Automated Scope Pose Estimation in Real-world Robotic-Assisted Bronchoscopy with Integrated C-Arm Imaging

Dingzhong Zhang[1*], Hedyeh Rafii-Tari[1], and Babak Matinfar[1]

Researchs&Development, Johnson&Johnson MedTech Surgery, Santa Clara, CA, USA
{dzhan111, hrafiita, mmatinfa}@its.jnj.com

Abstract. Robotic-assisted bronchoscopy enables physicians to access pulmonary lesions with enhanced stability and precision; however, navigation accuracy can be compromised by CT-to-body divergence due to respiratory motion, anatomical deformation during the procedure and tracking system errors. Cone-beam computed tomography (CBCT) has therefore emerged as a key intra-operative imaging modality for biopsy verification and navigation updates. A critical prerequisite for integrating CBCT into the workflow is accurate registration between the navigation system and the CBCT coordinate frame, which relies on precise bronchoscope pose estimation. Existing semi-automated solutions require manual scope tip localization and potential correction of pose parameters, introducing variability and prolonging procedure time. In this work, we propose a fully automated pipeline that estimates the bronchoscope pose directly from CBCT images, which consists of: (1) a two-stage model inferencing strategy for rapid and accurate scope tip segmentation, and (2) a context-aware robust pose estimation algorithm. Evaluated on 25 diverse real-world clinical cases, the method achieved a mean position error of 0.55 ± 0.28 mm and orientation error of $2.74\pm1.47°$, with an average processing time of 4.47 s, approximately $20\times$ faster than the current approach. This combination of accuracy and efficiency enables safer and more reliable navigation updates while reducing clinician workload and minimizing user-induced errors.

Keywords: Robotic-assisted Bronchoscopy · C-Arm Integration · Bronchoscope Pose Estimation · Surgical Navigation · Image Segmentation.

1 Introduction

Robotic-assisted bronchoscopy (RAB) has emerged as an effective modality for diagnosing and treating pulmonary lesions; however, navigation accuracy is often degraded by CT-to-body divergence caused by respiratory motion, anatomic deformation during the procedure and tracking system errors such as electromagnetic distortion [4, 5, 7, 13]. As a result, substantial discrepancies between pre-operative plans and true intra-operative lesion locations are frequently observed, leading to reduced tool-in-lesion rates and diagnostic yield if not corrected intraoperatively [4,5,14]. Cone-beam computed tomography (CBCT) has become a critical intra-operative imaging modality in RAB, providing real-time 3D volumetric information for tool-in-lesion confirmation and navigation updates [4,8,9]. Clinical studies have shown that integrating CBCT with RAB systems significantly improves navigation success and diagnostic yield [1–7, 10–12]. RAB systems incorporate CBCT to update navigation by registering the CBCT coordinate frame with the tracking system. Accurate estimation of the bronchoscope pose, including both position and orientation, from CBCT images is therefore a prerequisite for computing a reliable transformation matrix. However, current semi-automated workflows require bronchoscopists to manually localize the scope tip and refine pose parameters when inaccuracies are present, a process that is time-consuming, operator-dependent, and susceptible to inter-operator variability and user-induced error [15–19].

To address these limitations, we propose a fully automated workflow for fast and accurate bronchoscope pose estimation from CBCT images. The proposed workflow comprises two key components: (1) a two-stage inferencing strategy for precise scope tip segmentation, and (2) a robust pose estimation algorithm that leverages contextual information from both the segmented scope geometry and surrounding anatomical structures. Unlike prior CBCT-guided RAB workflows that require manual or semi-automated scope pose identification, our approach eliminates manunal intervention while remaining robust to metal- and motion-induced artifacts, indistinct scope boundaries, and working channel instruments. The method is validated on a diverse set of real-world RAB cases acquired during routine clinical procedures. Quantitative results demonstrate high accuracy and substantially reduced pose estimation time, thereby improving procedural reliability, efficiency, and scalability for CBCT-guided robotic bronchoscopy.

2 Methods

2.1 Scope Tip Segmentation

Segmenting the scope tip from the bronchoscopic shaft is challenging due to their continuous physical structure and shared metallic composition, which produce similarly high image intensities. This task is further complicated by artifacts, working channel instruments, and over-segmentation caused by metallic implants in real-world patient CBCT. To address these challenges, we propose a two-stage inferencing strategy, as illustrated in Fig. 1.

In the first stage, the CBCT volume is resampled to a lower resolution and processed by a coarse segmentation model to localize the scope tip. The resulting mask is used to estimate the scope tip center, which defines the center of a volume of interest (VOI). In the second stage, the original CBCT image is cropped to this VOI and provided as input to a fine segmentation model for precise scope tip delineation. Mirror padding is applied to the cropped image when necessary.

The two models in the proposed inferencing strategy share the same architecture of U-Net with residual encoders [20], which has been shown to outperform Transformer-based and Mamba-based networks across a range of medical image

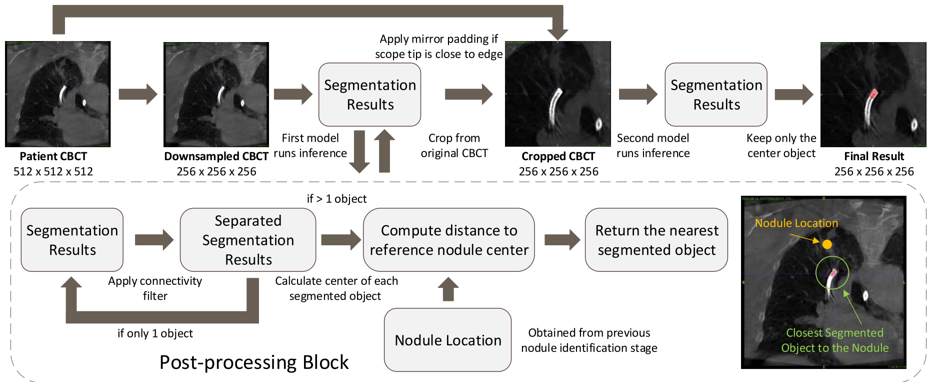


Fig. 1. Workflow of the two-stage inferencing strategy with post-processing steps.

segmentation tasks, while having relatively low GPU VRAM cost and shorter runtime compared with other CNN-based architectures [21]. While processing fewer patches on smaller volumes enables both models to effectively reduce isolated false positives due to metal implants, additional post-processing is employed to ensure clean final segmentations. Specifically, the first stage retains only the connected component nearest to the target nodule, whereas the second stage preserves only the centrally located segmented object.

2.2 Scope Pose Estimation

Scope Heading Estimation To compute the scope heading vector, principal component analysis (PCA) is first applied to the segmented scope tip mask to extract the direction of maximum variance. The first principal component is interpreted as the principal axis of the scope tip. Since this axis is bidirectional, a context-based sampling strategy is employed to resolve the correct heading, leveraging the fact that the bronchoscope exhibits substantially higher CBCT intensity than surrounding anatomical structures. As illustrated in Fig. 2(a), two sampling spheres are placed along the principal axis. Within each sphere, a set of sampling points is generated, and the voxel intensities at these points are accumulated. The sphere with the lower total intensity is identified as containing the distal tip of the scope, thereby determining the heading direction.

While computing relative intensity differences between the two sampling regions helps reduce the impact of global intensity variations across scans, real-world CBCT images may exhibit noises, working channel instruments, metal and motion artifacts that can obscure scope boundaries and reduce inter-region contrast. To enhance robustness, sampling points are generated using a spherical coordinate grid with uniform azimuthal and polar angle sampling, yielding a well-distributed point set, as shown in Algorithm 1. Sampling parameters, including sphere radius, center offset, and number of points, are derived from the bronchoscope tip CAD model and selected to maximize contrast.

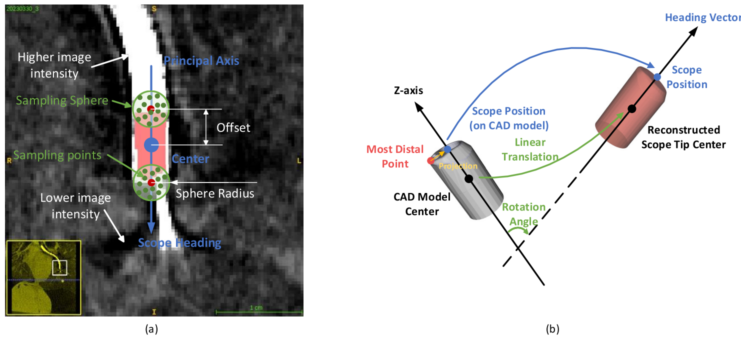


Fig. 2. Scope pose estimation. (a) Scope heading estimation using a context-based sampling method. (b) Scope position estimation using a registration-based method.

Algorithm 1 Spherical Coordinate Grid with Uniform Volume Density

---

Input: N (sampling points), R (sphere radius) **Definitions:** n_r (radial steps), n_θ (polar steps), n_ϕ (azimuthal steps); i (radial index), j (polar index), k (azimuthal index); r_i (radius at layer i), θ_j (polar angle), ϕ_k (azimuth angle); $p_{i,j,k} = (x, y, z)$ (Cartesian point) **Procedure:**

1. Set $n_r = \text{round}(N^{1/3})$, $n_\theta = 2n_r$, $n_\phi = 2n_r$
2. **for** $i = 1, \ldots, n_r$ **do** $r_i = R\left(\frac{i}{n_r}\right)^{1/3}$ (cube-root for uniform volume density)
3. **for** $j = 0, \ldots, n_\theta - 1$ **do** $\theta_j = \pi\frac{j+0.5}{n_\theta}$ (offset avoids pole clustering)
4. **for** $k = 0, \ldots, n_\phi - 1$ **do** $\phi_k = 2\pi\frac{k}{n_\phi}$
5. Compute $(x, y, z) = (r_i \sin\theta_j \cos\phi_k,\ r_i \sin\theta_j \sin\phi_k,\ r_i \cos\theta_j)$
6. Emit $p_{i,j,k} = (x, y, z)$
7. Return $\{p_{i,j,k}\}$

Scope Position Estimation The scope tip position is defined as the most distal point projected onto the principal axis. Directly extracting this point from the segmention mask can be challenging due to complex tip geometry and potential segmentation inaccuracies. To improve robustness, a registration-based method is employed that utilizes the known geometry of the bronchoscope tip CAD model. Specifically, the CAD model is rigidly aligned with the reconstructed scope tip by rotating its Z-axis to be parallel with the estimated heading vector and translating its center to match the center of the reconstructed scope tip. The most distal point of the aligned CAD model is then projected onto the principal axis, yielding the estimated scope tip position, as shown in Figure 2(b).

3 Data Preparation

A total of 225 patient CBCT scans were collected during the robotic bronchoscopy procedures using MONARCH™ Platform across multiple hospitals (the dataset meets HIPAA compliance). All scans were acquired using GE OCE 3D

CBCT systems and reconstructed at a resolution of $512 \times 512 \times 512$. The scope tip segmentation ground truth was generated using a convex hull representation of the scope tip CAD model, as sub-surface structures cannot be captured by CBCT image. The convex hull model was aligned to the scope tip in CBCT using the semi-automated workflow from the MONARCH™ Platform and then manually refined to achieve precise boundary alignment. The finalized model was converted into a binary segmentation mask. Ground truth for scope heading and position was derived from physician operation logs recorded by the intra-operative planning software, which served as the reference standard for pose estimation evaluation. All ground truth data were reviewed and validated by an experienced clinical team.

4 Experiments and Results

4.1 Scope Tip Segmentation

Among the 225 patient CBCT scans, 200 were used for training and validation and 25 for testing. For the first-stage model, CBCT volumes were resampled to $256 \times 256 \times 256$ using linear interpolation, while segmentation masks were resampled using nearest-neighbor interpolation. For the second stage, a $256 \times 256 \times 256$ VOI centered on the scope tip was cropped from the original CBCT volume, with mirror padding applied when necessary. All models were trained for 1000 epochs on dual RTX 4500 ADA GPUs (24 GB each) with data augmentation applied to improve generalization.

Table 1. Comparison of baseline U-Net performance using different loss functions.

Loss Function	Training Dataset		Test Dataset	
	Dice Score (%)	False Positive	Dice Score (%)	False Positive
Focal	89.74	300.59	80.82	1258.48
Dice + Focal	92.15	303.12	80.13	1477.40
Focal-Tversky	91.68	408.18	80.25	1618.80
CE + Focal-Tversky	91.65	164.93	79.51	1210.88
Dice + CE	92.71	219.96	81.16	1176.04
Dice + TopK10	**93.52**	**216.57**	**83.01**	**917.68**
Dice + TopK10 + Focal	93.40	228.56	80.65	1368.76
Dice + TopK10 + CE	93.23	234.69	80.84	1335.36

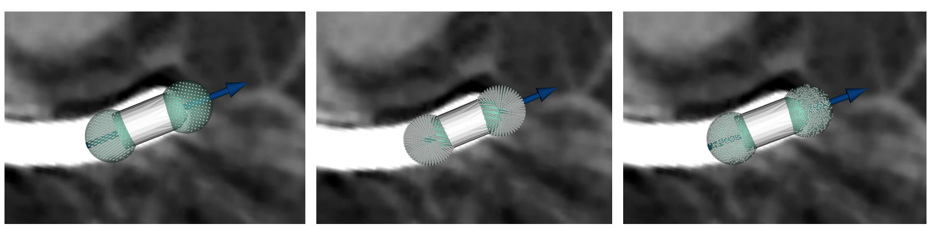
Loss functions were first evaluated using a baseline U-Net (Table 1). Dice+CE and Dice+TopK10 achieved higher Dice scores and lower false positives. Therefore, the first-stage model employed Dice+CE to emphasize global localization accuracy, while the second-stage model used Dice+TopK10 to specifically address challenging cases. We then compared different inferencing strategies using

Table 2. Comparison of model performance across different inferencing strategies before and after post-processing. Results are averaged over 25 real-world RAB cases.

Model	Original Results		Post-processed Results	
	Dice Score (%)	False Positive	Dice Score (%)	False Positive
First-stage	87.07	21.88	87.89	15.24
Second-stage	**86.11**	**348.16**	**89.26**	**102.12**
Single-stage	84.78	621.96	/	/
Cascaded	84.90	530.44	/	/

a large-scale U-Net with residual encoders to further enhance segmentation performance. As shown in Table 2, the proposed two-stage strategy outperformed single-stage full resolution inference, achieving the highest Dice score of 89.26% and the lowest false positive of 102.12 after post-processing. The approach also outperformed cascaded U-Net, which similarly adopt a coarse-to-fine segmentation strategy. Because in cascaded U-Net, the first-stage prediction is used directly to train the second-stage model, allowing errors to propagate and degrade the final performance. Notably, the two-stage approach achieved a higher Dice score and a lower false positive than other strategies even before post-processing due to the smaller, scope-centered input volume.

4.2 Scope Heading Identification



(a) (b) (c)

Fig. 3. Scope heading identification using context-based sampling methods. (a) Uniform Cartesian grid sampling. (b) Spherical coordinate grid sampling. (c) Spherical Fibonacci lattice sampling.

The proposed spherical coordinate grid sampling was compared with two alternatives: a uniform Cartesian grid and spherical Fibonacci lattices (Fig. 3). The Cartesian grid samples points uniformly along the X-, Y-, and Z-axes within the sphere, whereas the spherical Fibonacci lattice [22] uses golden-angle angular spacing with randomized radial distances to achieve quasi-uniform volumetric coverage. All methods were evaluated under identical settings: a 6.5 mm offset between the sphere and scope tip centers, a sphere radius of 3 mm, and 2000

Table 3. Comparison of sampling strategies for scope heading identification. Results are averaged over 25 real-world RAB cases. High/Low denote spheres with higher/lower accumulated voxel intensity.

Sampling Method	High ($\times 10^6$)	Low ($\times 10^6$)	Diff. ($\times 10^6$)	Time (ms)
Uniform Cartesian Grid	1.02	-0.03	1.04	**0.11**
Spherical Coordinate Grid	19.12	1.42	**17.70**	1.60
Spherical Fibonacci Lattices	4.35	0.13	4.22	0.65

sampling points, based on the scope tip size of approximately 7.8 mm in height and 4.4 mm in diameter. As summarized in Table 3, the spherical coordinate grid produced the largest inter-sphere intensity difference, with an average of 17.70×10^6 intensity units, substantially outperforming the Cartesian grid and Fibonacci lattice methods. The larger intensity contrast enables more robust distinction between the distal and proximal ends of the scope. All sampling strategies incurred negligible computational overhead and satisfied real-time requirements.

4.3 Scope Pose Estimation

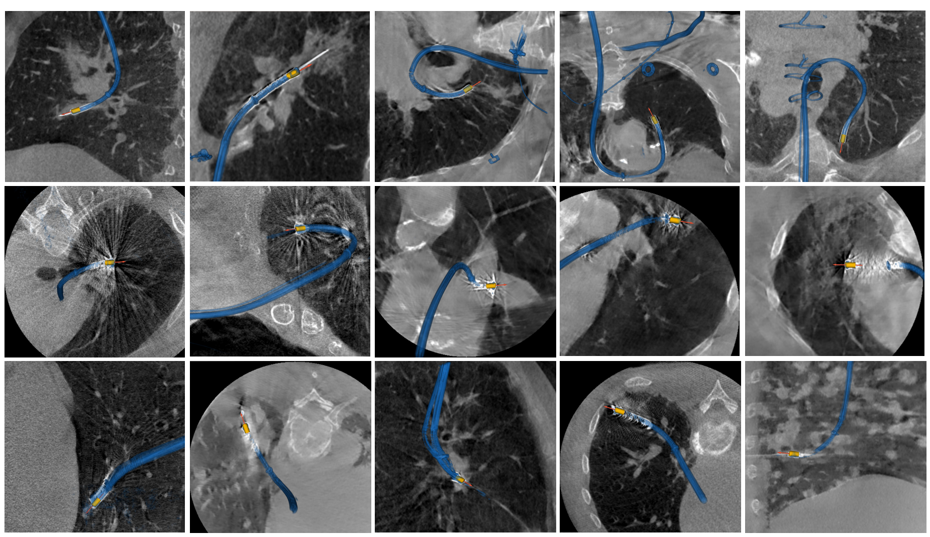


Fig. 4. Scope pose estimation results for normal cases with metal implants, working channel instruments (top row), and challenging cases with metal- and motion-induced artifacts (second and third rows). The pose is shown using the scope tip CAD model with an arrow indicating the heading; the bronchoscope is volume-rendered in blue.

Fig. 4 presents qualitative results for scope pose estimation under both normal and challenging conditions, including working channel instruments, metal implants, metal- and motion-induced artifacts, and indistinct scope boundaries. The proposed method consistently identified the correct scope position and heading across all cases, demonstrating strong robustness to perturbations commonly encountered in intra-operative CBCT.

Table 4. Comparison of scope pose estimation strategies. Results are averaged over 25 real-world RAB cases.

Strategy	Position Error (mm)	Heading Error (°)	Total Time (s)
Single-stage Inference	0.58 ± 0.31	3.12 ± 1.64	27.62
Two-stage Inference	$\mathbf{0.55 \pm 0.28}$	$\mathbf{2.74 \pm 1.47}$	**4.47** (2.26 + 2.21)

Quantitative results are summarized in Table 4. The two-stage strategy achieved a mean position error of 0.55 ± 0.28 mm and an orientation error of $2.74 \pm 1.47°$, outperforming the single-stage approach (0.58 ± 0.31 mm and $3.12 \pm 1.64°$). This improved accuracy is attributed to the higher-quality segmentation produced by the two-stage pipeline, which preserves fine scope-tip details while suppressing false positives. Additionally, both with the acceleration of NVIDIA TensorRT and parallel computation, the two-stage method reduced average processing time to 4.47 s, compared with 27.62 s for single-stage inference. When integrated into the clinical workflow, this translates to a reduction of 15% total average navigation time from previously reported 9.8 min [4] to 8.3 min, underscoring the clinical feasibility for real-time robotic bronchoscopy.

5 Conclusion and Discussion

In this work, we presented a fully automated workflow for precise and real-time bronchoscope pose estimation during robotic-assisted bronchoscopy with C-arm integration. By combining a two-stage inferencing strategy for accurate scope tip segmentation with a context-aware pose estimation algorithm, the proposed approach eliminates manual intervention for CBCT-guided navigation updates, which are time-consuming and prone to inter-operator variability. Evaluation on real-world clinical CBCT data demonstrates that the method achieves sub-millimeter position accuracy and low orientation error, while reducing total processing time to under 4.5 s, approximately 20× faster than existing semi-automated approach. The two-stage strategy achieves both high accuracy and efficiency by suppressing over-segmentations in challenging scenarios and minimizing computational cost. In addition, the proposed context-based heading determination and registration-based position estimation methods show strong robustness to metal artifacts, motion-induced distortions, and working channel instruments commonly encountered in intra-operative CBCT. While this study focuses on bronchoscope pose estimation for navigation updates, the framework is generalizable to other interventional devices and CBCT-guided procedures, enabling practical and scalable real-time C-arm-integrated navigation.

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

References

1. Liu, Y., Seides, B.J., Rowell, J.H., White, B., Kim, J., Greenhill, S., Johnson, J., Harvey, J., Guiance, I.R., Smith, S.B., Wahidi, M.M., Kapp, C.: Intraprocedural cone-beam CT during robotic-assisted bronchoscopy improves diagnostic yield and sensitivity for malignancy. CHEST 168(4), A5926-A5927 (2025). https://doi.org/10.1016/j.chest.2025.07.3321
2. Shaller, B.D., Duong, D.K., Swenson, K.E., Free, D.R., Bedi, H.: Added value of a robotic-assisted bronchoscopy platform in cone-beam computed tomography-guided bronchoscopy for the diagnosis of pulmonary parenchymal lesions. Journal of Bronchology & Interventional Pulmonology 31(3), e0971 (2024). https://doi.org/https://doi.org/10.1097/LBR.0000000000000971
3. Chen, J., McNierney, D., Zein, J.G., Vaszar, L.T., Swanson, K.L., Azadeh, N., Sakata, K.K.: Diagnostic performance of Monarch robotic bronchoscopy with and without mobile cone-beam CT support. Lung Cancer 213, 108910 (2026). https://doi.org/10.1016/j.lungcan.2026.108910
4. Cumbo-Nacheli, G., Velagapudi, R. K., Enter, M., Egan, J. P., 3rd, Conci, D.: Robotic-assisted Bronchoscopy and Cone-beam CT: A Retrospective Series. Journal of bronchology & interventional pulmonology **29**(4), 303–306 (2022). https://doi.org/10.1097/LBR.0000000000000860
5. Ho, A. T. N., Ma, X., Christos, P. J., Pedoto, A., Lee, R. P., Oberg, C. L., Rakočević, R., Bott, M. J., Chawla, M., Husta, B. C., Kalchiem-Dekel, O.: Virtual vs. reality in cone-beam CT-guided robotic-assisted bronchoscopy. Journal of thoracic disease **17**(10), 9051–9062 (2025). https://doi.org/10.21037/jtd-2025-1399
6. Abdelghani, R., Espinoza, D., Uribe, J., Becnel, D., Herr, R., Villalobos, R., Kheir, F.: Cone-beam computed tomography-guided shape-sensing robotic bronchoscopy vs. electromagnetic navigation bronchoscopy for pulmonary nodules. Journal of Thoracic Disease **16**(9), 5529–5538 (2024). https://doi.org/10.21037/jtd-24-178
7. Bashour, S. I., Khan, A., Song, J., Chintalapani, G., Kleinszig, G., Sabath, B. F., Lin, J., Grosu, H. B., Jimenez, C. A., Eapen, G. A., Ost, D. E., Sarkiss, M., Casal, R. F.: Improving Shape-Sensing Robotic-Assisted Bronchoscopy Outcomes with Mobile Cone-Beam Computed Tomography Guidance. Diagnostics **14**(17), 1955 (2024). https://doi.org/10.3390/diagnostics14171955
8. Duke, J.D., Sanborn, D., Reisenauer, J.: Enhancing nodule biopsy through technology integration. Innovations, 15569845231153639 (2023). https://doi.org/10.1177/15569845231153639
9. Ortiz-Jaimes, G., Reisenauer, J.: Real-world impact of robotic-assisted bronchoscopy on the staging and diagnosis of lung cancer: the shape of current and potential opportunities. Precision Oncology Reviews, 75-94 (2023). https://doi.org/10.2147/POR.S395806

10. Brown, M. V., Badiei, A., Arnold, M., Jersmann, H., Sullivan, T., Fielding, D., Nguyen, P.: The Diagnostic Yield of Cone Beam CT Combined With Radial-Endobronchial Ultrasound for the Diagnosis of Peripheral Pulmonary Nodules: Systematic Review and Meta-Analysis. CHEST Pulmonary **2**(2), 100037 (2024). https://doi.org/10.1016/j.chpulm.2024.100037
11. Fernandez-Bussy, S., Chandra, N., Koratala, A., Lee-Mateus, A., Barrios-Ruiz, A., Garza-Salas, A., Koirala, T., Funes-Ferrada, R., Balasubramanian, P., Patel, N., Chadha, R., Hazelett, B., Robertson, K., Reisenauer, J., Abia-Trujillo, D.: Robotic-assisted bronchoscopy: a narrative review of systems. Journal of Thoracic Disease 16(8), 5422-5434 (2024). https://doi.org/0.21037/jtd-24-456
12. Prado, R.M.G., Cicenia, J., Almeida, F.A.: Robotic-assisted bronchoscopy: a comprehensive review of system functions and analysis of outcome data. Diagnostics 14(4), 399 (2024). https://doi.org/10.3390/diagnostics14040399
13. Chen, A. C., Pastis, N. J., Machuzak, M. S., Gildea, T. R., Simoff, M. J., Gillespie, C. T., Mahajan, A. K., Oh, S. S., Silvestri, G. A.: Accuracy of a Robotic Endoscopic System in Cadaver Models with Simulated Tumor Targets: ACCESS Study. Respiration **99**(1), 56–61 (2020). https://doi.org/10.1159/000504181
14. Pritchett, M. A., Schampaert, S., de Groot, J. A. H., Schirmer, C. C., van der Bom, I.: Cone-beam CT with augmented fluoroscopy for transbronchial biopsy. Journal of Bronchology & Interventional Pulmonology **25**(4), 274–282 (2018). https://doi.org/10.1097/LBR.0000000000000536
15. Bhadra, K., Rickman, O.B., Mahajan, A.K., Hogarth, D.K.: "Tool-in-lesion" accuracy of Galaxy System—A robotic electromagnetic navigation bronchoscopy with integrated tool-in-lesion tomosynthesis technology (The MATCH Study). Journal of Bronchology & Interventional Pulmonology 31(1), 23-29 (2024). https://doi.org/10.1097/LBR.0000000000000923
16. Aboudara, M., Roller, L., Rickman, O., Lentz, R.J., Pannu, J., Chen, H., Maldonado, F.: Improved diagnostic yield for lung nodules with digital tomosynthesis-corrected navigational bronchoscopy: initial experience with a novel adjunct. Respirology 25(2), 206-213 (2020). https://doi.org/10.1111/resp.13609
17. Pritchett, M.A., Bhadra, K., Mattingley, J.S.: Electromagnetic navigation bronchoscopy with tomosynthesis-based visualization and positional correction: three-dimensional accuracy as confirmed by cone-beam computed tomography. Journal of Bronchology & Interventional Pulmonology 28(1), 10-20 (2021). https://doi.org/10.1097/LBR.0000000000000687
18. Leira, O., Hofstad, E.F., Bø, L.E., Langø, T., Amundsen, T.: Navigated bronchoscopy with electromagnetic tracking—cone-beam computed tomography influence on tracking and registration accuracy. Journal of Bronchology & Interventional Pulmonology 18(4), 329-336 (2011). https://doi.org/10.1097/LBR.0b013e318235d471
19. Bhadra, K., Setser, R.M., Condra, W., Bader, B.A., David, S.: A cone-beam CT bronchoscopy study of the ultrathin cryoprobe for biopsy of peripheral lung lesions. Journal of Bronchology & Interventional Pulmonology 31(2), 117-125 (2024). https://doi.org/10.1097/LBR.0000000000000936
20. Isensee, F., Maier-Hein, K.: An attempt at beating the 3D U-Net. arXiv preprint arXiv:1908.02182 (2019). https://doi.org/10.48550/arXiv.1908.02182
21. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.: nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation. Medical Image Computing and Computer Assisted Intervention **15009**, 488-498 (2024). https://doi.org/10.1007/978-3-031-72114-4_47

22. González, Á.: Measurement of areas on a sphere using Fibonacci and latitude - longitude lattices. Mathematical Geosciences **42**, 49–64 (2010). https://doi.org/10.1007/s11004-009-9257-x