2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

SpReg: An Autonomous Image-to-Patient Registration Framework for Robotic Bronchoscopy

Shucheng Ye[1,†], Peibo Sun[1,†], Jiajun Ma[2], Wenbo She[1], Hongen Liao[1], Dingpei Han[3], Tianqi Huang[1], Fang Chen[1,*]

***Abstract*— Robotic bronchoscopy offers transformative potential for the precise diagnosis and treatment of pulmonary diseases, yet its clinical adoption is bottlenecked by the challenge of rapidly and accurately registering preoperative CT images to the patient's anatomy. Current methods, which rely on manual expert operation, are laborious and time-intensive with a steep learning curve, posing inherent risks to patients due to the lack of navigational support. Here, we present SpReg, an autonomous spatial registration framework that, for the first time, enables a robotic bronchoscope to perform autonomous driving and registration without manual intervention. The SpReg framework leverages a deep learning network that uniquely incorporates physicians' eye-tracking data as prior information. This allows the system to identify key anatomical regions, using this as a foundation for autonomous path planning. The system then drives the robot to autonomously navigate along multi-level bronchial centerlines, recording its three-dimensional path, which is subsequently aligned with the preoperative CT model to complete the registration. Simulation experiments demonstrate that the trajectory error of SpReg during motion is below 1.6 mm. In vivo experiments in a porcine model further show that, compared to manual operation, SpReg produces smoother motion trajectories and reduces navigation error by 19% (2.1 mm vs. 2.6 mm). Notably, its final registration accuracy shows no statistically significant difference from the manual method. These findings demonstrate that SpReg has the potential to substantially reduce the surgeon's workload while enhancing procedural safety and efficiency, paving the way for the development of more advanced human-robot collaborative intelligent surgical systems.**

I. INTRODUCTION

Pulmonary diseases are an important global health concern, with lung cancer being the most lethal due to its high prevalence and typically late-stage diagnosis [1]. Bronchoscopy is a foundational tool in pulmonary care, offering direct visualization and access for diagnostic and therapeutic procedures [2]. Recent innovations in minimally invasive techniques have enabled the development of robotic bronchoscopy [3]–[5], which provides superior dexterity and precision to navigate peripheral bronchial branches that are often inaccessible with conventional methods. By extending the reach to distal pulmonary nodules, robotic bronchoscopy holds significant promise for the early detection and targeted treatment of lung cancer [6]. Among them, electromagnetic navigated bronchoscopy (ENB) [7] is the most widely adopted, achieving three-dimensional real-time localization through integration of a distal electromagnetic sensor with an external field generator.

Precise localization in this technique relies on the accurate spatial registration between the CT and patient coordinate systems [8], which ensures reliable procedural guidance. Nevertheless, this registration step remains time-consuming, labor-intensive, and highly dependent on operator expertise. In clinical practice, centerline-based registration is commonly used [9]. The operator navigates an electromagnetic tracking (EMT) catheter through the main bronchial landmarks, including the main carina, the primary bronchi, and the lobar or segmental branches, and the resulting trajectory is aligned with the preoperative CT airway centerline [10], establishing correspondence between patient anatomy and imaging data. Centerline-based registration suffers from limited localization and registration accuracy, unstable trajectories, and substantial operator-dependent errors that collectively constrain its robustness and reliability. Wegner et al. [11] proposed a projection-based method to align the tip of the bronchoscope with the nearest bronchial centerline. While effective for peripheral navigation, it lacks the precision required for high-accuracy procedures such as biopsies [12]. Hautmann et al. [13] integrated EMT coils into the bronchoscope and registered the tracking data with the CT-derived airway model to improve access to peripheral lesions. However, the dependence on fiducial markers mounted on the chest adds to the procedural complexity. Mori et al. [14] introduced a markerless approach using an unlabeled electromagnetic tracker (UEMT), combining branch information with tracker orientation and incorporating branch radius as a normalization factor, effectively resolving branch-matching ambiguities at bifurcations. Building on this, Deguchi et al. [15] developed a real-time markerless framework that iteratively refines the CT-to-EMT transformation using multiple sets of EMT points and Powell optimization. Despite these advances, challenges persist regarding trajectory stability and reliance on manual operation.

In recent years, significant progress has been made in bronchoscopic registration due to the rapid advancement of artificial intelligence. Researchers have increasingly explored the integration of deep learning [16], [17] and reinforcement learning [18], [19], aiming to enable data-driven autonomous localization and path guidance. However, existing studies remain limited to sub-tasks such as path planning or position estimation, as challenges posed by respiratory motion and incomplete coverage of the traversed regions hinder

[1]School of Biomedical Engineering, Shanghai Jiao Tong University. [2]School of Biomedical Engineering, Tsinghua University, Beijing, China. [3]Department of Thoracic Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine.
[†]These authors contributed to this paper equally.
[*]Corresponding Author. chen-fang@sjtu.edu.cn

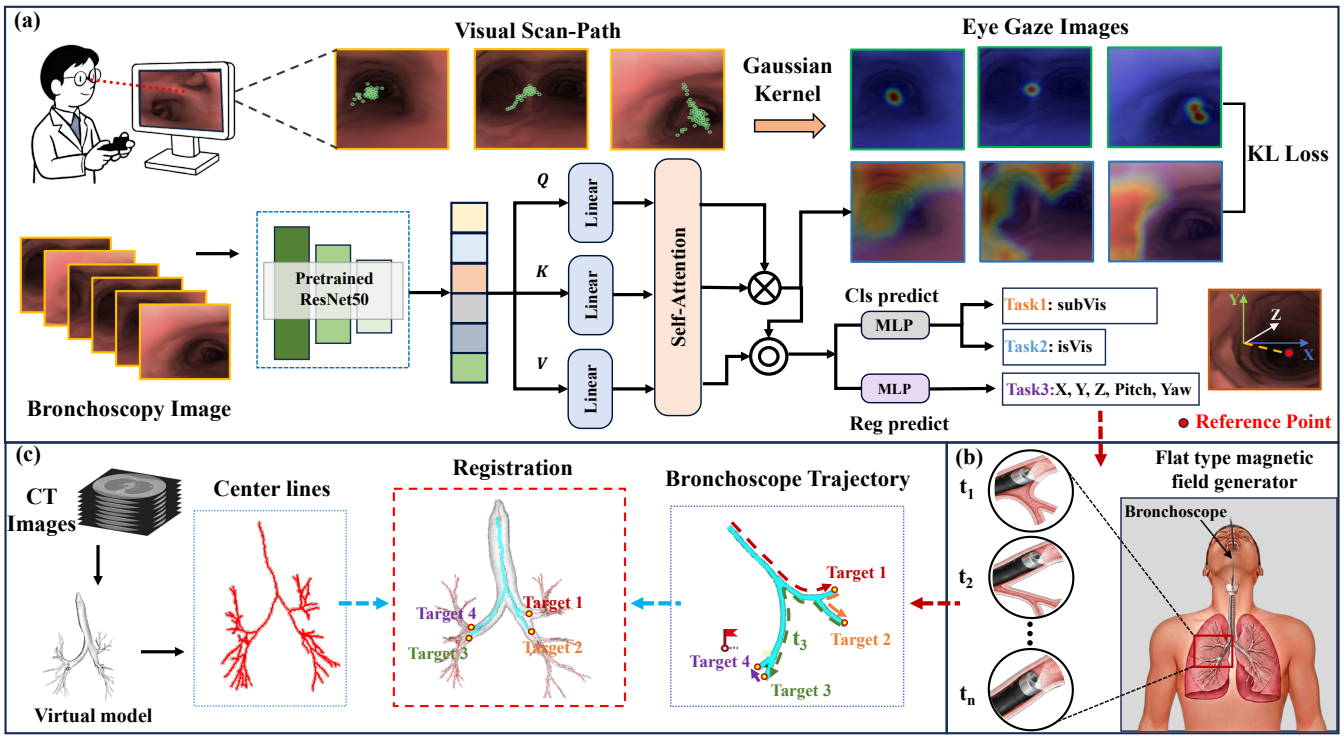

Fig. 1. Overview of the proposed SpReg. (a) The Autonomous Main Airway Path Traversal model integrates bronchoscope images and eye-tracking data to predict both classification tasks and regression tasks. A pretrained ResNet50 backbone extracts image features, which are fused with physician's eye-tracking data via a self-attention module. (b) During bronchoscope advancement, an EMT system continuously records the bronchoscope pose. (c) The registration process aligns EM-recorded bronchoscope trajectories with CT-derived airway centerlines, ultimately accomplishing the CT-to-patient registration.

the establishment of a unified, closed-loop framework for fully automated spatial registration. Furthermore, current autonomous navigation approaches often neglect the integration of domain-specific medical expertise and clinical experience, which are crucial for enhancing the interpretability of model decisions and improving both stability and accuracy [20], [21].

To address these challenges, this paper presents the first automated solution that achieves fully automatic CT-to-patient registration without manual intervention by physicians. By integrating the electromagnetic tracking system with real-time bronchoscope imaging, the system enables automated navigation of the bronchoscope through key anatomical landmarks in the lung. Furthermore, prior knowledge from clinicians, acquired through eye-tracking, is incorporated to facilitate smoother and more natural motion trajectories. In summary, this paper presents the following contributions:

- An automated spatial registration framework, termed SpReg, is proposed for the first time. It comprises two core components: an Autonomous Main Airway Path Traversal (AMAPT) mechanism integrating prior knowledge from physicians' eye-tracking data and a Direction Aware Iterative Closest Point (DA-ICP) registration algorithm with directional guidance.
- The proposed method is validated via simulations and six animal experiment groups, demonstrating smoother motion and achieving comparable accuracy to manual registration.

II. Method

In this section, we present SpReg, an autonomous framework for aligning preoperative CT images to the patients' anatomy, as shown in Fig. 1. The proposed framework comprises two core components: 1) AMAPT model that utilizes surgeon eye-tracking data as prior information to guide the network's focus, and 2) DA-ICP algorithm, incorporating directional cues to enhance trajectory-to-centerline alignment.

Given a bronchoscopic image $I_{\text{cam}} \in \mathbb{R}^{H\times W\times 3}$, AMAPT model first extracts feature representations $F \in \mathbb{R}^{h\times w\times d}$ using a convolutional neural network (CNN) backbone. A self-attention module refines these features by integrating expert eye-gaze priors. Two task-specific multi-layer perceptron heads then produce decision vectors corresponding to the predicted visibility for bronchoscope motion and the coordinates of the distal reference point, respectively. Concurrently, the bronchoscope trajectory is captured using electromagnetic tracking, while a virtual airway model and its centerline are reconstructed from preoperative CT scans. Trajectory-to-centerline registration is then performed using the DA-ICP algorithm, producing the transformation matrix $T \in SE(3)$ for CT-to-patient alignment.

A. Autonomous Main Airway Path Traversal

Traditional bronchoscopic registration methods rely on labor-intensive manual traversal of the airway, which is both

time-consuming and susceptible to operator variability. These limitations present a significant bottleneck in clinical workflows, often compromising the consistency and reliability of the registration outcome. To overcome this limitation, we propose the AMAPT model, which replaces manual exploration with automated path traversal, thereby improving efficiency and consistency.

Surgeons' eye movements during bronchoscopy provide critical cues about regions of visual attention. Inspired by visual saliency prediction [22], [23], we incorporate eye-tracking measurements to model the surgeon's attention distribution during bronchoscope operation. Specifically, a ResNet50 [24] backbone extracts hierarchical feature maps from the input image, which are then processed by a self-attention module to capture multi-scale and long-range contextual dependencies. The resulting attention output is fused with eye-tracking data to generate a spatial attention map that emphasizes clinically relevant regions, ensuring alignment between the network's focus and that of expert operators. The overall architecture of the proposed AMAPT model is shown in Fig. 1(a).

Given an input bronchoscopic image I_{cam}, high-level feature maps $F \in \mathbb{R}^{H/32 \times W/32 \times f}$ are first extracted. To prepare for the self-attention mechanism, F is flattened into a one-dimensional sequence, and positional encodings are added to preserve spatial information. The resulting feature sequence $N \in \mathbb{R}^{n \times f}$ (where $n = \frac{H}{32} \times \frac{W}{32}$) is then linearly projected to form the query, key, and value matrices:

$$Q = f_Q(N), \quad K = f_K(N), \quad V = f_V(N), \tag{1}$$

where $Q, K, V \in \mathbb{R}^{n \times d}$ are derived through the learnable linear transformations f_Q, f_K, f_V, and d denote the latent spatial dimension of each token. At the same time, we model the machine attention using the following formula:

$$M = \text{softmax}\left(\frac{QK^\top}{\sqrt{d}}\right). \tag{2}$$

In this formula, the matrix $QK^T \in \mathbb{R}^{n \times n}$ is used to characterize the pairwise attention relationships between elements. To enhance numerical stability during training, the attention scores are scaled $1/\sqrt{d}$ and then normalized into a probability distribution using the softmax function. The resulting matrix $M \in \mathbb{R}^{n \times n}$ is defined as the machine attention map, which represents the relative importance among elements. To incorporate expert prior knowledge into the model, we use eye-tracking data collected from an eye tracker (Tobii 4C eye tracker) to constrain the machine attention map.

During bronchoscopy, the physician's gaze is recorded as a sequence of fixation points $\{(x_i, y_i)\}_{i=1}^{N}$, which are projected onto bronchoscopy images and convolved with a 2D Gaussian kernel $G_\sigma(x, y)$ to form an eye-gaze heatmap:

$$G(u, v) = \sum_{i=1}^{N} G_\sigma(u - x_i, v - y_i), \tag{3}$$

where $G(u, v)$ denotes the probability of visual attention at each pixel. From this heatmap, a physician attention map $D \in \mathbb{R}^{n \times n}$ is constructed. To align it with the model's attention map M, the size of M is adjusted to correspond to D through linear interpolation. Each row of D subsequently encodes a probability distribution over spatial regions attended by the expert, enabling the computation of the Kullback-Leibler divergence [25]:

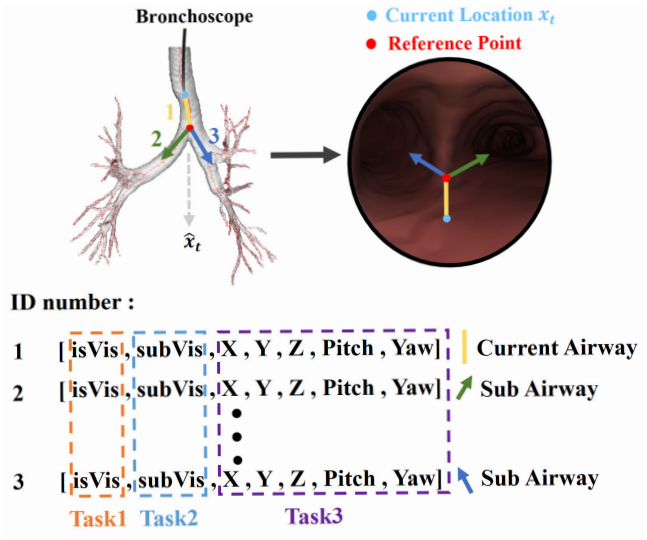

Fig. 2. Bronchoscope position estimation. The camera image I_{cam} from the bronchoscope's current location is processed by a trained AMAPT to output a feature vector for each airway in the lung skeleton. This vector includes airway visibility and bifurcation visibility. If an airway is visible, the model estimates its reference point's position relative to the current camera coordinate system.

$$L_{KL} = \sum_{i=1}^{n} \text{KL}(D_i || M_i) = \sum_{i=1}^{n} \sum_{j=1}^{n} D_{ij} \log\left(\frac{D_{ij}}{M_{ij} + \varepsilon}\right), \tag{4}$$

where ε is a small constant for numerical stability. Finally, these attention weights are multiplied by V to obtain the output of the final self-attention module:

$$\text{SelfAttention} = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V = MV. \tag{5}$$

The output of the self-attention layer is passed into two branches, with each row of the branch representing a unique airway ID number i, as shown in Fig. 2. The first branch contains two visibility boolean values: isVis and subVis. When any point on the airway centerline is visible, isVis is set to true, indicating that the point is within the camera's 60° field of view and the maximum visible distance (4 cm). If the bifurcation of the airway is visible, subVis is set to true. The second branch contains position and orientation information for the furthest visible point of the airway centerline in the camera frame. The number of airway IDs is a hyperparameter that represents the upper limit of the possible number of airways in the lungs. In this context, the upper limit is set to 300 to account for various lung conditions. The loss function for this portion of the model is as follows:

$$L_{Location} = L_{vis} + L_{reg}. \tag{6}$$

The definitions of L_{vis} and L_{reg} are given as follows:

$$L_{\text{vis}} = \sum_{i=1}^{M} f(y_{\text{p}}^{(i)}) \Big[- c_1 \cdot y_{\text{Vis}}^{(i)} \log(\hat{y}_{\text{Vis}}^{(i)}) - c_2 \cdot y_{\text{subVis}}^{(i)} \log(\hat{y}_{\text{subVis}}^{(i)}) \Big], \tag{7}$$

$$L_{\text{reg}} = \sum_{i=1}^{M} f(y_{\text{p}}^{(i)}) \cdot y_{\text{Vis}}^{(i)} \left[c_3 \left\| \hat{y}_{\text{p}}^{(i)} - y_{\text{p}}^{(i)} \right\|_2^2 + c_4 \left\| \hat{y}_{\text{o}}^{(i)} - y_{\text{o}}^{(i)} \right\|_2^2 \right], \tag{8}$$

where the adaptive weight is defined as:

$$f(y_{\text{p}}^{(i)}) = \max(c_5,\ c_6 - c_7 \cdot \|y_{\text{p}}^{(i)}\|_2), \tag{9}$$

Here, c denotes the set of hyperparameters that balance the classification and regression objectives. The depth-dependent scaling term is designed to prioritize errors in proximal airways over those in more distal regions, and the regression term is evaluated only when airway i is visible. Following the strategy outlined in [26], the hyperparameters were selected as $c_1 = 2$, $c_2 = 2$, $c_3 = 1$, $c_4 = 10$, $c_5 = 0$, $c_6 = 6$, and $c_7 = 0.2$.

Finally, the total loss of the model is constructed as follows:

$$L_{total} = L_{Location} + L_{KL}. \tag{10}$$

In autonomous traversal, the bronchoscope follows a predefined airway-ID sequence toward the target registration landmark at an insertion velocity v (manually tuned). The trajectory switches to the next airway ID when the distal-bifurcation distance is below a threshold d and the next-airway reference point is visible; d is a tunable hyperparameter set to 4 cm in this study. At that point, the terminal reference point of the current airway is treated as reached, and the trajectory is updated accordingly. The distal tip orientation is then adjusted using the predicted pitch and yaw to guide forward motion. If the reference point is temporarily lost, the system retracts until it re-enters the field of view.

B. Direction Aware Iterative Closest Point Algorithm

Following the autonomous traversal phase, the framework initiates a registration process to align the preoperative CT coordinate system with the patient's intraoperative anatomy. To achieve this, a trajectory of 6-Degree-of-Freedom (6-DOF) poses is collected at a 20 Hz sampling rate from an EMT sensor mounted at the distal tip of the bronchoscope, as shown in Fig. 1(b). To ensure a balanced spatial distribution, the raw trajectory is spatially down-sampled by enforcing a minimum distance of 1 mm between consecutive points.

Let the filtered bronchoscope trajectory be defined as the source point set $\mathbf{B} = \{b_i\}_{i=1}^{M} \subset \mathbb{R}^3$, and the preoperative airway centerline be the target point set $\mathbf{C} = \{c_j\}_{j=1}^{N} \subset \mathbb{R}^3$. The objective of the registration is to solve for the optimal rigid-body transformation $T \in SE(3)$ that minimizes the distance between the transformed source points and their corresponding target points.

Traditional nearest-neighbor matching methods often struggle to establish reliable correspondences, particularly in areas with complex bifurcations. To address this limitation, we propose an enhanced ICP algorithm [27], termed DA-ICP. This method constrains the correspondence search by considering both Euclidean distance and local orientation consistency. Specifically, we first associate each point $b_i \in \mathbf{B}$ and $c_j \in \mathbf{C}$ with an orientation vector, $\vec{v}_{b_i}$ and $\vec{v}_{c_j}$ respectively, which is calculated as the principal direction of the local point neighborhood.

The correspondence search is then based on a combined, weighted distance metric:

$$d_{\text{comb}}(b_i, c_j) = d_{\text{pos}}(b_i, c_j) + \alpha \cdot d_{\text{orient}}(\vec{v}_{b_i}, \vec{v}_{c_j}), \tag{11}$$

where d_{pos} is the Euclidean distance between points and d_{orient} is the distance between their orientation vectors:

$$\begin{cases} d_{\text{pos}}(b_i, c_j) = \|b_i - c_j\|_2^2 \\ d_{\text{orient}}(\vec{v}_{b_i}, \vec{v}_{c_j}) = \|\vec{v}_{b_i} - \vec{v}_{c_j}\|_2^2. \end{cases} \tag{12}$$

The weighting parameter α balances the influence of position and orientation. It is dynamically updated at each iteration to reflect the local structural consistency between the point sets and is defined as follows:

$$\alpha = \frac{1}{N} \sum_{j=1}^{N} \frac{d_{\text{position}}^{i,j}}{d_{\text{orientation}}^{i,j}} \tag{13}$$

The registration is solved within an Expectation-Maximization (EM) framework, which alternates between finding the optimal correspondences based on the combined distance (E-step) and computing the transformation T that minimizes the registration error (M-step). This iterative process continues until a predefined convergence criterion is met.

Upon convergence of this EM framework, the optimal rigid-body transformation is determined. The DA-ICP algorithm thus successfully registers the intraoperatively acquired trajectory point set $\mathbf{B}$ with the preoperative centerline point set $\mathbf{C}$, yielding the final transformation matrix T that precisely represents the spatial relationship between the real-time tracking coordinate system and the preoperative CT coordinate system, as illustrated in Fig. 1(c).

III. Experiments and results

A. Experimental Setup

To validate the proposed method, a series of simulation experiments was conducted, followed by animal experiments. All simulations were performed on a desktop workstation equipped with a 3.2 GHz Intel CPU and an NVIDIA RTX 4070 Super GPU (12 GB memory). The simulation environment was implemented using the SOFA platform [28], and a high-fidelity pulmonary airway model was constructed. This model was constructed with a spatial resolution of 0.2 mm, encompassing five generations of bronchial branches to approximate the anatomical complexity of the human lung, as shown in Fig. 3(a). During the simulation, the physics

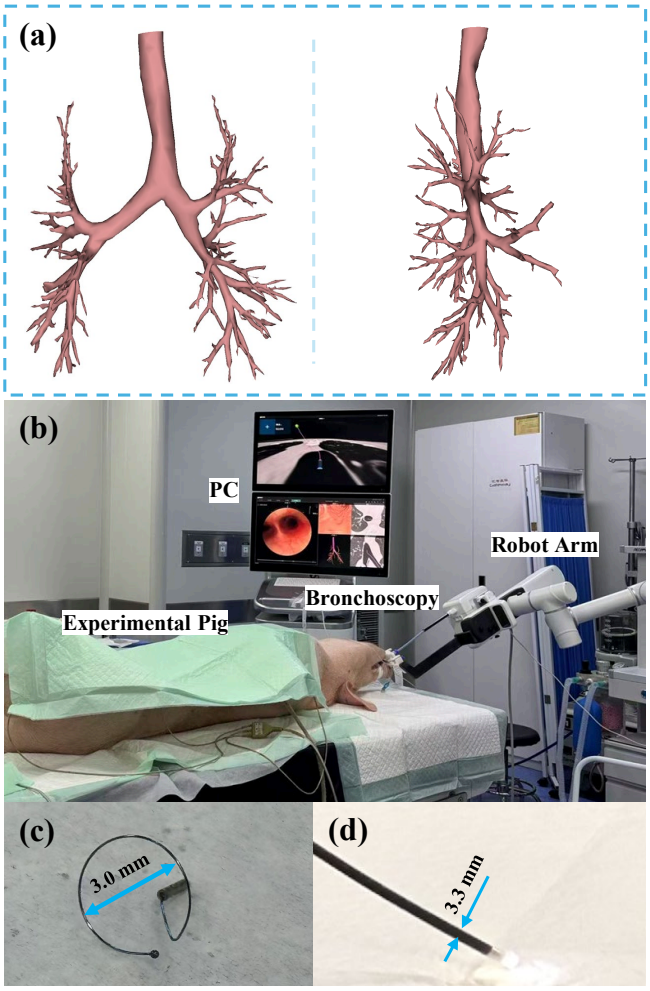

Fig. 3. Robotic bronchoscopy system and experimental setup. (a) Different viewpoints of the 3D airway model in the simulation environment. (b) Physical animal experiment of robotic bronchoscopy, with the setup comprising the robotic arm, endoscopy display, and the experimental pig. (c) Mock lesion marker employed in the animal experiment. (d) Close-up view of the bronchoscope.

engine was configured with a time step of 0.01 s to ensure numerical stability while maintaining computational efficiency. The average frame rate ranged from 15 to 45 FPS. To ensure accurate bronchoscope navigation and operation, the SpReg system was integrated into the simulation platform to autonomously navigate to the anatomical landmarks required for registration and to record six degrees of freedom (DOFs) of trajectory data throughout the process. The training dataset was generated using the bronchoscope's camera within the SOFA environment and comprised 3,641 images from seven IDs, accompanied by the corresponding physician's eye-tracking data. The SpReg model was implemented in PyTorch and trained with the Adam optimizer for 120 epochs, using a batch size of 32 and a learning rate of 0.0001.

As illustrated in Fig. 3(b), (c), and (d), we further conducted automatic registration experiments on six porcine subjects using a robotic bronchoscopy system (Unicorn *QiLin*^{TM}, LungHealth, China). During the experiments, the physical manipulation of the bronchoscope is executed by a robotic platform, while the SpReg provides real-time autonomous path traversal by integrating intraoperative endoscopic images with preoperative CT scan data. The training dataset is collected from in vivo porcine experiments, where an experienced physician manually navigates the bronchoscope throughout the bronchial tree. During the procedures, intraoperative videos are continuously captured by the front-end camera of the bronchoscope at a frame rate of 30 FPS, while the corresponding six degrees of freedom (6-DOF) trajectory data of the bronchoscope are simultaneously recorded from the robotic platform. The average duration of each trial is approximately 15–25 minutes. The subsequent training process is consistent with that under the simulation conditions.

B. Evaluation Metrics

To assess the performance of the proposed SpReg, a set of metrics is defined to characterize the system from multiple perspectives.

- **Path Traversal Error**: The precision of autonomous path traversal is evaluated using the Euclidean distance between the bronchoscope tip trajectory $\{p_i\}_{i=1}^{N}$ and the nearest points $\hat{c}(p_i)$ on the airway centerline.
- **Registration Accuracy**: Spatial alignment is evaluated by comparing manual and SpReg-based registration. A physician first performed manual registration to obtain the reference transformation $\mathbf{T}_{\text{Manual}}$, and subsequently performed navigation tasks based on both $\mathbf{T}_{\text{Manual}}$ and the SpReg estimated transformation $\mathbf{T}_{\text{Auto}}$. Six anatomical target points $\{q_j\}_{j=1}^{6}$ are selected, and the following metrics are computed for each:

  1) Target Distance Error: The Euclidean distance between the bronchoscope tip p_j and the target q_j:

$$E_{\text{target}}^{(j)} = \|p_j - q_j\|_2, \quad j = 1, \ldots, 6. \tag{14}$$

  2) Statistical Comparison: Paired differences between trajectory deviations are computed as:

$$d_k^{(j)} = E_{\text{traj,Auto}}^{(j,k)} - E_{\text{traj,Manual}}^{(j,k)}, \quad k = 1, \ldots, N_j. \tag{15}$$

  A non-significant p-value ($p > 0.05$) indicates comparable performance between SpReg-based and manual registration, validating the effectiveness of the proposed method.
- **Task Completion Time**: Operational efficiency is assessed using the total completion time (TCT) of the registration task, defined for each trial as:

$$\text{TCT} = t_{\text{end}} - t_{\text{start}}, \tag{16}$$

  where t_{start} and t_{end} denote the timestamps that mark the beginning and end of the registration process.

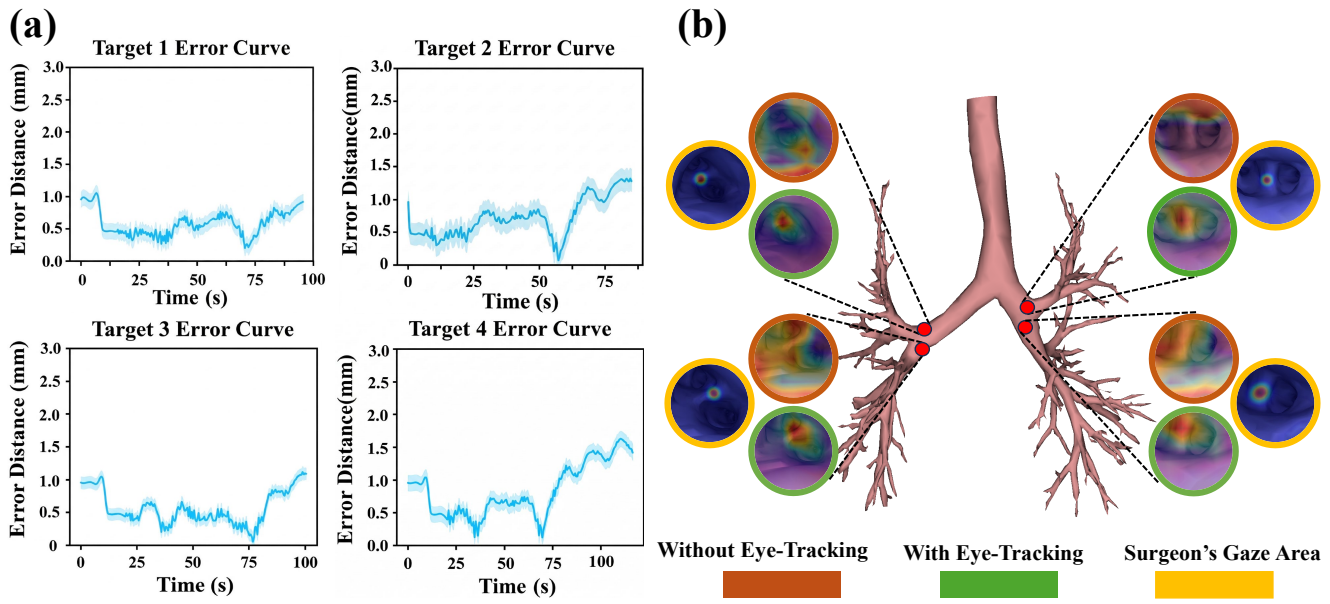

Fig. 4. Performance analysis of the SpReg. (a) The error between the bronchoscope tip and the nearest point over time during path traversal to four distinct targets. The solid line and shaded area represent the mean and standard deviation, respectively. (b) CAM visualizations at the target locations, comparing the model's attention with and without the eye-tracking module.

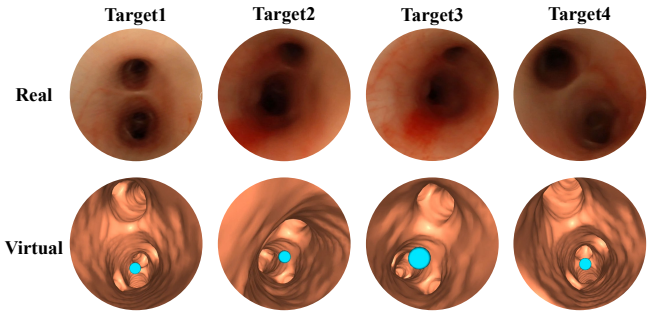

Fig. 5. Comparison of real (top row) and virtual (bottom row) bronchoscopic images at four preset registration points.

TABLE I

DISTANCES TO TARGET LESION

Cases	1	2	3	4	5	6	mean
Manual (mm)	2.4	1.8	2.8	2.5	2.7	3.4	2.6
SpReg (mm)	2.0	1.1	3.0	1.2	3.2	2.1	2.1

TABLE II

P VALUES FOR MANUAL AND AUTOMATIC TRAJECTORY ERROR

Cases	1	2	3	4	5	6
p-value	0.626	0.254	0.368	0.547	0.354	0.214

C. Simulation Experiments

Four anatomical landmarks (Targets 1–4), corresponding to key sites required for registration, were selected to evaluate the capability of the SpReg to autonomously traverse the bronchial tree. A virtual bronchoscope was autonomously guided by the framework to each landmark, and trajectory data were recorded to quantitatively assess path traversal accuracy and stability.

Path Traversal Error: Fig. 4(a) presents the error curve between the bronchoscope tip and the nearest points on the airway centerline during path traversal.

To further elucidate this stability, Class Activation Mapping (CAM) [29] was employed on models both with and without eye-tracking constraints, as depicted in Fig. 4(b). The results demonstrate a high degree of spatial alignment between the model's attention and the clinicians' focal regions, suggesting that SpReg successfully captures semantically meaningful visual features. Consequently, this enhances the smoothness and reliability of bronchoscope path traversal throughout the complex airway anatomy.

D. Animal Experiments

Registration Accuracy: The alignment of anatomical landmarks between real and virtual bronchoscopic environments is presented in Fig. 5. Real bronchoscopy images, captured when the target landmarks were reached under SpReg guidance, are displayed in the first row, whereas the corresponding virtual bronchoscopy images, in which blue markers indicate the same targets, are presented in the second row. Despite variations in illumination and tissue morphology, the virtual markers remain aligned with the anatomical structures observed in the real images.

To evaluate navigation performance, a target lesion was predefined in each trial. All targets were successfully reached by physicians using the SpReg-generated virtual airway map. Fig. 6 illustrates high anatomical consistency between real bronchoscopic views (top) and virtual projections (bottom). SpReg achieved a mean target localization error of 2.1 mm, representing a 19% improvement over manual registration

TABLE III

REGISTRATION TASK COMPLETION TIME

Cases	1	2	3	4	5	6	mean
TCT_{Manual} (s)	76.1	73.3	72.1	82.8	73.5	79.6	76.2
TCT_{SpReg} (s)	105.6	110.8	74.7	115.1	110.0	96.6	102.1

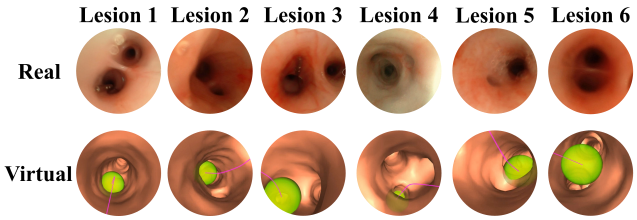

Fig. 6. Real-to-virtual correspondence at six target lesions, achieved via registration-guided manual navigation.

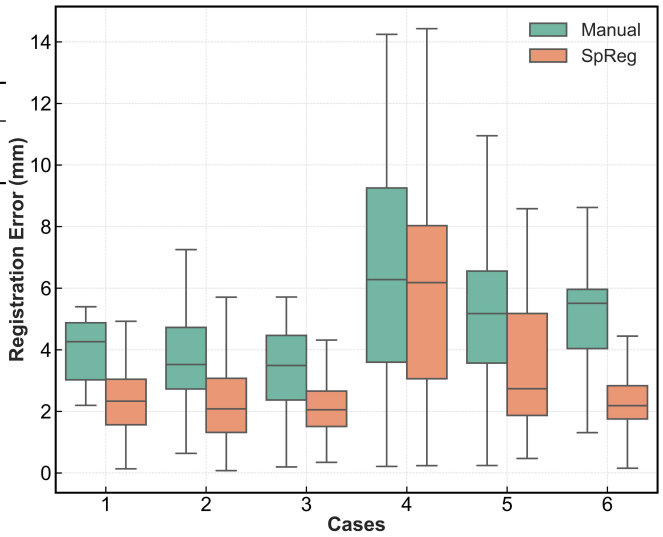

Fig. 7. Box plots comparing the registration error between the manual and automatic methods across six different cases.

(2.6 mm), with a minimum error of 1.1 mm, as shown in Table I. Fig. 7 presents a comparison of registration errors between manual and automatic methods across six experimental trials. The box plots indicate that the automatic method generally outperforms the manual method in both accuracy and consistency. The median registration error for SpReg remains consistently lower than that of the manual method across all trials, suggesting that the automatic approach yields more reliable results. SpReg also exhibits a smaller dispersion of errors, indicating that the automatic method produces more stable outcomes. In contrast, the manual method exhibits substantial variability, particularly in trials 4 and 6, where error ranges were notably wider, indicating potential instability in manual execution.

Finally, we further compared the navigation trajectories obtained after manual registration and SpReg registration. As shown in Fig. 8, the trajectory visualization results across six experiments demonstrate that the SpReg-guided paths remain highly consistent with those based on manual registration. No significant difference was observed between the two methods ($p > 0.05$), as reported in Table II. These findings indicate that SpReg achieves registration-and-navigation performance comparable to expert manual practice, while reducing operator-dependent variability and improving overall workflow consistency.

Task Completion Time: Table III compares task completion time between the two registration methods. Manual registration yielded an average TCT of 76.2 s. In contrast, SpReg required an average of 102.1 s, with times ranging from 74.7 s to 115.1 s. SpReg demonstrated a 19% improvement in localization accuracy over manual registration, reducing the mean target localization error from 2.6 mm to 2.1 mm. This enhanced accuracy was achieved by allocating additional computational resources to the optimization of the path traversal. Although this process increases computational overhead, it ensures accurate and reliable path traversal through complex anatomical structures and reduces operator-dependent variability.

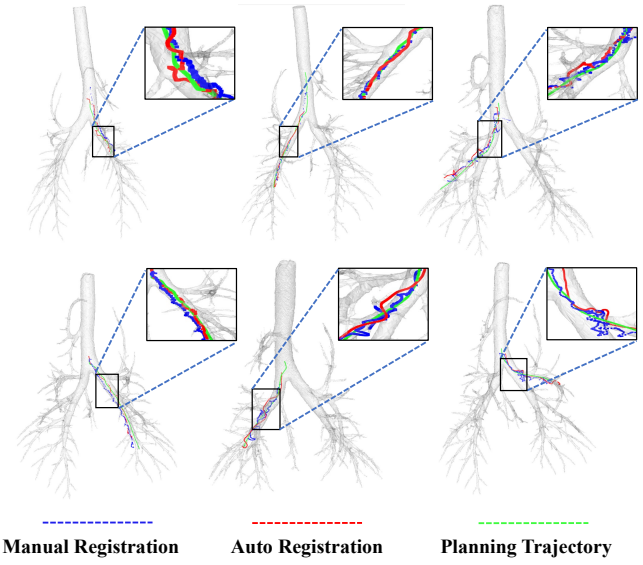

Fig. 8. Visualization results of navigation trajectories based on manual registration and SpReg registration. The paths are overlaid on 3D airway models for six representative cases.

IV. CONCLUSIONS

To address the inefficiency and subjectivity of manual registration in robotic bronchoscopy, we introduce SpReg, an end-to-end autonomous registration framework. The framework is centered on AMAPT model that identifies and acquires corresponding feature points. The principal innovation lies in the integration of clinician eye-tracking data, through which expert prior knowledge is incorporated to constrain the automated process. Through this fusion of expert gaze and autonomous control, more stable and clinically relevant registration is achieved, thereby translating clinical expertise into actionable guidance for the robotic system.

Extensive evaluations conducted on a high-fidelity simulation platform and in six in vivo porcine studies demonstrated

that SpReg attained registration accuracy comparable to manual methods while substantially reducing clinician workload. These results highlight the potential of SpReg to provide reliable registration for robotic bronchoscopy. Future work will focus on optimizing the navigation model and improving its generalization across diverse anatomical and procedural scenarios, thereby enhancing robustness and practical utility for clinical deployment.

V. ACKNOWLEDGEMENTS

This work was supported by National Key Research and Development Program of China (2025YFC2426300), the Science and Technology Commission of Shanghai Municipality (Nos.24511104100, 25ZR1402225, 24ZR1439800); National Nature Science Foundation of China Grants (82572314, 62477031, 62403307, 62271246), the Open Research Fund of The State Key Laboratory of Multimodal Artificial Intelligence Systems. This manuscript was partially assisted by DeepSeek and ChatGPT for the preparation of the Method section, specifically in translation and language refinement.

REFERENCES

[1] A. Leiter, R. R. Veluswamy, and J. P. Wisnivesky, "The global burden of lung cancer: current status and future trends," *Nature Reviews Clinical Oncology*, vol. 20, no. 9, pp. 624–639, 2023.

[2] G. J. Criner, R. Eberhardt, S. Fernandez-Bussy, D. Gompelmann, F. Maldonado, N. Patel, P. L. Shah, D.-J. Slebos, A. Valipour, M. M. Wahidi, *et al.*, "Interventional bronchoscopy," *American Journal of Respiratory and Critical Care Medicine*, vol. 202, no. 1, pp. 29–50, 2020.

[3] Z. Niu, Y. Cao, M. Du, S. Sun, Y. Yan, Y. Zheng, Y. Han, X. Zhang, Z. Zhang, Y. Yuan, *et al.*, "Robotic-assisted versus video-assisted lobectomy for resectable non-small-cell lung cancer: the rvlob randomized controlled trial," *EClinicalMedicine*, vol. 74, 2024.

[4] S. Fu, S. Dong, H. Shen, Z. Chen, G. Ma, M. Cai, C. Huang, Q. Peng, C. Bai, Y. Dong, *et al.*, "Multifunctional magnetic catheter robot with triaxial force sensing capability for minimally invasive surgery," *Research*, vol. 8, p. 0681, 2025.

[5] P. J. Kneuertz, M. Abdel-Rasoul, D. M. D'Souza, J. Zhao, and R. E. Merritt, "Segmentectomy for clinical stage i non–small cell lung cancer: National benchmarks for nodal staging and outcomes by operative approach," *Cancer*, vol. 128, no. 7, pp. 1483–1492, 2022.

[6] J. R. Rojas-Solano, L. Ugalde-Gamboa, and M. Machuzak, "Robotic bronchoscopy for diagnosis of suspected lung cancer: a feasibility study," *Journal of Bronchology & Interventional Pulmonology*, vol. 25, no. 3, pp. 168–175, 2018.

[7] Y. Li, W. Chen, F. Xie, R. Huang, X. Liu, Y. Xiao, L. Cao, Y. Hu, M. Ke, S. Wu, *et al.*, "Novel electromagnetic navigation bronchoscopy system for the diagnosis of peripheral pulmonary nodules: a prospective, multicentre study," *Thorax*, vol. 78, no. 12, pp. 1197–1205, 2023.

[8] E. F. Hofstad, H. Sorger, J. B. L. Bakeng, L. Gruionu, H. O. Leira, T. Amundsen, and T. Langø, "Intraoperative localized constrained registration in navigated bronchoscopy," *Medical Physics*, vol. 44, no. 8, pp. 4204–4212, 2017.

[9] J. Cicenia, S. K. Avasarala, and T. R. Gildea, "Navigational bronchoscopy: a guide through history, current use, and developing technology," *Journal of Thoracic Disease*, vol. 12, no. 6, p. 3263, 2020.

[10] H. Sorger, E. F. Hofstad, T. Amundsen, T. Langø, and H. O. Leira, "A novel platform for electromagnetic navigated ultrasound bronchoscopy (ebus)," *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 8, pp. 1431–1443, 2016.

[11] I. Wegner, J. Biederer, R. Tetzlaff, I. Wolf, and H.-P. Meinzer, "Evaluation and extension of a navigation system for bronchoscopy inside human lungs," in *Medical Imaging 2007: Visualization and Image-Guided Procedures*, vol. 6509, pp. 522–533, SPIE, 2007.

[12] S. A. Merritt, J. D. Gibbs, K.-C. Yu, V. Patel, L. Rai, D. C. Cornish, R. Bascom, and W. E. Higgins, "Image-guided bronchoscopy for peripheral lung lesions: a phantom study," *Chest*, vol. 134, no. 5, pp. 1017–1026, 2008.

[13] H. Hautmann, A. Schneider, T. Pinkau, F. Peltz, and H. Feussner, "Electromagnetic catheter navigation during bronchoscopy: validation of a novel method by conventional fluoroscopy," *Chest*, vol. 128, no. 1, pp. 382–387, 2005.

[14] K. Mori, D. Deguchi, T. Kitasaka, Y. Suenaga, Y. Hasegawa, K. Imaizumi, and H. Takabatake, "Improvement of accuracy of marker-free bronchoscope tracking using electromagnetic tracker based on bronchial branch information," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 535–542, Springer, 2008.

[15] D. Deguchi, M. Feuerstein, T. Kitasaka, Y. Suenaga, I. Ide, H. Murase, K. Imaizumi, Y. Hasegawa, and K. Mori, "Real-time marker-free patient registration for electromagnetic navigated bronchoscopy: a phantom study," *International Journal of Computer Assisted Radiology and Surgery*, vol. 7, no. 3, pp. 359–369, 2012.

[16] J. Sganga, D. Eng, C. Graetzel, and D. Camarillo, "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.

[17] Q. Tian, H. Liao, X. Huang, J. Chen, Z. Zhang, B. Yang, S. Ourselin, and H. Liu, "Dd-vnb: A depth-based dual-loop framework for real-time visually navigated bronchoscopy," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 12979–12986, IEEE, 2024.

[18] J. Zhang, L. Liu, P. Xiang, Q. Fang, X. Nie, H. Ma, J. Hu, R. Xiong, Y. Wang, and H. Lu, "Ai co-pilot bronchoscope robot," *Nature Communications*, vol. 15, no. 1, p. 241, 2024.

[19] J. Zhao, H. Chen, Q. Tian, J. Chen, B. Yang, Z. Zhang, and H. Liu, "Bronchocopilot: Towards autonomous robotic bronchoscopy via multimodal reinforcement learning," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6923–6930, IEEE, 2024.

[20] L. Chen, Z. Cao, W. Zhang, X. Tang, D. Zhao, D. Zhang, H. Liao, and F. Chen, "Thinking like sonographers: Human-centered cnn models for gout diagnosis from musculoskeletal ultrasound," *IEEE Transactions on Biomedical Engineering*, 2024.

[21] J. Lee, T. Lim, and W. Kim, "Investigating the usability of collaborative robot control through hands-free operation using eye gaze and augmented reality," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4101–4106, IEEE, 2023.

[22] J. Lou, H. Lin, D. Marshall, D. Saupe, and H. Liu, "Transalnet: Towards perceptually relevant visual saliency prediction," *Neurocomputing*, vol. 494, pp. 455–467, 2022.

[23] H. Lee and S. Kim, "Sspnet: Learning spatiotemporal saliency prediction networks for visual tracking," *Information Sciences*, vol. 575, pp. 399–416, 2021.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[25] Z. Zhou and Y. Zhu, "Kldet: Detecting tiny objects in remote sensing images via kullback–leibler divergence," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–16, 2024.

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

[27] P. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[28] H. Talbot, N. Haouchine, I. Peterlik, J. Dequidt, C. Duriez, H. Delingette, and S. Cotin, "Surgery training, planning and guidance using the sofa framework," in *Eurographics*, 2015.

[29] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618–626, 2017.