Navigational Bronchoscopy in Critical Care via End-to-End Pose Regression

Emile Mackute[1 [0009−0004−4535−4619]], Francis Xiatian Zhang[1,2 [0000−0003−0228−6359]], Kevin Dhaliwal[2 [0000−0002−3925−3174]], and Mohsen Khadem[1,2 [0000−0002−6873−273X]]

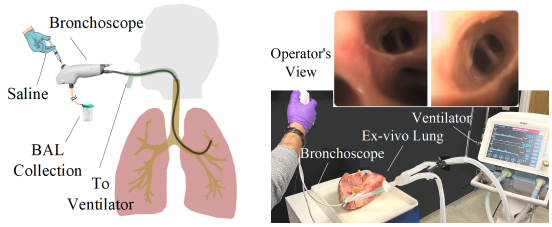
[1] School of Informatics, University of Edinburgh, Edinburgh, United Kingdom
[2] Baillie Gifford Pandemic Science Hub, Institute of Regeneration and Repair, University of Edinburgh, Edinburgh, United Kingdom
{s1657385, francis.zhang, kev.dhaliwal, mohsen.khadem}@ed.ac.uk

Abstract. Bronchoscopy is a minimally invasive procedure for diagnosing and treating lung conditions, but accurate navigation remains challenging and resource-intensive due to reliance on preoperative imaging, sensor-based tracking, and the low-saliency visual environment of the airways. To address these limitations, we propose a novel Navigational Bronchoscopy framework that enables real-time guidance and repeatable interventions without requiring external sensors or CT scans, making it particularly suitable for mechanically ventilated patients in critical care units with limited access to preoperative imaging. Our approach leverages deep learning, combining airway landmark recognition with deep visual features and a Vision Transformer (ViT)-based pose regression network to track bronchoscope motion. The framework is deployed on a commercially available bronchoscope and validated through trials in both a phantom lung model and a mechanically ventilated ex-vivo human lung. Results show that our ViT-based model achieves the lowest pose estimation errors among tested methods. Furthermore, in ex-vivo trials, our system successfully guided the bronchoscope to predefined targets, achieving high similarity scores for reliable landmark identification. These findings highlight the feasibility of our approach for real-world clinical applications.

Keywords: Navigational Bronchoscopy · Vision Transformer · Pose Estimation.

1 Introduction

Bronchoscopy is a diagnostic and therapeutic tool in critical care, used for examining the airways with a flexible tube equipped with an endoscopic camera. Common interventions include bronchoalveolar lavage and biopsy. However, in critically ill and mechanically ventilated patients, airway sampling is particularly challenging due to the absence of navigation tools or preoperative images, making procedures like Bronchoalveolar Lavage (BAL) [8], shown in Fig. 1, difficult



(a) (b)

Fig. 1. (a) Schematic of Bronchoalveolar Lavage (BAL) in critical care. Saline is introduced via the bronchoscope to wash the airways, and a suction pump retrieves the fluid for analysis. The lack of repeatability in re-sampling or targeted therapy delivery limits its clinical impact. We propose a navigation platform to assist operators in accurately returning to the same location for re-sampling or localized therapy. (b) Experimental setup using ex-vivo human lungs and a commercial bronchoscope (Ambu aScope 4, Ambu Ltd.) to simulate BAL.

to accurately repeat and interpret. The lack of repeatability for resampling or local therapy delivery further limits its impact on patient management [13].

Navigational Bronchoscopy (NB) systems, such as Illumisite (Medtronic, USA) [5], and robotics-assisted platforms like Ion (Intuitive Surgical, USA) [18], Monarch (Auris Health, USA) [19], and Galaxy (Noah Medical, USA) [21], have advanced lung cancer diagnosis by integrating sensors (e.g., electromagnetic tracking, optical fibers) with pre- or intra-operative imaging for precise guidance. However, these systems are unsuitable for critical care, where imaging is often unfeasible, and robotic platforms are not designed for ICU use. As a result, NB remains largely confined to cancer and infection diagnostics outside critical care.

Vision-based methods, estimating bronchoscope pose from sequential images, offer an alternative to sensor-based navigation. However, feature-based approaches [11, 20] struggle due to the lack of salient features in bronchoscopic videos [4, 24]. Learning-based methods, including CNNs [10, 25] and GNNs [6], show promise in endoscopy but remain unvalidated in bronchoscopy due to feature sparsity, occlusions, and motion blur.

Supervised [1,22] and self-supervised CNN models [9,16] have been developed for direct pose estimation but struggle with reliable tracking in real human lungs [12]. The Probabilistic Airway Navigation System (PANS) [23] improves 6-DOF localization using depth-based motion inference and bronchial semantic landmark detection but relies on preoperative airway segmentation, limiting its applicability in critical care.

Our main contribution is a novel NB framework designed for mechanically ventilated patients in critical care, operating without preoperative imaging. CT-based NB platforms are prone to errors from patient and lung motion—such as breathing and peristalsis—which can shift bronchial structures by up to 25 mm per breathing cycle [14]. To address this, we leverage a vision foundation model

for trajectory generation and introduce visual landmark-based loop closure, enabling accurate and repeatable access to sampled regions. This facilitates reliable re-sampling and targeted drug delivery following infection or inflammation detection. We validate our method using a phantom lung model (Fig. 3) and an ex-vivo ventilated human lung (Fig. 1(b)). Although the bronchoscope is manually controlled and landmark recognition enables loop closure, pose estimation transforms recognition into actionable navigation by providing real-time directional context [23], especially in low-texture or distorted airway regions where CT maps are unavailable. To our knowledge, this is the first NB pipeline developed for bronchoscopy in critical care without reliance on preoperative imaging.

2 Methodology

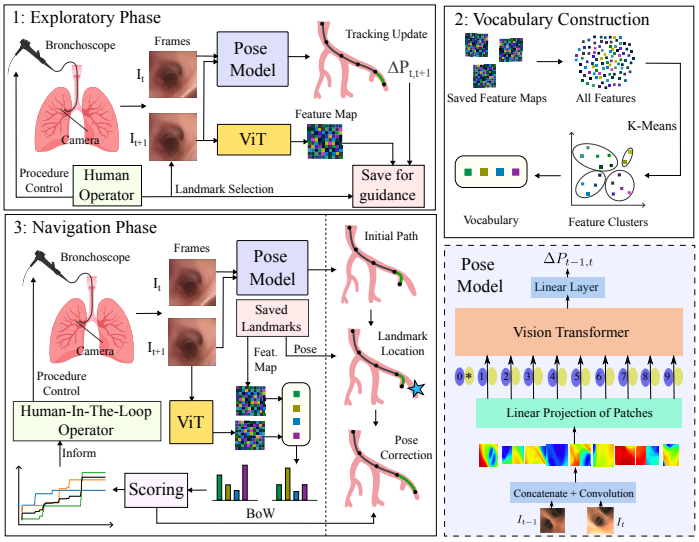


Fig. 2. Navigational pipeline comprising four stages: (1) The operator inspects the lungs, marking relevant landmarks or areas of interest. Visual features are extracted from images and stored alongside tracked camera positions. (2) Extracted features are clustered using K-Means to construct a Bag-of-Words (BoW) vocabulary. (3) During subsequent bronchoscopy rounds, the bronchoscope is guided through the lungs while image features and camera pose are computed in real time. BoW vectors identify previously marked landmarks, enabling pose correction to reduce tracking error. A navigation score provides feedback on proximity to the target location. (4) A pose regression network, shown in the bottom right, is used in both exploratory and navigation rounds. It processes image pairs I_{t-1} and I_t, which are concatenated and passed through a convolutional layer. The output is reshaped into patches and fed into a Vision Transformer (ViT) followed by a feedforward layer to predict pose.

Our navigation pipeline consists of three stages, illustrated in Fig. 2:

1. **Exploratory Phase:** The operator inspects the airways, marking landmarks or areas of interest while tracking the pose of the bronchoscope. This phase mirrors the initial round of critical care bronchoscopy, where the goal is to sample from as many locations as possible.
2. **Vocabulary Construction:** A Bag-of-Words (BoW) visual vocabulary is built from deep features extracted from recorded frames.
3. **Navigation Phase:** During resampling or drug delivery, real-time pose predictions and landmark recognition guide the bronchoscope. A navigation score provides feedback on target proximity, with landmarks serving as visual anchors to minimize pose drift.

The backbone of the NB pipeline in both exploratory and navigation phases is a Visual Foundation Model, specifically DINO, a pre-trained Vision Transformer (ViT) [3]. ViT is utilized for:
(i) Camera Pose Prediction: The predicted camera pose is visually displayed to the operator during both the exploratory and navigation phases, aiding guidance.
(ii) Deep Feature Extraction: Feature maps extracted from camera frames are stored as a Bag-of-Words (BoW) after the exploratory phase and later used to minimize pose prediction error and drift during navigation. The following subsections provide further details on these steps.

2.1 6D Pose Estimation

To achieve real-time bronchoscope pose tracking, we fine-tune a ViT for relative camera pose regression, as illustrated in Fig. 2. The network processes consecutive frames (I_{t-1}, I_t) and predicts a 12-dimensional pose vector:

$$\Delta P^C_{t-1,t} = \left(\Delta R^C_{t-1,t}, \Delta r^C_{t-1,t}\right), \tag{1}$$

where $\Delta R^C_{t-1,t}$ and $\Delta r^C_{t-1,t}$ represent the predicted rotation matrix and the translation vector between the time steps $t-1$ and t. The network is initialized with pre-trained weights [3] and fine-tuned on publicly available bronchoscopy sequences with ground truth poses [12]. Each training sample consists of two consecutive frames and their relative transformation, with annotations derived from an electromagnetic tracking system used only during training, not at inference.

A weighted loss function supervises translation and rotation simultaneously. Translation loss is computed using Mean Squared Error (MSE) as $\mathcal{L}_T = \frac{1}{3}\sum_{i=1}^{3}\left(\Delta T_i - \Delta\hat{T}_i\right)^2$, where $\Delta\hat{T}$ and ΔT are the predicted and target translations, respectively. For rotation loss, we use a Procrustes orthonormalization approach [2], mapping a 3×3 matrix onto the closest valid rotation matrix in SO(3) by minimizing the Frobenius norm difference, ensuring continuity and differentiability:

$$\Delta R = \arg\min_{R\in SO(3)} ||R - M||_F^2, \tag{2}$$

The final rotation error loss function is measured using quaternion-based angular difference:

$$\mathcal{L}_R = 1 - q_w(\Delta R^{-1}\Delta\hat{R})^2, \tag{3}$$

where $q(\cdot)$ maps the rotation matrix to a unit quaternion, with w representing the scalar term. This mapping ensures continuous optimization during training. We note that the pose regression model operates separately from the ViT used for image feature extraction in the next stage.

2.2 Landmark Recognition and Pose Refinement

During the exploratory phase, a pre-trained ViT [3] extracts deep feature representations from camera frames. The final transformer layer encodes frame-specific information, which is utilized for landmark recognition. Unlike traditional vision-based SLAM algorithms [7,15] that rely on handcrafted local feature descriptors, our approach employs a deep-learning-based Bag-of-Words (BoW) model. Specifically, deep image features extracted by ViT are clustered using the K-Means algorithm, where the resulting cluster centers define a landmark set: $\mathcal{L} = \{\mathbf{v}_l \mid l \in \mathbb{N}, l \text{ generation of airway branch}\}$, where each landmark $\mathbf{v}_l$ corresponds to a feature descriptor associated with a specific airway branch.

The landmarks recorded in exploratory phase serve as reference points for pose refinement during navigation in serial bronchoscopy, where feature vectors from incoming frames are extracted using the same ViT model and mapped to their closest K-Nearest Neighbor (KNN) match within the BoW vocabulary. This process generates a histogram representation of the frame's BoW vector. Landmarks are identified by computing the normalized dot product similarity between the current frame vector $\mathbf{v}_t$ and reference landmark vectors $\mathbf{v}_l$:

$$\mathrm{s}(\mathbf{v}_t, \mathbf{v}_l) = \frac{\mathbf{v}_t \cdot \mathbf{v}_l}{\|\mathbf{v}_t\|\|\mathbf{v}_l\|}, \quad \forall \mathbf{v}_l \in \mathcal{L}. \tag{4}$$

This similarity metric ensures robust landmark matching and triggers pose refinement when the similarity exceeds a threshold, thereby facilitating loop closure and drift correction in pose estimation.

To enhance robustness, similarity scores are averaged across matching frames, while the lowest 10% of values are excluded to mitigate outliers. Furthermore, navigation accuracy is quantified using a dynamic navigation score (S_t), aggregating weighted maximum similarity scores across all identified landmarks:

$$S_t = \sum_{l=0}^{L} w_l \max_t \mathrm{s}(\mathbf{v}_t, \mathbf{v}_l), \tag{5}$$

where S_t represents the navigation score at time t. The weight w_l assigns exponentially increasing significance to landmarks detected later in the procedure, ensuring a higher influence on pose refinement. The weights are normalized that:

$$w_l = \frac{2^{l-L}}{\sum_{l=0}^{L} w_l}. \tag{6}$$

Table 1. Comparison of different end-to-end pose regression methods.

Method	Params (M)	RPE^{trans} (mm)	RPE^{rot} (°)	ATE^{trans} (cm)	ATE^{rot} (°)
[16]	27.9	0.120 ± 0.112	0.85 ± 0.58	7.10 ± 1.30	126.80 ± 25.50
[22]	43.6	1.790 ± 1.510	11.95 ± 9.85	8.30 ± 2.50	126.10 ± 28.00
Ours	86.4	**0.095 ± 0.070**	**0.45 ± 0.32**	**2.10 ± 0.90**	**19.00 ± 5.90**

The predicted pose is used to visualize the bronchoscope's trajectory, while detected landmarks provide loop closure corrections. When a valid landmark match is confirmed, the trajectory is refined by computing the relative pose transformation between the current frame and the closest matched landmark. The pose model then estimates the transformation update, which is propagated backward using the interpolation method from [17], stopping at the nearest previously matched frame or the trajectory's starting point. This approach reduces accumulated drift and enhances pose consistency during serial bronchoscopy.

3 Experiments

To validate our navigation framework, we conducted bronchoscopy trials in two settings: a phantom lung model (Bronchoscopy training model, Koken Co.) and a mechanically ventilated ex-vivo human lung. The phantom model served as a baseline for assessing feasibility and identifying hyper parameters, including threshold for land mark detection, while the ex-vivo lung trials tested the system under anatomically realistic conditions. In our experiments, the inference pipeline achieved approximately 20Hz on a standard RTX 3060 GPU, which is sufficient for real-time clinical guidance during bronchoscopy procedures. Trials lasted approximately 12–15 minutes, reflecting the time needed for initial exploration, landmark acquisition, and repeat navigation. Drug delivery was deferred to isolate navigation performance.

We conducted three trials in a phantom lung, each targeting a different airway region (Fig. 3a). During the Exploratory Phase, an expert bronchoscopist navigated the bronchoscope to predefined targets in lobes, simulating BAL. Landmark locations were recorded at each branching level. After bronchoscopy, captured frames were processed into feature maps to construct a visual vocabulary.

We first evaluated our pose prediction network using this data set, comparing it with state-of-the-art methods. The 6D ground truth pose was collected using the Aurora electromagnetic (EM) tracking system (NDI, Canada). We report the relative pose error (RPE), which measures frame-to-frame accuracy, and the absolute trajectory error (ATE), which assesses cumulative drift over entire trajectories. Table 1 summarizes the results, showing that our ViT-based model achieves the lowest translation and rotation errors among the tested methods.

Next, we evaluated the NB framework as a whole during the Navigation Phase, where the bronchoscope was guided toward predefined targets from the exploratory phase. To simulate realistic conditions, the phantom lung was covered, preventing operators from seeing the lung anatomy. They relied solely on

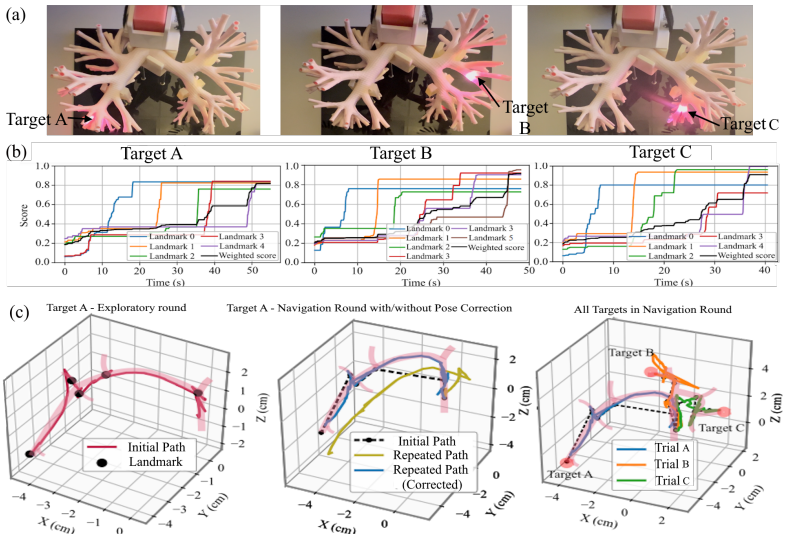


Fig. 3. (a) Phantom lung model at three airway target locations. The Target locations identified during exploratory phase are shown. (b) Landmark scoring during bronchoscope navigation trials of the three phantom lung model experiments. (c) Pose prediction results for phantom trials. From left to right: Camera pose during exploration for target A; Camera pose during navigation, before and after pose correction for target A; Initial and repeated paths of all three trials.

the endoscopic camera feed and navigation feedback including corrected pose predictions and the navigation score in Eq. (5). Previously identified landmarks were used to compute similarity scores, with a threshold of 0.7 indicating sufficient proximity for pose refinement.

Fig. 3b shows that similarity scores exceeded the threshold at multiple time points, ensuring reliable landmark identification at each bifurcation. Pose tracking results (Fig. 3c) illustrate camera pose tracking and trajectory refinement, showing the initial bronchoscope path and recorded landmarks. The results demonstrate that landmark-based pose correction significantly improves pose prediction in repeated trials, enabling the operator to consistently reach all predefined targets. Apparent off-lung paths in Fig. 3c and Fig. 4c are artifacts from accumulated drift during pose integration, not navigation errors, as the overall trajectory remains correct and is refined by landmark-based correction.

To assess system performance in a more realistic setting, we conducted three trials in a mechanically ventilated ex-vivo human lung, each targeting a different airway location in right lobe (Fig. 4a). A small bright fluosphere was deposited near each target location (6th to 8th airway generations) to confirm successful navigation in subsequent rounds. The trials followed the same three-phase navigation procedure as in the phantom experiments. In all experiments, the NB framework successfully guided a novice operator toward target locations. Fig. 4b

shows the computed scores for the Navigation Phase of each trial, demonstrating the system's ability to track progress toward intended targets. Similarity scores exceeding 0.7 (as in phantom trials) were used to refine camera pose, with corrected trajectories shown in Fig. 4c.

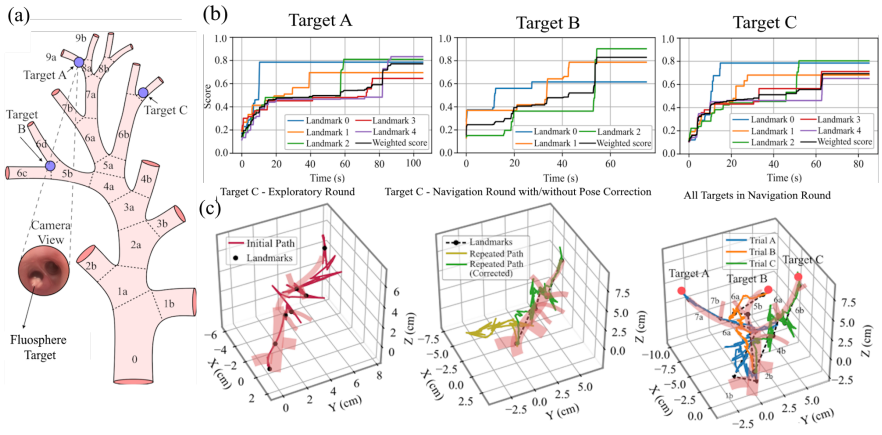


Fig. 4. (a) Diagram of the ex-vivo lung segment explored during navigation trials, with numbers indicating airway branching levels. Blue dots represent target locations (fluospheres) at different branches. (b) Landmark scoring during navigation trials for three ex-vivo lung experiments. (c) Pose prediction results for phantom trials. From left to right: Camera pose during exploration for target C; Camera pose during navigation before and after pose correction for target A; Initial and repeated paths for all trials.

Comparing the two experimental settings, bronchoscopy in the ex-vivo lung was more challenging due to dynamic lung tissue and occlusions from secretions. This is evident in varying locations of several branches (e.g., 1b and 6a) across trials. Landmark similarity scores were generally higher in the phantom lung, likely due to lower visual variability. In the ex-vivo lung, ventilation-induced movement, tissue secretions, and camera smudging affected performance. Further tuning of the navigation score may be needed for different clinical scenarios or datasets. Future work could explore alternative scoring or post-processing techniques to unify similar landmarks across trials, enhancing robustness.

Pose tracking accuracy could not be directly validated with ground truth in the ex-vivo model due to interference between the lung ventilator and the magnetic tracker. However, inconsistencies on the trajectory scale suggest a discrepancy between training data and the bronchoscopy sequences tested. In future work, to mitigate cumulative drift, shared landmark registration across trials could be implemented to construct a graph of visited landmarks for pose optimization. This approach may enhance trajectory accuracy and support multi-session navigation in real clinical applications.

4 Conclusion

This study introduced a novel NB framework designed to overcome key challenges in bronchoscopy, including reliance on preoperative imaging, sensor limitations, and the inherent difficulty of navigating feature-sparse airway environments. By leveraging vision foundation models, specifically adopted for (i) pose regression, and (ii) deep image feature mapping, our method enables real-time bronchoscope tracking and guidance for repeatable interventions. The results of our ex-vivo human lung trials have validated the practical feasibility of our NB framework for real-world applications in mechanically ventilated lungs. Future work will focus on refining and expanding the NB framework. Key areas for improvement include implementing automatic landmark detection, refining landmark registration and adopting pose graph optimization techniques for more robust and accurate navigation. Additionally, we plan to further investigate failure cases and implement adaptive hyperparameters to better handle anatomical variability beyond the current static phantom-derived settings.

Acknowledgments. This work was supported by the UKRI Medical Research Council (MR/T023252/1) and the Baillie Gifford Pandemic Science Hub.

Disclosure of Interests. The authors declare no conflicts of interest in this paper.

References

1. 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 (Jan 2023). https://doi.org/10.1016/j.cmpb.2022.107241
2. Bregier, R.: Deep regression on manifolds: A 3D rotation case study. In: 2021 International Conference on 3D Vision (3DV). pp. 166–174. IEEE Computer Society, Los Alamitos, CA, USA (dec 2021)
3. Caron, M., Touvron, H., Misra, I., Jegou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 9630–9640 (Apr 2021)
4. 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 (2023). https://doi.org/10.1109/IROS55552.2023.10342034
5. Dunn, B.K., Blaj, M., Stahl, J., Speicher, J., Anciano, C., Hudson, S., Kragel, E.A., Bowling, M.R.: Evaluation of electromagnetic navigational bronchoscopy using tomosynthesis-assisted visualization, intraprocedural positional correction and continuous guidance for evaluation of peripheral pulmonary nodules. Journal of Bronchology & Interventional Pulmonology **30**(1), 16–23 (Jan 2023)
6. Farhat, M., Chaabouni-Chouayakh, H., Ben-Hamadou, A.: Self-supervised endoscopic image key-points matching. Expert Systems with Applications **213**, 118696 (2023). https://doi.org/10.1016/j.eswa.2022.118696

7. Gálvez-López, D., Tardós, J.D.: Bags of binary words for fast place recognition in image sequences. IEEE Transactions on Robotics **28**(5), 1188–1197 (October 2012). https://doi.org/10.1109/TRO.2012.2197158
8. Kamel, T., Helms, J., Janssen-Langenstein, R., Kouatchet, A., Guillon, A., Bourenne, J., Contou, D., Guervilly, C., Coudroy, R., Hoppe, M.A., Lascarrou, J.B., Quenot, J.P., Colin, G., Meng, P., Roustan, J., Cracco, C., Nay, M.A., Boulain, T., Clinical Research in Intensive Care Sepsis Group (CRICS-TRIGGERSEP): Benefit-to-risk balance of bronchoalveolar lavage in the critically ill. A prospective, multicenter cohort study. Intensive Care Medicine **46**(3), 463–474 (Mar 2020)
9. Li, L., Li, X., Yang, S., Ding, S., Jolfaei, A., Zheng, X.: Unsupervised-learning-based continuous depth and motion estimation with monocular endoscopy for virtual reality minimally invasive surgery. IEEE Transactions on Industrial Informatics **17**(6), 3920–3928 (2021). https://doi.org/10.1109/TII.2020.3011067
10. Liu, X., Stiber, M., Huang, J., Ishii, M., Hager, G.D., Taylor, R.H., Unberath, M.: Reconstructing sinus anatomy from endoscopic video – towards a radiation-free approach for quantitative longitudinal assessment. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. pp. 3–13 (2020). https://doi.org/10.1007/978-3-030-59716-0_1
11. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision **60**(2), 91–110 (Nov 2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94
12. Mackutė, E., Abdalla, A., Dickson, S., Dhaliwal, K., Khadem, M.: On challenges of monocular pose estimation for endoluminal navigation. Journal of Medical Robotics Research **09**(03n04), 2440009 (2024)
13. Martin-Loeches, I., Chastre, J., Wunderink, R.G.: Bronchoscopy for diagnosis of ventilator-associated pneumonia. Intensive Care Medicine **49**(1), 79–82 (Sep 2022)
14. Miyawaki, S., Choi, S., Hoffman, E.A., Lin, C.L.: A 4DCT imaging-based breathing lung model with relative hysteresis. Journal of Computational Physics **326**, 76–90 (Dec 2016)
15. Mur-Artal, R., Montiel, J.M.M., Tardós, J.D.: ORB-SLAM: A versatile and accurate monocular SLAM system. IEEE Transactions on Robotics **31**(5), 1147–1163 (Aug 2015). https://doi.org/10.1109/TRO.2015.2463671
16. Ozyoruk, K.B., Gokceler, G.I., Bobrow, T.L., Coskun, G., Incetan, K., Almalioglu, Y., Mahmood, F., Curto, E., Perdigoto, L., Oliveira, M., Sahin, H., Araujo, H., Alexandrino, H., Durr, N.J., Gilbert, H.B., Turan, M.: EndoSLAM dataset and an unsupervised monocular visual odometry and depth estimation approach for endoscopic videos. Medical Image Analysis **71**, 102058 (Jul 2021). https://doi.org/10.1016/j.media.2021.102058
17. Pire, T., Fischer, T., Castro, G., De Cristóforis, P., Civera, J., Jacobo Berlles, J.: S-PTAM: Stereo parallel tracking and mapping. Robotics and Autonomous Systems **93**, 27–42 (Jul 2017). https://doi.org/https://doi.org/10.1016/j.robot.2017.03.019
18. Reisenauer, J., Simoff, M.J., Pritchett, M.A., Ost, D.E., Majid, A., Keyes, C., Casal, R.F., Parikh, M.S., Diaz-Mendoza, J., Fernandez-Bussy, S., Folch, E.E.: Ion: Technology and techniques for shape-sensing robotic-assisted bronchoscopy. The Annals of Thoracic Surgery **113**(1), 308–315 (Jan 2022)
19. Rojas-Solano, J.R., Ugalde-Gamboa, L., Machuzak, M.: Robotic bronchoscopy for diagnosis of suspected lung cancer: A feasibility study. Journal of Bronchology & Interventional Pulmonology **25**(3), 168–175 (2018). https://doi.org/10.1097/LBR.0000000000000499

20. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT or SURF. In: 2011 International Conference on Computer Vision. pp. 2564–2571 (2011). https://doi.org/10.1109/ICCV.2011.6126544
21. Saghaie, T., Williamson, J.P., Phillips, M., Kafili, D., Sundar, S., Hogarth, D.K., Ing, A.: First-in-human use of a new robotic electromagnetic navigation bronchoscopic platform with integrated tool-in-lesion tomosynthesis (TiLT) technology for peripheral pulmonary lesions: The FRONTIER study. Respirology **29**(11), 969–975 (2024). https://doi.org/https://doi.org/10.1111/resp.14778
22. Sganga, J., Eng, D., Graetzel, 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 (May 2019)
23. Tian, Q., Chen, Z., Liao, H., Huang, X., Yang, B., Li, L., Liu, H.: PANS: Probabilistic Airway Navigation System for Real-time Robust Bronchoscope Localization . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. vol. LNCS 15006. Springer Nature Switzerland (October 2024)
24. Wang, C., Oda, M., Hayashi, Y., Kitasaka, T., Honma, H., Takabatake, H., Mori, M., Natori, H., Mori, K.: Improved visual SLAM for bronchoscope tracking and registration with pre-operative CT images. In: Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling. vol. 11315, pp. 324–329 (2020)
25. Xu, Y., Feng, L., Xia, Z., Xiong, J.: Camera pose estimation based on feature extraction and description for robotic gastrointestinal endoscopy. In: Intelligent Robotics and Applications. vol. 13015, pp. 113–122 (2021)