

# BronchoCopilot: Towards Autonomous Robotic Bronchoscopy via Multimodal Reinforcement Learning

Jianbo Zhao<sup>1,2</sup>, Hao Chen<sup>2,1</sup>, Qingyao Tian<sup>1</sup>, Jian Chen<sup>1</sup>, Bingyu Yang<sup>1</sup>, Zihui Zhang<sup>1</sup> and Hongbin Liu<sup>1,3,4</sup>

**Abstract**—Bronchoscopy plays a significant role in the early diagnosis and treatment of lung diseases. This process demands physicians to maneuver the flexible endoscope for reaching distal lesions, particularly requiring substantial expertise when examining the airways of the upper lung lobe. With the development of artificial intelligence and robotics, reinforcement learning (RL) method has been applied to the manipulation of interventional surgical robots. However, unlike human physicians who utilize multimodal information, most of the current RL methods rely on a single modality, limiting their performance. In this paper, we propose BronchoCopilot, a multimodal RL agent designed to acquire manipulation skills for autonomous bronchoscopy. BronchoCopilot specifically integrates images from the bronchoscope camera and estimated robot poses, aiming for a higher success rate within challenging airway environment. We employ auxiliary reconstruction tasks to compress multimodal data and utilize attention mechanisms to achieve an efficient latent representation of this data, serving as input for the RL module. This framework adopts a stepwise training and fine-tuning approach to mitigate the challenges of training difficulty. Our evaluation in the realistic simulation environment reveals that BronchoCopilot, by effectively harnessing multimodal information, attains a success rate of approximately 90% in fifth generation airways with consistent movements. Additionally, it demonstrates a robust capacity to adapt to diverse cases.

## I. INTRODUCTION

Bronchoscopy has been instrumental in the inspection and diagnosis of lung diseases [1], [2]. It is a surgical procedure that allows medical professionals to visually examine the lungs and airways. Physicians are required to manipulate flexible, non-linear surgical instruments carefully through the airways to reach distal lesions, implying a requirement for extensive experience and skills. Robotic bronchoscopy platform [3] has emerged to alleviate difficulties of sensing and control for physicians, enhancing the diagnostic rate while reducing operational risks, such as discomfort or bleeding [4]. Nevertheless, due to the demands for precision and safety, mastering the platform still necessitates high training costs. As a result, current platform is expected to

\*This work was supported by the Institute of Automation, Chinese Academy of Sciences. (Corresponding author: Hongbin Liu)

<sup>1</sup>Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China ({zhaojianbo2022, chenhao2020, tianqingyao2021, chenjian2020, liuhongbin}@ia.ac.cn)

<sup>2</sup>School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup>Centre of AI and Robotics (CAIR), Hong Kong Institute of Science & Innovation, Chinese Academy of Sciences, Hongkong, China

<sup>4</sup>School of Engineering and Imaging Sciences, King's College London, UK

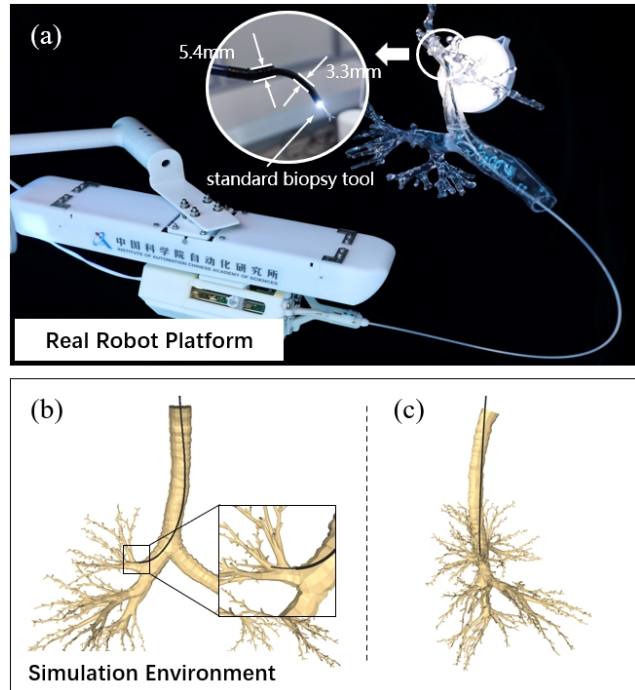


Fig. 1: (a) Real robot platform, the insertion part includes outer sheath and inner endoscope. (b), (c) The simulation environment, includes the 3D airway model and the simulated dual-segment flexible endoscopic robot.

have higher-level autonomy, performing more complex tasks with enhanced success rate and consistent motion [3].

Many navigation systems have been proposed for robot-assisted bronchoscopy, aiming to alleviate cognitive and physical stress on physicians. This allows them to redirect their focus from the manipulation of surgical instruments to making more advanced intervention decisions and diagnoses. Previous efforts include kinematics-based [5], [6] and image-based [7], [8] motion planners, which have significantly enhanced diagnostic rates. However, these methods require sophisticated manual designs, and they overlooked the interaction between the robot and the airway wall, but it is common because the robot needs support from the wall in tortuous airways. In recent years, propelled by advancements in artificial intelligence and robotics, reinforcement learning (RL) has been applied to surgical robots, enhancing the surgical autonomy [9]–[14]. While showing benefits such as safety and transferability, current RL methods often struggle in complex tasks. Contrary to human physicians who leverage multimodal information for decision-making,

RL methods typically concentrate on a single modality, either vision [9], [10] or proprioception (including position and orientation) [11]–[13]. This challenge arises due to the heterogeneity and inconsistency of multimodal information, and the indirect reward feedback in RL training further amplifies its complexity [14], [15].

Inspired by the strategies employed by experienced surgeons, in this paper, we propose **BronchoCopilot**, a RL-based agent leveraging multimodal information for autonomous robotic bronchoscopy. Specifically, BronchoCopilot devises manipulation strategy to a target by bronchoscope camera images and estimated robot pose, with enhanced success rate, consistency and safety in complex airway environments. To address the challenges posed by the heterogeneity of different data modalities while maximizing their complementary nature, auxiliary reconstruction tasks [16] are introduced to obtain low-dimensional representations of multimodal data. Furthermore, we employ a cross-modal attention mechanism to dynamically adjust the significance of different modalities. Through a staged training regime and stepwise fine-tuning, we tackle the convergence challenges inherent in multimodal reinforcement learning algorithms.

As a proof of concept, the training and evaluation of BronchoCopilot are carried out in a realistic simulation environment, which is elaborated in the section III. A. The agent is designed for the dual-segment flexible endoscopic robot platform derived from [17]. As depicted in Fig. 1(a), the robot features an outer sheath and an inner endoscope, each actuated by three cables. Coordinated actions between these components are crucial for navigating to distal lesions, particularly within the finer airways of the upper lung lobe. The main contributions of this work are as follows:

- To the best of our knowledge, BronchoCopilot is the first work to use multimodal reinforcement learning for autonomous interventional surgical robot manipulation.
- We propose a novel algorithm framework that employs cross-attention and stepwise training, to fuse modalities with heterogeneity and alleviate the convergence difficulty.
- Detailed experiments demonstrate that through our method, multimodal information can improve agent’s performance in complex scenarios, as well as its generalization capability across various cases.

## II. RELATED WORK

### A. RL Methods in Interventional Surgical Robots

Interventional surgical robots are widely used in minimally invasive surgeries (MIS), typically equipped with flexible catheters or needles to navigate through vessels, cavities, or surgical openings [18]. They enable precise and tremor-free continuous operations. Nevertheless, due to the non-linearity of the instruments and the constraints of narrow working spaces, traditional controllers struggle to overcome the operational challenges posed by surgical robots [19]. Numerous efforts have applied RL methods to the manipulation of interventional surgical robots, including path

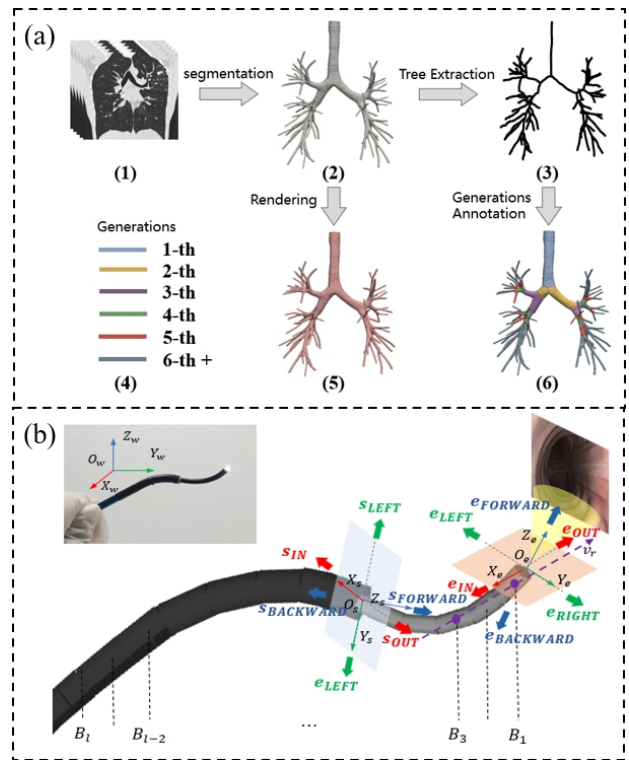


Fig. 2: The establishment of the simulation environment. (a) The creation of the airway model, including segmentation from preoperative CT scans, bronchial tree extraction and rendering in the simulator, and we visualize the airway’s generations based on tree branching. (b) The FEM modeling and the action space of the robot. The purple arrow  $v_r$  denotes the orientation vector of the robot.

planning [10], [11], [20], surgical training systems [9], [21], and surrogate surgeon operations [12], [13]. These methods typically rely on single-modal surgical information, unlike human physicians often requiring multimodal feedback to perform their manipulations. As a result, these methods are mostly confined to simpler tasks and scenarios, lacking the capability to learn in more complex surgical environments and instrument behaviors.

### B. Autonomous Robotic Bronchoscopy

The integration of artificial intelligence and control theory with bronchoscopy significantly enhances the autonomy of robot-assisted bronchoscopy platforms, reducing training costs and workloads on physicians [22]. For example, the Alterovitz team [23], [5] has made significant contributions to steerable needle lung robot, designing motion planners based on needle kinematics, with bronchoscope navigation being one stage in the overall process. However, these efforts did not specifically consider the robot deformation caused by bronchoscope contact with the airway wall, restricting the maneuverability of the bronchoscope. By leveraging technologies of computer vision, researchers have developed a range of image-based guiding systems [7], [8], [24]. These systems typically utilize endoscopic images or fluoroscopy to

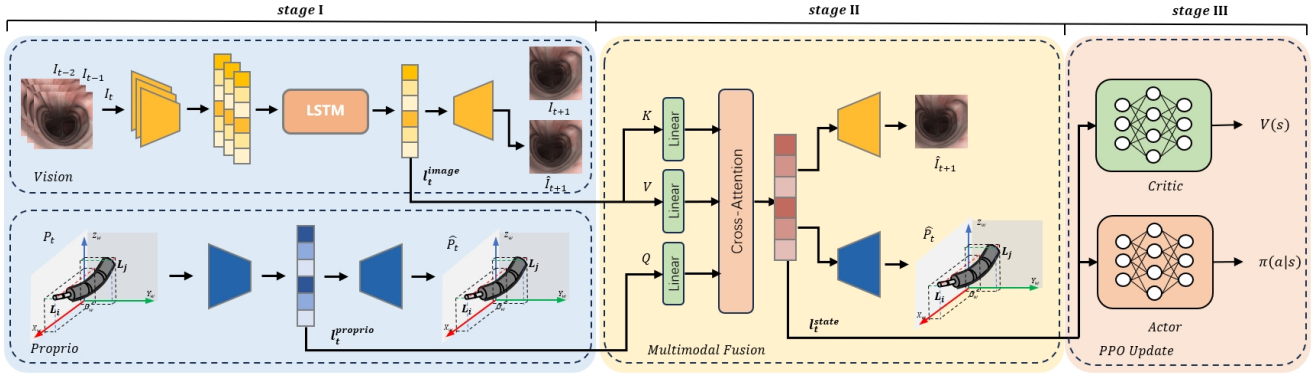


Fig. 3: The Architecture of our method. The network takes data from two different modalities as input and outputs the manipulation policy. The entire architecture is trained in stages. In **stage I**, it encodes multimodal information to low-dimensional embeddings through reconstruction tasks. In **stage II**, it fuses multimodal embeddings into state representation as the input of **stage III**, with the loss from subsequent tasks used to fine-tune the parameters of the front stage’s network.

estimate robot poses and devise insertion strategies, raising the precision and efficiency of the procedure. Nevertheless, these methods exhibit unsatisfactory performance in upper lung lobe interventions, which require pre-bending or retraction maneuvers [25]. Additionally, due to their complete reliance on visual information, limitations in perspective and estimation errors may raise safety concerns.

### C. Multimodal Reinforcement Learning

Multimodal models outperform single-modal ones, as confirmed by both theory and experiments [26], [27]. The complementarity of heterogeneous sensor modalities has been explored for training decision models [28], [29], [30]. The key to multimodal RL lies in how to obtain latent representations suitable for RL tasks [15]. It is easy when labeled data is available, but the indirect feedback in RL training makes it more challenging. We adapted the pre-training approach similar to [16], leveraging auxiliary reconstruction tasks to obtain compressed and unified representations of multimodal information. These representations are further fine-tuned during the RL training phase. Furthermore, many works encode multimodal inputs using various encoders and then fuse them through summation or concatenation [31]. This approach may mask and introduce ambiguity in inter-modal information. In this paper, we explored a cross-attention based multimodal fusion approach [32], [33] to integrate inter-modal information while addressing dynamic requirements.

## III. METHOD

### A. Realistic Simulation Environment Design

Due to extended training durations and potential risky behaviors, direct training in real-world settings is costly. As illustrated in Fig. 1(b), we have developed a simulation environment tailored to provide detailed and authentic descriptions of actual bronchoscopy procedures.

To construct the airway model, as shown in Fig. 2(a), we employ a segmentation approach [34] to extract lung

anatomy from preoperative CT scans. The model is refined to generate surface mesh as collision environment, which is composed of multiple triangles. Subsequently, the obtained surface mesh is displayed by OpenGL [35], where high-quality, realistic visual rendering is applied, including additional effects such as the reflection on the surface of the inner wall and the illumination from the front camera. We consider the bronchial model as rigid for the purpose of collision detection within the environment [36]. The airway centerlines are extracted by VMTK [37] to serve as reference paths, then the airway generations can be determined.

For the simulation of the robot, we use a finite element beam model to simulate the deformation of the robot and consider the effects of multiple contact points with the environment, similar to the SOFA [38], [39]. In the simulation we use parameters obtained from real robot: The Young’s modulus of the sheath is 510 MPa, the length is 0.7 m, the area is  $9.30 \times 10^{-6} \text{ m}^2$ , and the moment of inertia is  $19.233 \times 10^{-12} \text{ m}^4$ . Correspondingly, the corresponding parameters of the endoscope part are 307 MPa, 0.7 m,  $2.83 \times 10^{-6} \text{ m}^2$ ,  $1.817 \times 10^{-12} \text{ m}^4$ .

It is essential that multimodal data is easy to access in the real environment. In this work, the visual information comprises endoscopic images obtained through a camera mounted at the robot’s tip. The robot’s proprioceptive information can be acquired by bronchoscopic localization systems like electromagnetic (EM). Furthermore, to capture the nonlinear kinematic characteristics of the robot, advanced shape perception technique [40] is employed for overall pose estimation of the robot.

### B. Problem Statement

The problem can be stated as follows: In the realistic simulation environment, a continuous flexible robot moves within the airways and reaches distal target. BronchoCopilot needs to determine the optimal action set ( $A = \{a_0, a_1, \dots, a_j\}$ ) to reach the target by leveraging multimodal information gathered from the environment. The process is

defined as a partially observable Markov decision process  $\langle \mathcal{S}, \mathcal{O}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$ , where  $\mathcal{S}$  is the set of states,  $\mathcal{O}$  is the observation space,  $\mathcal{A}$  is the action space,  $\mathcal{P} : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}'$  is the transition probability,  $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{R}$  is the reward function, and  $\gamma$  is the discount factor. Each element is defined in detail as follows:

1) Observations: The observations include two parts as shown in Fig. 3. At timestep  $t$ , for visual data, it consists of a sequence of  $c$  consecutive frames  $I$  captured by the camera on the bronchoscope  $\mathcal{O}_t^v = \{I_{t-c+1}, \dots, I_{t-1}, I_t\}$ , and for proprioceptive data, it is an array composed of the concatenated coordinates of the bronchoscope backbone  $\mathcal{O}_t^p = \{B_1, B_2, \dots, B_n\}$ , where  $B_i = [x_i, y_i, z_i]$  denotes the position in world coordinate system.

2) Actions: The outer sheath and inner endoscope share the same driving mechanism. As shown in Fig. 2(b), we define discrete action elements. Specifically, at each time step, the sheath and endoscope can either move forward/backward by 3 mm, or bend by 0.2 radians in  $xOz$  and  $yOz$  planes relative to respective coordinate systems  $O_sxyz$  and  $O_exyz$ . Only one of the sheath and endoscope can execute an action at the same time step. Then the action space can be defined as  $\mathcal{A} = \mathcal{A}_s + \mathcal{A}_e$ , where  $s$  denotes the sheath and  $e$  denotes the endoscope, and  $\mathcal{A}_s = \{s_{FORWARD}, s_{BACKWARD}, s_{LEFT}, s_{RIGHT}, s_{IN}, s_{OUT}\}$ ,  $\mathcal{A}_e = \{e_{FORWARD}, e_{BACKWARD}, e_{LEFT}, e_{RIGHT}, e_{IN}, e_{OUT}\}$ .

3) Rewards: By leveraging the bronchial tree, the path to the target position can be uniquely determined. The design of the reward function aims to encourage the robot to move accurately, efficiently and safely along the reference path:

$$R_t = \omega_1 * r_d + \omega_2 * r_a + r_b + r_g, \quad (1)$$

where

$$r_d = -\|B_1 - g_k\|, \quad (2)$$

$$r_a = -\left(e^{\langle \mathbf{v}_1, \mathbf{v}_2 \rangle / \pi} - 1\right), \quad (3)$$

$$r_b = \begin{cases} -20, & \text{if break} \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

$$r_g = \begin{cases} 10, & \text{if reached} \\ 0, & \text{otherwise} \end{cases}, \quad (5)$$

where  $B_1$  denotes the location of the robot's tip,  $g_k$  denotes the endpoint coordinates of  $k$ -th generation airway of the reference path,  $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle$  signifies the angle between the robot's orientation and the airway's orientation. The robot's orientation is defined as the vector extending from the third to the first backbone of its tip, while the airway's orientation is determined by the vector from the starting to the ending point of its centerline within the segment.  $\omega_1, \omega_2$  are hyper-parameters. Additionally, to ensure safety and efficiency, we set thresholds for contact force, direction angle, and distance between the robot's tip and the target. Exceeding these thresholds results in the premature termination of the episode and a penalty  $r_b$  is applied as a consequence.

### C. Multimodal Information Extraction and Fusion

As previously discussed, directly acquired multimodal information differs in dimensions and values, making it unsuitable for direct input to the decision model. In this section, we delve into the representation and fusion of multimodal information.

As shown in Fig. 3, in **stage I**, we manually operate the simulator to thoroughly explore the airways. For visual data, continuous video frames are captured, and a dynamic prediction task is designed to understand the patterns of image changes and deduce the robot's motion state. In this paper we set  $c = 3$ . At timestep  $t$ , frames  $I_{t-2}$ ,  $I_{t-1}$ , and  $I_t$  are respectively encoded and input into the LSTM [41] model. The process can be formulated as  $l_t^{image} = LSTM_{\delta}(f_{\xi}(I_{t-2}), f_{\xi}(I_{t-1}), f_{\xi}(I_t))$ , and the decoder process aims to predict the next frame:  $\hat{I}_{t+1} = g_{\xi'}(r_t^{image})$ , where  $\delta, \xi, \xi'$  denotes the parameters of LSTM, visual encoder and decoder separately. For proprioceptive data, we use vanilla Autoencoder [42] for reconstruction task, at timestep  $t$ , the feature of proprioceptive data is  $l_t^{proprio} = f_{\psi}(P_t)$ , and the decoding process is  $\hat{P}_t = g_{\psi'}(l_t^{proprio})$ , where  $\psi, \psi'$  denotes the parameters of proprioceptive autoencoder. All parameters are updated by gradient descent on the reconstruction error:

$$\xi^*, \xi'^*, \delta^* = \arg \min_{\xi, \xi', \delta} \frac{1}{n} \sum_{i=1}^n \mathcal{L} \left[ I_{t+1}^{(i)}, \hat{I}_{t+1}^{(i)} \right], \quad (6)$$

$$\psi^*, \psi'^* = \arg \min_{\psi, \psi'} \frac{1}{n} \sum_{i=1}^n \mathcal{L} \left[ P_t^{(i)}, \hat{P}_t^{(i)} \right], \quad (7)$$

in which  $\mathcal{L}$  is a mean squared error loss function.

In **stage II**, we use cross-attention to fuse visual and proprioceptive information. Cross-attention enables a better capture of dynamic inter-modal relationships, which is calculated by:

$$l_t^{state} = \text{softmax} \left( \frac{\left( W_Q l_t^{proprio} \right) \left( W_K l_t^{image} \right)^T}{\sqrt{d}} \right) W_V l_t^{image}, \quad (8)$$

where  $l_t^{state}$  denotes the state feature,  $W_Q, W_K$  and  $W_V$  denote the weight matrices, and  $d$  is the dimension of the vector. We continue to use the reconstruction tasks mentioned above to train the attention model, ensuring that the fused vector can still individually reconstruct the original multimodal inputs. During the training process, the parameters of **stage I** are frozen.

### D. Reinforcement Learning Model

As shown in Fig. 3, in **stage III**, the agent receives inputs from the state representation model and updates parameters as described above. The gradient updates of RL module are also propagated to the state representation model for further fine-tuning of its parameters.

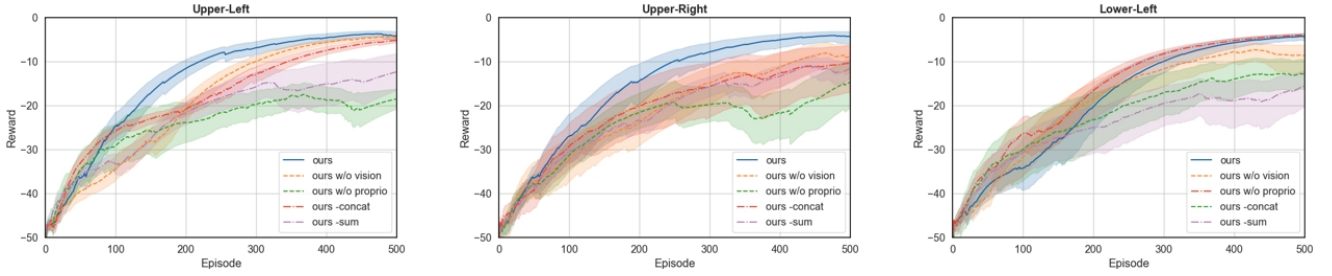


Fig. 4: RL learning curves for ablation and comparison experiments: (1) BronchoCopilot, (2) BronchoCopilot without visual data, (3) BronchoCopilot without proprioceptive data, (4) BronchoCopilot using concat for fusion, (5) BronchoCopilot using sum for fusion. All curves are smoothed by exponential smoothing with a factor of  $r=0.95$ .

The principle of RL is to train an agent following the policy  $\pi(a|s)$  which maximize the expected reward  $r$  by selecting an action  $a_t$  based the state  $s_t$  at time step  $t$ . The policy  $\pi(a|s)$  is parameterized by  $\theta$  and defined as  $\pi_\theta(a|s)$ . Following the setup,  $\theta$  is optimized to maximize the expected return in a policy gradient algorithm:

$$\theta^* = \arg \max_{\theta} \left[ \sum_t \gamma^t R(s_t, a_t) \right]. \quad (9)$$

We use Proximal Policy Optimization (PPO) [43] algorithm to maximize the loss function:

$$J_t^{CLIP'}(\theta, \theta') = \mathbb{E} [J_t^{CLIP}(\theta) - c_1 J_t^{VF}(\theta') + c_2 H(\pi_\theta | s_t)], \quad (10)$$

where  $H(\pi_\theta | s_t)$  is an entropy term to encourage exploration, and  $c_1, c_2$  are weights of loss, and  $J_t^{VF}$  is the error term on the value estimation with discount factor  $\gamma$  and target value function:

$$J_t^{VF}(\theta') = (V_{\theta'}(s_t) - (R(s_t, a_t) + \gamma V_{\theta'}(s_{t+1})))^2. \quad (11)$$

$J_t^{CLIP}(\theta)$  is the loss limited by a clipped ratio  $\epsilon$  to stabilize the update procedure:

$$J_t^{CLIP}(\theta) = \mathbb{E} \left[ \min \left( r_t(\theta) \hat{A}_t(s, a), \text{clip} \left( r_t(\theta), 1 - \epsilon, 1 + \epsilon \right) \hat{A}_t(s, a) \right) \right], \quad (12)$$

where  $r_t(\theta) = \pi_\theta(a|s)/\pi_{\theta_{old}}(a|s)$  is the probability ratio between current policy and old policy, and  $\hat{A}_t(s, a)$  is the advantage estimator calculated according to [44].

#### IV. EXPERIMENT

In this section, we perform qualitative and quantitative evaluation of our approach in the simulation environment.

##### A. Experimental Setup

We compare our BronchoCopilot method against prior collision-aware methods: AEA [8] and DQNN [13]. Since both methods were designed for traditional bronchoscope platform, we afford a convenience for these by providing prior information: only sheath actions are allowed before reaching the 4-th generation airways, and subsequently only endoscope actions are permitted. This adjustment more closely mirrors the typical procedural routines of clinicians [3]. In contrast, our method employs a unified action space for both the sheath and endoscope, challenging the agent to independently discern implicit operational strategies. The detailed experimental setup is as follows:

**AEA:** Given the camera image and visualized centerline, insertion or bending is determined based on the observed centerline position from the camera. We directly obtain actions without converting them into specific control signals.

**DQNN:** Given the camera image, the action space from [2] is reduced to the settings in this paper (i.e., eliminating

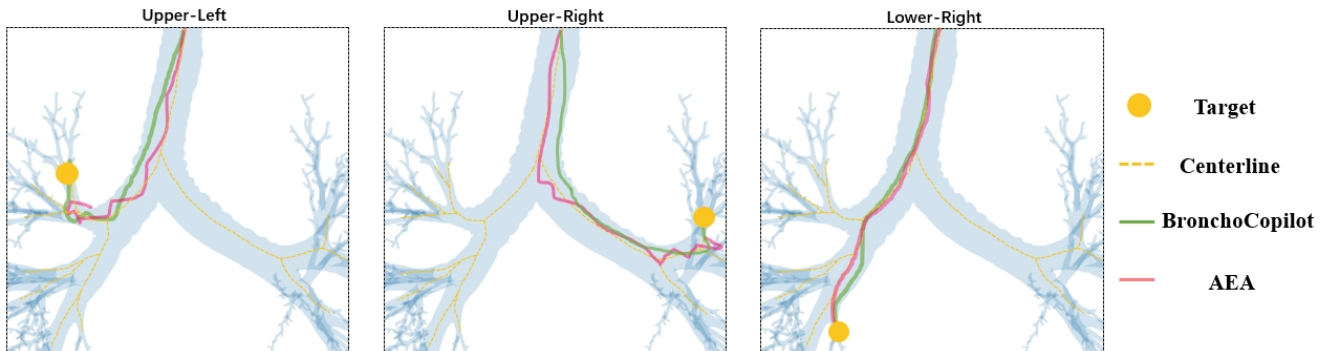


Fig. 5: The target positions in the fifth-level airways were selected for the upper left, upper right, and lower left lung lobes, respectively. The yellow, green, and pink lines represent the centerlines (reference paths), BronchoCopilot, and AEA robot tip trajectories. For visualization, all trajectories represent the average of three test runs.

TABLE I: **Overall performance comparison.** We evaluate 7 methods, including AEA, DQNN, BronchoCopilot without proprioception or vision, and using concat or sum for fusion in 3 tasks.

	Upper-Left				Upper-Right				Lower-Left			
	SR(%) $\uparrow$	NA $\downarrow$	TL $\downarrow$	F.M/A(10 $\wedge$ (-1)N) $\downarrow$	SR $\uparrow$	NA $\downarrow$	TL $\downarrow$	F.M/A(10 $\wedge$ (-1)N) $\downarrow$	SR $\uparrow$	NA $\downarrow$	TL $\downarrow$	F.M/A(10 $\wedge$ (-1)N) $\downarrow$
AEA	54.3( $\pm$ 6.4)	-	-	-	76.2( $\pm$ 3.3)	322.2( $\pm$ 12.5)	1.24( $\pm$ 0.15)	2.78/0.24	89.8( $\pm$ 1.8)	<b>251.4(<math>\pm</math>19.7)</b>	<b>0.94(<math>\pm</math>0.15)</b>	<b>3.34/0.22</b>
DQNN	0	-	-	-	0	-	-	-	23.8( $\pm$ 7.7)	-	-	-
ours w/o P	43.6( $\pm$ 6.7)	-	-	-	57.7( $\pm$ 4.9)	-	-	-	<b>92.6(<math>\pm</math>1.7)</b>	255.0( $\pm$ 30.7)	0.98( $\pm$ 0.16)	4.34/0.88
ours w/o V	91.2( $\pm$ 3.8)	272.2( $\pm$ 18.5)	1.39( $\pm$ 0.17)	<b>1.88/0.43</b>	73.3( $\pm$ 5.8)	332.2( $\pm$ 19.6)	1.15( $\pm$ 0.18)	11.22/1.44	86.7( $\pm$ 4.2)	298.6( $\pm$ 31.0)	1.03( $\pm$ 0.17)	12.10/1.57
ours-concat	91.5( $\pm$ 2.1)	255( $\pm$ 24.3)	0.99( $\pm$ 0.08)	3.44/0.25	82.7( $\pm$ 1.1)	344.4( $\pm$ 25.2)	1.03( $\pm$ 0.07)	1.55/0.38	63.3( $\pm$ 2.6)	-	-	-
ours-sum	83.4( $\pm$ 4.2)	352.2( $\pm$ 23.1)	1.32( $\pm$ 0.09)	9.87/0.74	85.6( $\pm$ 2.8)	328.1( $\pm$ 34.6)	1.07( $\pm$ 0.11)	2.94/0.77	28.6( $\pm$ 8.2)	-	-	-
ours	<b>97.1(<math>\pm</math>1.2)</b>	<b>223.2(<math>\pm</math>12.7)</b>	<b>0.96(<math>\pm</math>0.04)</b>	2.56/0.38	<b>95.5(<math>\pm</math>0.8)</b>	<b>289.3(<math>\pm</math>8.4)</b>	<b>0.95(<math>\pm</math>0.06)</b>	<b>1.67/0.25</b>	91.3( $\pm$ 2.4)	269.6( $\pm$ 23.8)	1.01( $\pm$ 0.13)	5.09/0.97

‘-’ indicates that the metric is not calculated due to a low success rate.

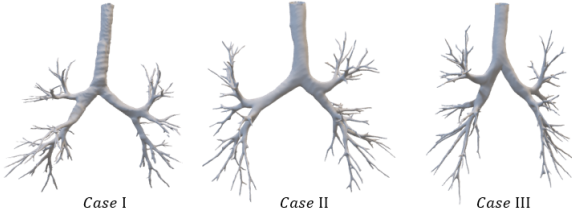


Fig. 6: Different airway models for transfer training.

actions numbered 2, 4, 6, 8 in the original setup, and adding a backward retreat action).

**Ablation on Multimodality:** This part involves training with only visual or proprioceptive input. The encoder is trained through the same reconstruction task and serves as the state vector, while the fusion module is removed.

**Ablation on Fusion Module:** We replace the cross-attention module with concatenation (concat) and summation (sum) individually while keeping other settings consistent.

Based on the reconstructed 3D model detailed in Section III. A, we selected three targets in the 5-th generation airways of the upper left, lower left, and upper right lung lobes for training. During the model training phase, we evaluated the agent’s accumulated reward and learning efficiency, which are depicted in the reward curve shown in Fig. 4. In the model evaluation phase, we conducted 80 insertion procedures, with the targets’ locations and insertion paths illustrated in Fig. 5. The metrics for evaluation include: (i) Success Rate (*SR*), defined when the robot’s tip is within 7mm of the target. This threshold corresponds to the length of the standard biopsy tool attached to the bronchoscope, as shown in Fig. 1(a). A task is considered a failure if, after 500 actions, the robot has not reached the target or has triggered the exit criteria described in Section III. (ii) Number of Actions (*NA*), representing the total actions taken to reach the target. (iii) Trajectory Length (*TL*), recording the distance the robot’s tip travels. To facilitate comparison across different experiments, all lengths are normalized by the reference path length to the target. (iv) The Maximum and Average contact Forces ( $F_{M/A}$ ), while contact is unavoidable, minimizing force is preferable. The dynamic insertion processes for targets in the upper left, lower left, and upper right lung lobes are shown in the attached video.

TABLE II: Mean value of Success Rate for different airway structures and its generations.

	2-th	3-th	4-th	5-th	6-th
Case I	100.0%	96.6%	91.2%	83.3%	59.8%
Case II	100.0%	92.4%	83.5%	79.3%	43.7%
Case III	100.0%	94.2%	88.7%	85.4%	76.6%

For **Transfer Experiment**, we fixed the parameters of *stage I & II* and exclusively trained the RL module. We randomly selected CT images of three cases from the EX-ACR’09 dataset [45] and performed the segmentation process as described in Section III. A. The anatomical structures of the cases are depicted in Fig. 6. We randomly selected three targets on the centerlines of airways at different levels for training. Subsequently, we conducted 80 evaluation trials to determine the success rate, with the findings presented in Table II.

### B. Implementation Details

For all networks involved in our experiments, we employed Kaiming initialization [46] for the weight matrices. The camera images were captured at a resolution of 512x512, and we utilized ResNet34 [47] as the visual encoder alongside a Multilayer Perceptron (MLP) for the proprioceptive encoder. The state representation model outputs a vector of dimension 64 (or dimension 128 when concatenated). In RL model, we implemented two MLPs for the actor and critic networks, respectively. Each network consists of five layers with 256 nodes and employs Tanh as the activation function. Our experiments were conducted using the PyTorch framework on a workstation equipped with an Intel i7-13700KF CPU and an NVIDIA RTX 4070 GPU. During *stages I & II*, we trained for 50 epochs on a dataset comprising approximately 30,000 images, and we set  $c = 3$  for consecutive frames of the visual observation. For *stage III*, the training extended over 500 episodes, with each episode capped at a maximum of 1000 steps. The average training duration was 2.2 hours, with the model typically reaching convergence between the 120th and 190th episodes.

## V. RESULT AND DISCUSSION

**The BronchoCopilot outperforms previous image-guided and RL-based bronchoscopy agents, with mul-**

**timodal information significantly contributes to the improved performance.** As shown in Table I and Fig. 5, in all three tasks, BronchoCopilot performs excellently. Compared to AEA and DQNN, our approach demonstrates nearly the best performance across four metrics in the upper lobe task. While AEA shows commendable results in the lower lobe task, where the centerline remains visible throughout the insertion without major turns, it struggles with the pre-bend strategy essential for accessing the upper lobe. DQNN, with its overly simplistic network design, fails to adequately capture image change patterns, leading to convergence issues during our training sessions.

For AEA and DQNN, we observed severe algorithmic failures when the robot's tip is very close to the airway wall. Comparatively, BronchoCopilot leverages proprioceptive data to bolster decision-making processes, achieving a success rate exceeding 90% for insertions into the 5-th generation airways. The LSTM-based visual module plays a crucial role in detecting shifts in imagery, which minimizes oscillations during insertion and contributes to a reduction in both *NA* and *TL*. Our method exhibits minimal variance across most evaluation metrics, indicating that BronchoCopilot can execute highly consistent actions. This consistency is vital for minimizing the risk of unforeseen events during surgeries. In tasks involving the lower lobe, where the pathway is smoother and requires less bending, our method is slightly inferior to more intuitive approach (AEA) due to considering more factors.

**The fusion approach of visual and proprioceptive information outperforms traditional fusion methods.** Through comparison tests between concatenation and summation methods, it's evident that BronchoCopilot's incorporation of cross-attention mechanisms significantly enhances performance. This cross-attention functionality enables the model to autonomously discern the interplay between modalities and dynamically modulate the weight assigned to each, ensuring an optimal blend of information for decision-making. The efficacy of this approach is clearly validated in the results presented in Table I, showcasing the robust advantage of cross-attention in multimodal information fusion.

**Our method is capable of rapid end-to-end training transfer across diverse surgical cases.** By freezing the parameters of the state representation model and focusing on training the decision model in an end-to-end fashion, we've shown that BronchoCopilot consistently achieves high success rates across diverse anatomical structures. Furthermore, the average training duration for tasks targeting the 5-th generation airways stands at merely 0.68 hours. This efficiency in end-to-end transfer training underscores our method's swift adaptability to different bronchoscopy cases, emphasizing its significant practical utility in the field.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we introduced BronchoCopilot, a multimodal reinforcement learning algorithm and training framework designed for the autonomous robotic bronchoscopy. Leveraging

the synergistic potential of visual and proprioceptive information, BronchoCopilot represents a significant advancement in the manipulation of dual-segment flexible bronchoscopy robot, particularly in more complex airway environments. Our staged training approach not only simplifies the complexity associated with training multimodal RL models but also enables rapid adaptation to diverse surgical scenarios through end-to-end transfer learning. In the future, the translation of BronchoCopilot to real-world clinical settings stand as the next frontier in our research. For example, the potential noises in real proprioceptive data may raise other issues, which needs more validation on physical robotic platforms.

## REFERENCES

- [1] 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.
- [2] R. J. Miller, R. F. Casal, D. R. Lazarus, D. E. Ost, and G. A. Eapen, "Flexible bronchoscopy," *Clinics in Chest Medicine*, vol. 39, no. 1, pp. 1–16, 2018.
- [3] A. Agrawal, D. K. Hogarth, and S. Murgu, "Robotic bronchoscopy for pulmonary lesions: a review of existing technologies and clinical data," *Journal of thoracic disease*, vol. 12, no. 6, p. 3279, 2020.
- [4] M. Davoudi and H. G. Colt, "Bronchoscopy simulation: a brief review," *Advances in health sciences education*, vol. 14, pp. 287–296, 2009.
- [5] J. Hoelscher, M. Fu, I. Fried, M. Emerson, T. E. Ertop, M. Rox, A. Kuntz, J. A. Akulian, R. J. Webster III, and R. Alterovitz, "Backward planning for a multi-stage steerable needle lung robot," *IEEE robotics and automation letters*, vol. 6, no. 2, pp. 3987–3994, 2021.
- [6] J. Rosell, A. Pérez, P. Cabras, and A. Rosell, "Motion planning for the virtual bronchoscopy," in *2012 IEEE International Conference on Robotics and Automation*, pp. 2932–2937, IEEE, 2012.
- [7] 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.
- [8] 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.
- [9] X. Tan, C.-B. Chng, Y. Su, K.-B. Lim, and C.-K. Chui, "Robot-assisted training in laparoscopy using deep reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 485–492, 2019.
- [10] J. Kweon, K. Kim, C. Lee, H. Kwon, J. Park, K. Song, Y. I. Kim, J. Park, I. Back, J.-H. Roh, *et al.*, "Deep reinforcement learning for guidewire navigation in coronary artery phantom," *IEEE Access*, vol. 9, pp. 166409–166422, 2021.
- [11] A. Segato, L. Sestini, A. Castellano, and E. De Momi, "Ga3c reinforcement learning for surgical steerable catheter path planning," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2429–2435, IEEE, 2020.
- [12] L. Karstensen, J. Ritter, J. Hatzl, T. Pätz, J. Langejürgen, C. Uhl, and F. Mathis-Ullrich, "Learning-based autonomous vascular guidewire navigation without human demonstration in the venous system of a porcine liver," *International Journal of Computer Assisted Radiology and Surgery*, vol. 17, no. 11, pp. 2033–2040, 2022.
- [13] S. Athinotiis, R. Srivatsan, and H. Choset, "Deep q reinforcement learning for autonomous navigation of surgical snake robot in confined spaces," in *Proceedings of the The Hamlyn Symposium on Medical Robotics, London, UK*, pp. 23–26, 2019.
- [14] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," *IEEE signal processing magazine*, vol. 34, no. 6, pp. 96–108, 2017.
- [15] J. Ma, F. Wu, Y. Chen, X. Ji, and Y. Ding, "Effective multimodal reinforcement learning with modality alignment and importance enhancement," *arXiv preprint arXiv:2302.09318*, 2023.

- [16] M. A. Lee, Y. Zhu, P. Zachares, M. Tan, K. Srinivasan, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg, "Making sense of vision and touch: Learning multimodal representations for contact-rich tasks," *IEEE Transactions on Robotics*, vol. 36, no. 3, pp. 582–596, 2020.
- [17] J. Chen, M. Chen, Q. Zhao, S. Wang, Y. Wang, Y. Xiao, J. Hu, D. T. M. Chan, K. T. L. Yeung, D. Y. C. Chan, *et al.*, "Design and visual servoing control of a hybrid dual-segment flexible neurosurgical robot for intraventricular biopsy," *arXiv preprint arXiv:2402.09679*, 2024.
- [18] K. Cleary, A. Melzer, V. Watson, G. Kronreif, and D. Stoianovici, "Interventional robotic systems: applications and technology state-of-the-art," *Minimally Invasive Therapy & Allied Technologies*, vol. 15, no. 2, pp. 101–113, 2006.
- [19] M. Yip and N. Das, "Robot autonomy for surgery," in *The Encyclopedia of MEDICAL ROBOTICS: Volume 1 Minimally Invasive Surgical Robotics*, pp. 281–313, World Scientific, 2019.
- [20] A. Segato, M. Di Marzo, S. Zucchelli, S. Galvan, R. Secoli, and E. De Momi, "Inverse reinforcement learning intra-operative path planning for steerable needle," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 6, pp. 1995–2005, 2021.
- [21] Y. Wang, J. Wang, Y. Li, T. Yang, and C. Ren, "The deep reinforcement learning-based vr training system with haptic guidance for catheterization skill transfer," *IEEE Sensors Journal*, vol. 22, no. 23, pp. 23356–23366, 2022.
- [22] A. Pore, Z. Li, D. Dall'Alba, A. Hernansanz, E. De Momi, A. Menciassi, A. C. Gelpí, J. Dankelman, P. Fiorini, and E. Vander Poorten, "Autonomous navigation for robot-assisted intraluminal and endovascular procedures: A systematic review," *IEEE Transactions on Robotics*, 2023.
- [23] A. Kuntz, L. G. Torres, R. H. Feins, R. J. Webster, and R. Alterovitz, "Motion planning for a three-stage multilumen transoral lung access system," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3255–3261, IEEE, 2015.
- [24] R. Khare, R. Bascom, and W. E. Higgins, "Hands-free system for bronchoscopy planning and guidance," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 12, pp. 2794–2811, 2015.
- [25] C. F. Graetzel, A. Sheehy, and D. P. Noonan, "Robotic bronchoscopy drive mode of the auris monarch platform," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 3895–3901, IEEE, 2019.
- [26] Y. Huang, C. Du, Z. Xue, X. Chen, H. Zhao, and L. Huang, "What makes multi-modal learning better than single (provably)," *Advances in Neural Information Processing Systems*, vol. 34, pp. 10944–10956, 2021.
- [27] S. Lacey and K. Sathian, "Crossmodal and multisensory interactions between vision and touch," in *Scholarpedia of touch*, pp. 301–315, Springer, 2015.
- [28] G.-H. Liu, A. Siravuru, S. Prabhakar, M. Veloso, and G. Kantor, "Learning end-to-end multimodal sensor policies for autonomous navigation," in *Conference on Robot Learning*, pp. 249–261, PMLR, 2017.
- [29] D. S. Chaplot, L. Lee, R. Salakhutdinov, D. Parikh, and D. Batra, "Embodied multimodal multitask learning," *arXiv preprint arXiv:1902.01385*, 2019.
- [30] D. Misra, J. Langford, and Y. Artzi, "Mapping instructions and visual observations to actions with reinforcement learning," *arXiv preprint arXiv:1704.08795*, 2017.
- [31] J. Hansen, F. Hogan, D. Rivkin, D. Meger, M. Jenkin, and G. Dudek, "Visuotactile-rl: learning multimodal manipulation policies with deep reinforcement learning," in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 8298–8304, IEEE, 2022.
- [32] S. Omidshafiei, D.-K. Kim, J. Papis, and J. P. How, "Crossmodal attentive skill learner," *arXiv preprint arXiv:1711.10314*, 2017.
- [33] A. Manchin, E. Abbasnejad, and A. Van Den Hengel, "Reinforcement learning with attention that works: A self-supervised approach," in *Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part V 26*, pp. 223–230, Springer, 2019.
- [34] H. Zheng, Y. Qin, Y. Gu, F. Xie, J. Yang, J. Sun, and G.-Z. Yang, "Alleviating class-wise gradient imbalance for pulmonary airway segmentation," *IEEE transactions on medical imaging*, vol. 40, no. 9, pp. 2452–2462, 2021.
- [35] D. Hearn, M. P. Baker, and M. P. Baker, *Computer graphics with OpenGL*, vol. 3. Pearson Prentice Hall Upper Saddle River, NJ., 2004.
- [36] A. Pore, M. Finocchiaro, D. Dall'Alba, A. Hernansanz, G. Ciuti, A. Arezzo, A. Menciassi, A. Casals, and P. Fiorini, "Colonoscopy navigation using end-to-end deep visuomotor control: A user study," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 9582–9588, IEEE, 2022.
- [37] R. Izzo, D. Steinman, S. Manini, and L. Antiga, "The vascular modeling toolkit: a python library for the analysis of tubular structures in medical images," *Journal of Open Source Software*, vol. 3, no. 25, p. 745, 2018.
- [38] J. Allard, S. Cotin, F. Faure, P.-J. Bensoussan, F. Poyer, C. Duriez, H. Delingette, and L. Grisoni, "Sofa-an open source framework for medical simulation," in *MMVR 15-Medicine Meets Virtual Reality*, vol. 125, pp. 13–18, IOP Press, 2007.
- [39] F. Faure, C. Duriez, H. Delingette, J. Allard, B. Gilles, S. Marchesseau, H. Talbot, H. Courtecuisse, G. Bousquet, I. Peterlik, *et al.*, "Sofa: A multi-model framework for interactive physical simulation," *Soft tissue biomechanical modeling for computer assisted surgery*, pp. 283–321, 2012.
- [40] X. Liu, J. Chen, J. Hu, H. Chen, Y. Huang, and H. Liu, "Multi-interface strain transfer modelling for flexible endoscope shape sensing," *IEEE Robotics and Automation Letters*, 2024.
- [41] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," *Advances in neural information processing systems*, vol. 28, 2015.
- [42] M. Tschannen, O. Bachem, and M. Lucic, "Recent advances in autoencoder-based representation learning," *arXiv preprint arXiv:1812.05069*, 2018.
- [43] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
- [44] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, "High-dimensional continuous control using generalized advantage estimation," *arXiv preprint arXiv:1506.02438*, 2015.
- [45] P. Lo, B. Van Ginneken, J. M. Reinhardt, T. Yavarna, P. A. De Jong, B. Irving, C. Fetita, M. Ortner, R. Pinho, J. Sijbers, *et al.*, "Extraction of airways from ct (exact'09)," *IEEE Transactions on Medical Imaging*, vol. 31, no. 11, pp. 2093–2107, 2012.
- [46] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034, 2015.
- [47] 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.