, H., Natori, H.: Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. *Medical Image Analysis* **6**(3), 321–336 (2002)
16. Recasens, D., Lamarca, J., Fácil, J.M., Montiel, J., Civera, J.: Endo-depth-and-motion: Reconstruction and tracking in endoscopic videos using depth networks and photometric constraints. *IEEE Robotics and Automation Letters* **6**(4), 7225–7232 (2021)
17. Sganga, J., Eng, D., Graetz, C., Camarillo, D.: Offsetnet: Deep learning for localization in the lung using rendered images. In: 2019 international conference on robotics and automation (ICRA). pp. 5046–5052. IEEE (2019)
18. Sganga, J., Eng, D., Graetz, C., Camarillo, D.B.: Autonomous driving in the lung using deep learning for localization. *arXiv preprint arXiv:1907.08136* (2019)
19. Shen, M., Gu, Y., Liu, N., Yang, G.Z.: Context-aware depth and pose estimation for bronchoscopic navigation. *IEEE Robotics and Automation Letters* **4**(2), 732–739 (2019)
20. Tan, Z., Feng, J., Zhou, J.: Sgnet: Structure-aware graph-based network for airway semantic segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 153–163. Springer (2021)
21. Thrun, S.: Probabilistic robotics. *Communications of the ACM* **45**(3), 52–57 (2002)
22. Tian, Q., Liao, H., Huang, X., Chen, J., Zhang, Z., Yang, B., Ourselin, S., Liu, H.: Dd-vnb: A depth-based dual-loop framework for real-time visually navigated bronchoscopy. *arXiv preprint arXiv:2403.01683* (2024)

23. Tian, Q., Liao, H., Huang, X., Yang, B., Wu, J., Chen, J., Li, L., Liu, H.: Bronchotrack: Airway lumen tracking for branch-level bronchoscopic localization. arXiv preprint arXiv:2402.12763 (2024)
24. Wang, C., Oda, M., Hayashi, Y., Villard, B., Kitasaka, T., Takabatake, H., Mori, M., Honma, H., Natori, H., Mori, K.: A visual slam-based bronchoscope tracking scheme for bronchoscopic navigation. International Journal of Computer Assisted Radiology and Surgery **15**, 1619–1630 (2020)
25. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7464–7475 (2023)
26. Xie, W., Jacobs, C., Charbonnier, J.P., van Ginneken, B.: Structure and position-aware graph neural network for airway labeling. arXiv preprint arXiv:2201.04532 (2022)
27. Zhao, C., Shen, M., Sun, L., Yang, G.Z.: Generative localization with uncertainty estimation through video-ct data for bronchoscopic biopsy. IEEE Robotics and Automation Letters **5**(1), 258–265 (2019)
28. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)