Lightweight Handheld Detachable Compliant Robotic Laryngoscope with Lightweight Intelligent Visual Guidance

Tao Zhang, Sattar Kadir, Hongyu Geng, Huilin Pan, An Wang, Jiewen Lai* and Hongliang Ren*

***Abstract*— Tracheal intubation (TI) is a routine procedure in hospitals and is often life-saving for patients who need respiratory assistance. However, the current intubation approach requires the physicians to have extensive experience, or there will be severe time waste or possible airway damage. Aiming to assist the physician in performing better and easier TI, this paper proposes a novel, lightweight, portable, detachable compliant robotic laryngoscope with a lightweight intelligent guidance module to assist with transoral tracheal intubation. The robotic laryngoscope comprises an endoscope-equipped steerable flexible segment on a detachable module and an ergonomic handle homed to the control units, a micro-computer, a joystick, a battery, and an LCD monitor. The steerable flexible segment with a tip camera is designed to fit the shape of the upper airway and identify the glottis in real-time. An additional anchoring mechanism at the base of the steerable segment has been specially designed to enhance the robot's intraluminal motion and laryngoscopic vision stability. A self-contained, lightweight learning-based glottis detection algorithm (within 5 MB parameters) is deployed in the portable device without the need to access additional servers or clouds. The system also adopts a modular design, where the robotic section and the driving unit can be detached and reassembled swiftly between the TI procedures. Finally, the working performance is verified by experiments. The result shows that the motion error of the steerable segment is less than 2.7% over its length, and the steerable segment can be inserted into the upper airway easily while the glottis can be detected in real-time.**

I. Introduction

Tracheal intubation (TI) is one of the most effective ways to ensure the airway patency of patients requiring respiratory assistance [1] (e.g., in intensive care units, ICU). However, the lack of dexterity and intelligent visual feedback of the traditional blade-like laryngoscope would require a well-trained physician. A novice may need a longer operation time and sometimes cause misplacement (e.g., inserting a tracheal tube into the esophagus instead of the respiratory tract), which would lead to asphyxia, hypoxia, and pulmonary aspiration [2]. When TI is performed under difficult and stressful circumstances, these medical accidents occur more frequently [3]. To improve the situation, we propose a compliant robotic laryngoscope with intelligent visual guidance to assist with the TI approach. Unlike the rigid laryngoscope with a blade-like arc only responsible for airway opening and glottis visualization, our newly proposed compliant laryngoscope is flexible and steerable. It can be used as an endotracheal guide wire that allows the tracheal tube (TT) to be placed into the patient's trachea through the compliant laryngoscope and also provides a direct view at its steerable tip. Due to the direct view of the glottis and reduction of involved equipment (i.e., intubation guide wire is no longer needed), this approach can improve the efficiency of the TI process.

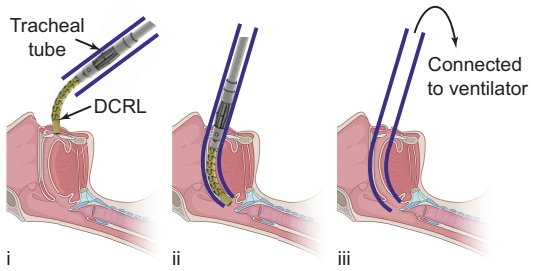

Fig. 1. Key procedures for transoral tracheal intubation when using our robotic laryngoscope. (i) Place and steer the compliant laryngoscope (steerable flexible segment) into the upper airway and fix its pose; (ii) Insert the tracheal tube along the flexible segment; (iii) Remove the robotic laryngoscope. The intubation is done without using an intubation guide wire.

With the development of transoral surgical robots [4], [5], robotic TI has also drawn the attention of the flexible robots community. For instance, a 3D-printed curved laryngoscope that integrated three flexible manipulators was proposed for the treatment of early-stage laryngeal cancer [6]. The curvature of the laryngoscope was predetermined to fit the shape of the human upper airway. However, the rigid structure of the laryngoscope still needs to be more adaptive to different individual patients. To improve the capacity of the laryngoscope to navigate the upper airways of various patients, new designs of laryngoscope incorporate flexible mechanisms [7], [8]. Particularly, an intelligent steerable endoscope, with feature recognition and autonomous tip adjustment, was adopted with a laryngoscope together to lead the tracheal tube into the glottis [9]. However, the system only robotized the steerable guide wire, and three elements still need to work together to carry out the intubation task. Zhao *et al.* [10] introduced a parallel continuum robot (PCR)

This work was supported in part by the Hong Kong Research Grants Council (RGC) Collaborative Research Fund under grant CRF C4026-21GF; in part by CUHK Direct Grant for Research under grant 4055213; in part by IdeaBooster Fund Award under grant 3230391; in part by the Hong Kong Research Grants Council (RGC) Research Impact Fund under grant R4020-22. (*Corresponding authors: Jiewen Lai and Hongliang Ren.*)

T. Zhang, H. Geng, A. Wang, J. Lai, and H. Ren are with the Department of Electronics Engineering, The Chinese University of Hong Kong, Hong Kong, China {tzhang,hygeng,wa09}@link.cuhk.edu.hk; {jwlai,hlren}@ee.cuhk.edu.hk

S. Kadir is with the Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong, China sattar.kadir@link.cuhk.edu.hk

H. Pan is with the Department of Electronics Engineering, The Chinese University of Hong Kong, and the Department of Artificial Intelligence, Huazhong University of Science and Technology, Wuhan, China huilinpan@hust.edu.cn

and applied it to a transoral laser phonosurgery robot, where three Bowden cables are connected in parallel with a curved laryngeal blade with a scaffold to adjust the curvature of the blade directly, while the long PCR possesses a large swept area when increasing the blade curvature. Yet, the swept area could interfere with the upper airway and affect the surgery performance.

In this paper, we propose a detachable compliant robotic laryngoscope (DCRL) with intelligent visual guidance. In our DCRL, a steerable flexible segment with a central spring is employed to replace the rigid blade-like laryngoscope. The design significantly enhances the flexibility and passage capacity of the existing laryngoscope. Moreover, a detachable and modular design [11] is adopted to realize the rapid assembly of the flexible part that directly contacts the patient's airway. A flexible anchoring mechanism is designed and utilized to stabilize the flexible segment whenever necessary. An LCD touchscreen is integrated into the portable handle. It can provide an endoscope view and the detected glottis enabled by a learning-based method, which makes intubation more intuitive and convenient. The standalone DCRL comprises a micro-computer and a joystick powered by a built-in battery that all fit into a portable handle. Using our developed system, the TI is now divided into three key steps (as shown in Fig. 1). First, the physician will need to place and steer the flexible segment into the upper airway with the assistance of our intelligent visual guidance and fix its pose when it reaches the target site. After that, the physician needs to insert the tracheal tube along the flexible segment. Lastly, the DCRL shall be removed, and the tracheal tube should stay put. With our system, the intubation can be done without using an intubation guide wire.

The core contributions of this work include the following:

- **Lightweight Compliant Robotic Laryngoscope:** An all-in-one lightweight detachable compliant robotic laryngoscope (0.65 kg) is developed. The robotic laryngoscope comprises a steerable flexible segment with an endoscopic camera and a three-DoF (bending, axial rotation, and anchoring) flexible segment that can be steerable in one hand.
- **Detachability:** A detachable design is adopted to allow the disassembly and assembly of the steerable flexible segment swiftly to ensure sanitation.
- **Lightweight Intelligent Visual Guidance:** A self-contained, lightweight glottis detection network displays the target glottis at the laryngoscope monitor in real time is proposed.

II. Prototype Design

As shown in Fig. 2, the prototype of the DRCL system can be decomposed into two subsystems, viz., the compliant manipulator (patient side) and the driving and control system (physician side). The patient-side robotic system is designed to be consumable for the sake of sanitation, and it comprises only mechanical parts (excluding the endoscopic camera), including a gearbox, actuation wires, a flexible anchor, and a flexible segment. The physician-side system is reusable, as it contains an LCD monitor, a joystick, a microcomputer, a

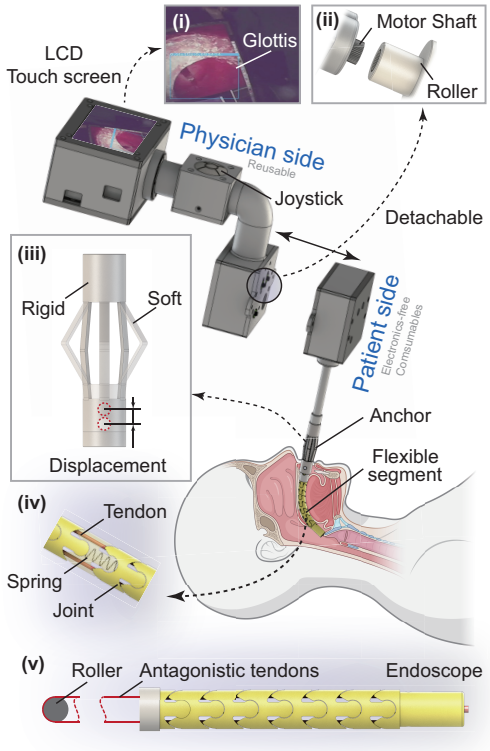

Fig. 2. A schematics of the overall structure of the portable compliant laryngoscope. (i) View of the camera; (ii) The detachable mechanism; (iii) Working principle of the proposed flexible anchoring mechanism; (iv) Mechanism of the steerable flexible segment that is specially designed for TI application; (v) Driving mechanism for the steerable segment.

motor control unit, and two batteries—all packed inside a single portable handle. The patient-side and physician-side structures can be rapidly connected and detached from each other through interference fit join (see Fig. 2(ii)) so that the physician-side structures can be more durable as it does not require standard sterilization. In addition, different compliant manipulators can be interchanged conveniently.

The detachable flexible manipulator has three DoFs, including (1) the bending and (2) continuous axial rotation of the continuum arm, as well as (3) the anchor deployment. The continuum arm is designed to fit the normal human pharyngeal cavity shape, and the flexible anchor stabilizes the laryngoscopic vision in the intraluminal environment without harming the surrounding soft tissue.

A. Patient-side Compliant Manipulator

The patient-side manipulator consists of a rigid shaft, a flexible anchor, and a steerable flexible segment. The flexible segment constitutes a series of 3D-printed vertebrae with an outer diameter of 8 mm and an inner diameter of 4 mm, a central spring, an endoscopic camera, and a pair of antagonistic driving wires. The vertebrae and the central spring work together as the backbone to provide the stiffness of the compliant structure (length: 74 mm). The endoscopic camera is employed to capture the eye-in-hand view at the distal tip to identify the human glottis, facilitating the physician to steer the robot. The flexible segment has two DoFs, and the shape of the vertebrae is explicitly designed to program the maximum bending angle and to fit the curvature

Fig. 3. The worm gears mechanism at the patient side allows continuous rotational motion ($\pm$ 180°) of the manipulator. Such a continuous rotation would help insert the flexible robot into narrow and frictional cavities.

of the general upper airway. The flexible anchor is designed to stabilize the DCRL within the respiratory tract when the physicians try to insert the flexible segment into the trachea. As shown in Fig. 2(ii), the anchor's structure comprises a series of deployable soft circumferential strips. This mechanism exhibits two distinct operational modes: the retracted and deployed configurations. The transition between these two configurations is facilitated through the linear actuation assembly. The actuation unit consists of a cable-driven push-pull rod integrated with a fixed spring, which enables linear forward and reverse motions. In addition, the overall flexible robot and the rigid shaft can be axially rotated up to $\pm$180°, enabled by the worm gears mechanisms as shown in Fig. 3. The continuous rotation would help insert the robot into narrow and frictional cavities [12].

B. Physician-side Driving and Control System

The driving system has three servo motors (DC, Tower-Pro), and their motor shaft can be firmly attached to the roller of the patient-side unit by interference fit. The antagonistic cables are actuated in pairs by a single motor through a roller to achieve an equal amount of cable pulling and releasing (as shown in Fig. 2(v)). The worm gears control the rigid shaft's rotation without altering the bending cables within a specific range of rotation, which is controlled by the second motor joint. The flexible anchor can be deployed and retracted by another cable driven by the third motor.

A Raspberry Pi powered by two 1865 batteries is used for the micro-computer. The micro-computer controls the servo motors through an STM32 microcontroller, and the command signals are assigned by a 5-key joystick (up, down, left, right, and OK). The left-right keys are responsible for the bending, and the up-down keys are for rotation. The OK key controls the anchor deployment, with the deploying and retracting motion distinguished by short press and long press, respectively. The tip-mounted endoscopic camera (1.6 mm, OV6946, OmniVision) is connected to the micro-computer directly, and its vision can be streamed on the LCD monitor. All the components are jam-packed into a carefully designed frame (handle). At the prototype stage, we printed the handle using PLA material.

III. Kinematic Model

A. Forward Kinematics

As shown in Fig. 4, l_i and d are the length of each joint and the distance between two wire holes. $\{C_0\}$ is

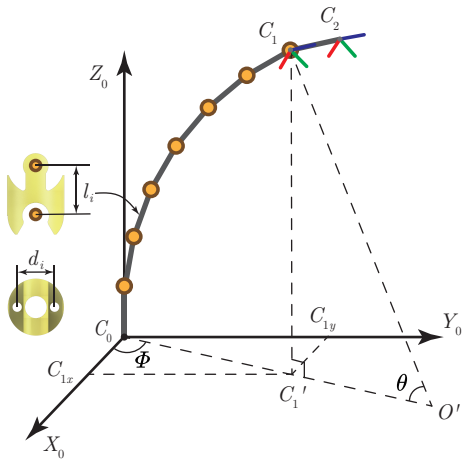

Fig. 4. Simplified structure of the flexible segment and the fixed coordinate systems.

the coordinate system fixed on the proximal tip center of the flexible segment. $\{C_1\}$ denotes the coordinate system fixed at the base of the distal joint, and $\{C_2\}$ represents the coordinate frame fixed at the distal tip (i.e., the camera frame). $\{C_1{}'\}$ is the projection of $\{C_1\}$ w.r.t. $\{C_0\}$, and c_{1x} and c_{1y} are the X and Y coordinates of $\{C_1\}$, respectively. ϕ and θ denote the deflection and bending angles.

We can assume that the flexible segment follows the constant curvature assumption. The kinematic model can be divided into two segments: the relationship between the driving space and the configuration space and the relationship between the configuration space and the working space. The former relationship can be expressed as follows:

$$l_i = nl_i + (-1)^i \cdot \frac{\theta d_i}{2}, \quad \text{given} \quad i \in \{1, 2\} \tag{1}$$

such that

$$\theta = (-1)^i \cdot \frac{2\,(l_i - nl_i)}{d_i}, \tag{2}$$

where l_1 and l_2 are the lengths of the pulling and the pushing wires, respectively.

Considering that the deflection and bending of the flexible segment is decoupled, the transfer matrix between the $\{C_0\}$, $\{C_1\}$, and $\{C_2\}$ can be expressed as

$$\begin{cases} {}^0_1T = \text{Trans}\,(P_{10})\,\text{Rot}\,(z, \phi)\,\text{Rot}\,(y, \theta) \\ {}^1_2T = \text{Trans}\,(P_{21}) \\ {}^0_2T = {}^0_1T \cdot {}^1_2T \end{cases}, \tag{3}$$

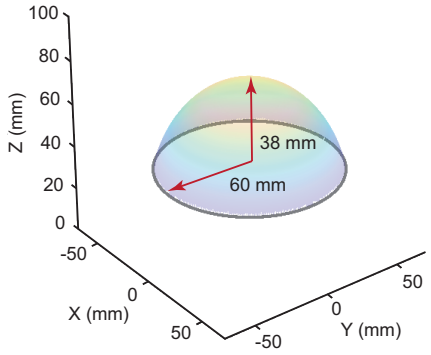

Fig. 5. The distal workspace of the steerable flexible segment with a length of 74 mm.

where

$$P_{10} = \left[\frac{nl_i}{\theta}(1-\mathrm{c}\theta)\,\mathrm{c}\phi \quad \frac{nl_i}{\theta}(1-\mathrm{c}\theta)\,\mathrm{s}\phi \quad \frac{nl_i}{\theta}\mathrm{s}\theta\right]^\top,$$
$$P_{21} = \begin{bmatrix}0 & 0 & l_t\end{bmatrix}^\top$$

where c and s refer to $\cos(\,)$ and $\sin(\,)$, respectively; l_t is the distance between the camera and the former joint; $\mathrm{Trans}(\,)$ and $\mathrm{Rot}(\,)$ denote the transformation and rotation matrices.

B. Inverse Kinematics

For the inverse kinematic model, we assume the position of the camera is $P = [P_x, P_y, P_z]^\top$, and the deflection angle can be obtained by

$$\phi = \arctan\left(\frac{P_y}{P_x}\right), \tag{4}$$

Where the bending angle θ can be obtained by solving the following equation.

$$P_z = \frac{nl_i}{\theta}\sin\theta + l_t\cos\theta, \tag{5}$$

which can be further simplified to avoid infinity caused by $\lim_{\theta\to 0}\frac{\sin\theta}{\theta}$. Here, we rewrite the terms with the Maclaurin series such that

$$\begin{cases}\sin\theta = x - \dfrac{\theta^3}{3!} + \dfrac{\theta^5}{5!} + o\left(\theta^7\right)\\ \cos\theta = 1 - \dfrac{\theta^2}{2!} + \dfrac{\theta^4}{4!} + o\left(\theta^6\right)\end{cases}. \tag{6}$$

By substituting (6) into the relevant terms in (5), we can obtain

$$P_z = nl_i\left(1 - \frac{\theta^2}{3!} + \frac{(\theta^2)^2}{5!}\right) + l_t\left(1 - \frac{\theta^2}{2!} + \frac{(\theta^2)^2}{4!}\right), \tag{7}$$

which can be solved directly by using the quadratic formula.

C. Workspace Analysis

To simulate the workspace of the flexible segment, we first assume that the bending angle can range from -90° to 90°, and the axial rotation can range from -180° to 180°. Note that the bending angle is designed to have a maximum of 90° bending through controlling the vertebra parameters and number of amounts. As shown in Fig. 5, the reachable workspace demonstrates a semi-hemisphere with a base radius of 60 mm and a height of 38 mm, which can mostly cover the oral and respiratory cavities.

IV. Glottis Detection

Glottis detection algorithms have been studied in various literature [7], [9], [13]. To further enhance the glottis detection performance, we introduce a novel backbone network and an enhanced strategy for the neck section, thus improving the network feature extraction performance. Besides, a dynamic label assignment strategy and a hybrid loss function tailored for the characteristics of glottis images are also adopted, making the training process more stable and efficient. The framework of our glottis detection network is depicted in Fig. 6.

A. Detection Backbone

On mobile devices, ShuffleNetV2 [14] has demonstrated superior robustness as the backbone component through extensive studies. In our glottis detection task, we have adopted optimization approaches based on PP-LCNet [15] and an improved backbone network to realize the lightweight glottis detection, i.e., Enhanced ShuffleNet (ES-Net) [16], to construct a more robust feature extractor. As shown in Fig. 6, the ES-Net consists primarily of several ES blocks, with the output feature maps of the C3, C4, and C5 layers serving as inputs to the neck section. Notably, we incorporate the Squeeze-and-Excitation (SE) modules [17] across all ES blocks, which can effectively weigh the network channels and obtain more discriminative features.

The channel shuffle in ShuffleNetV2 [14] facilitates information exchange between channels, but this operation would result in a loss of certain features during feature fusion. To dissolve this issue, when the stride is set to 2, depthwise and pointwise convolution are introduced to amalgamate the feature information from diverse channels. In addition, the ghost block adopted in GhostNet [18] can generate diverse feature maps, thus improving network learning efficiency. Therefore, when the stride is set to 1, the ghost block is integrated into the ES Block to learn more representative features.

B. Detection Neck

In the neck segment, Cross Stage Partial (CSP) is employed to construct the CSP-PAN (Path Aggregation network) architecture. In specific, the PAN structure can capture feature maps at different layers, while the CSP structure realizes the concatenation and fusion of adjacent feature maps. The initial CSP-PAN has identical channel numbers in each output feature map and the input feature maps of the backbone. However, the design with a large number of channels incurs a high computational cost on resource-constrained mobile devices. To address this issue, we adopt a 1×1 convolution to unify the input channel numbers of all branches, setting them to the minimum (i.e., 96). Then, the CSP module can achieve the feature fusion in both a top-down and a bottom-up manner. In this way, we can reduce the number of parameters meanwhile strengthening feature consolidation in the detection neck. Finally, we replaced all convolutional layers, except for the 1×1 convolutions,

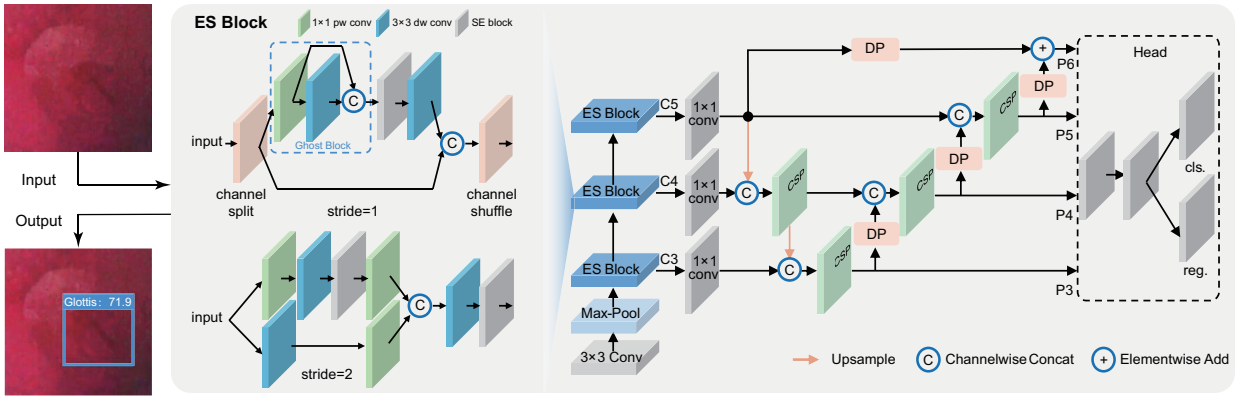

Fig. 6. The algorithm framework of the glottis detection.

with 5×5 depthwise separable convolutions to substitute the Channel-Shuffle operation in ShuffleNetV2 [14], thereby expanding the reception field with minimal parameters.

C. Label Assignment and Loss

Traditional label assignment strategies often employ static methods, making it challenging to adapt to diverse generalized data during the global training process. RetinaNet [19] differentiates positive and negative samples based on the IoU values between anchors and ground truth boxes relative to a predefined threshold. In contrast, FCOS [20] determines positive and negative samples by checking whether the anchor center points fall within the ground truth boxes. To address this issue, we adopt the SimOTA [21], a dynamic label assignment strategy, to optimize the model's training process. Unlike conventional static label assignment methods, SimOTA [21] operates online for label allocation. In each iteration, it dynamically adjusts label assignments based on the current models predictions and the similarity to the ground truth boxes, thereby selecting the most appropriate target boxes and enhancing the attention given to hard-to-classify or ambiguously bounded targets. The similarity calculation does not solely focus on the IoU; it also considers the category information and spatial relationships of the target boxes, which helps the model better assess the quality and significance of each box, leading to more informed label assignment decisions.

To facilitate dynamic allocation during the training process with SimOTA [21], we employ Varifocal Loss [22] in the detection head to couple classification predictions with quality predictions. Besides, GIoU Loss and Distribution Focal Loss are adopted for the regression tasks, i.e.,

$$\mathcal{L} = \mathcal{L}_{vfl} + 2\mathcal{L}_{giou} + 0.5\mathcal{L}_{dfl}, \tag{8}$$

where $\mathcal{L}_{vfl}$ denotes Varifocal loss, $\mathcal{L}_{giou}$ refers to GIoU loss, and $\mathcal{L}_{dfl}$ represents distribution focal loss.

D. Network Training and Inference

During training, stochastic gradient descent (SGD) was utilized as the optimizer, with a momentum of 0.9 and a weight decay of 4e-5. We set the batch size to 64 and trained all the models for 300 epochs on an NVIDIA RTX 3090 GPU. The learning rate started from 0.1 and was updated following the cosine decay scheduling. To deploy the network on the Raspberry Pi, which is only 85 mm × 56 mm and equipped with a 2.4 GHz CPU and an 800 MHz GPU, the Paddle-Lite inference framework is adopted to achieve real-time glottis detection on the proposed DCRL system.

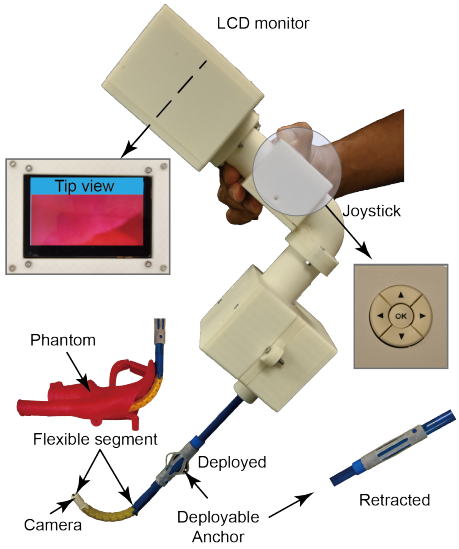

Fig. 7. The prototype of the proposed all-in-one portable Detachable Compliant Robotic Laryngoscope (DCRL) system to assist with transoral tracheal intubation. With the help of our embedded intelligent vision guidance, physicians can use this standalone device with one hand to steer the flexible segment into the trachea. The entire system weighs 0.65 kg.

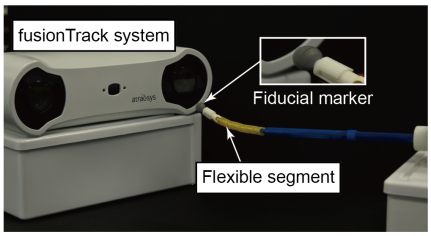

Fig. 8. The setup of kinematic model verification experiment.

V. Experiments

A. Model Verification

As shown in Fig. 7, the prototype has been fabricated. More details can be found at Sec. II. As shown in Fig. 8, the tip position of the flexible segment was obtained by a desktop motion tracking system (fusionTrack 250, Atracsys

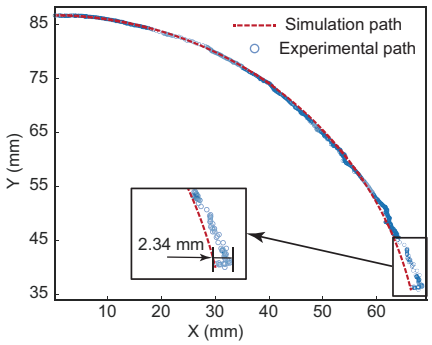

Fig. 9. Comparison of the experimental and simulation bending motion paths.

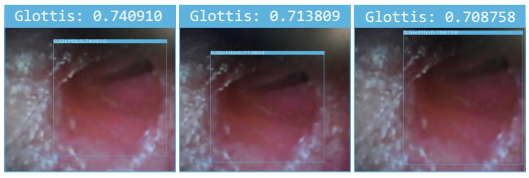

Fig. 10. Glottis detection (with zoomed-in captions) from different views.

LLC, Switzerland). A silver fiducial marker (radius: 5.85 mm) was attached to the distal tip to track the 3D position. Then, we can obtain the kinematic model performance by comparing the experimental and the simulation path (Fig. 9). The experiment result indicates that the position error is smaller than 2.34 mm, which is 2.7% over the total length of the flexible segment and the fiducial marker.

B. Verification on Glottis Detection

The preliminary results of the glottis detection network are shown in Fig. 10. We used an upper respiratory tract with lifelike glottis to test our system. The blue frames show the position of the glottis in the 2D image frame. Under the blurry condition, our detector could identify the glottis with over 70% confidence coefficient and a dynamic bounding box, with a model size of less than 5 MB and a time cost of around 50 ms. The result shows that the proposed detection algorithm can effectively detect the glottis in real time, which can potentially guide physicians and beginners in aligning the robot tip with the target site.

VI. Conclusions

In this paper, an all-in-one lightweight, portable, and detachable compliant robotic laryngoscope with a lightweight visual guidance network has been proposed. A flexible segment is designed to fit the curvature of the upper airway, an anchor is adopted to enhance the stability of the laryngoscope, an endoscopic camera is used to obtain the tip view, a glottis detection network is deployed to a Raspberry Pi, and an LCD screen is incorporated for on-device visualization. All these components are packed in a portable size so that the system can be held and steered single-handed. In addition, our system practices a modular design in which the patient-side tool can be replaced swiftly by the physician-side tool. However, our current system remains a functional preliminary prototype, and extra work is required for clinical trials, such as dimension optimization of the steerable segment and ergonomic optimization for the system.

References

[1] E. B. Thomas and S. Moss, "Tracheal intubation," *Anaesth. Intensiv. Care Med.*, vol. 15, no. 1, pp. 5–7, 2014.

[2] J. L. Apfelbaum, C. A. Hagberg *et al.*, "Practice guidelines for management of the difficult airway: an updated report by the american society of anesthesiologists task force on management of the difficult airway," *Anesthesiology*, vol. 118, no. 2, pp. 251–270, 2013.

[3] P. E. Pepe, L. P. Roppolo, and R. L. Fowler, "Prehospital endotracheal intubation: elemental or detrimental?" *Crit. Care*, vol. 19, pp. 1–7, 2015.

[4] X. Gu and H. Ren, "A survey of transoral robotic mechanisms: Distal dexterity, variable stiffness, and triangulation," *Cyborg Bionic Syst.*, vol. 4, p. 0007, 2023.

[5] C. Li, X. Gu, X. Xiao, C. M. Lim, and H. Ren, "Compliant and flexible robotic system with parallel continuum mechanism for transoral surgery: A pilot cadaveric study," *Robotics*, vol. 11, no. 6, p. 135, 2022.

[6] L. L. Kienle, L. R. Schild, F. Böhm, R. Grässlin, J. Greve, T. K. Hoffmann, and P. J. Schuler, "A novel 3d-printed laryngoscope with integrated working channels for laryngeal surgery," *Front. Surg.*, vol. 10, p. 906151, 2023.

[7] J. Lai, T.-A. Ren, W. Yue, S. Su, J. Y. Chan, and H. Ren, "Sim-to-real transfer of soft robotic navigation strategies that learns from the virtual eye-in-hand vision," *IEEE Trans. Ind. Inform.*, vol. 20, no. 2, pp. 2365–2377, 2024.

[8] F. Feng, Y. Zhou, W. Hong, K. Li, and L. Xie, "Development and experiments of a continuum robotic system for transoral laryngeal surgery," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 17, no. 3, pp. 497–505, 2022.

[9] Q. Boehler, D. S. Gage, P. Hofmann, A. Gehring, C. Chautems, D. R. Spahn, P. Biro, and B. J. Nelson, "Realiti: A robotic endoscope automated via laryngeal imaging for tracheal intubation," *IEEE Trans. Med. Robotics Bionics*, vol. 2, no. 2, pp. 157–164, 2020.

[10] M. Zhao, T. J. O. Vrielink, A. A. Kogkas, M. S. Runciman, D. S. Elson, and G. P. Mylonas, "Laryngotors: A novel cable-driven parallel robotic system for transoral laser phonosurgery," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1516–1523, 2020.

[11] H. Gao, X. Yang, X. Xiao, X. Zhu, T. Zhang, C. Hou, H. Liu, M. Q.-H. Meng, L. Sun, X. Zuo *et al.*, "Transendoscopic flexible parallel continuum robotic mechanism for bimanual endoscopic submucosal dissection," *Int. J. Robot. Res.*, vol. 43, no. 3, pp. 281–304, 2024.

[12] Q. Zhao, J. Lai, X. Hu, and H. K. Chu, "Dual-segment continuum robot with continuous rotational motion along the deformable backbone," *IEEE/ASME Trans. Mechatron.*, vol. 27, no. 6, pp. 4994–5004, 2022.

[13] E. Kruse, M. Döllinger, A. Schützenberger, and A. M. Kist, "Glottisnetv2: temporal glottal midline detection using deep convolutional neural networks," *IEEE J. Transl. Eng. Health Med.*, vol. 11, pp. 137–144, 2023.

[14] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Proc. of Euro. Conf. Comput. Vis. (ECCV)*, 2018, pp. 116–131.

[15] C. Cui, T. Gao, S. Wei, Y. Du, R. Guo, S. Dong, B. Lu, Y. Zhou, X. Lv, Q. Liu *et al.*, "Pp-lcnet: A lightweight cpu convolutional neural network," *arXiv preprint arXiv:2109.15099*, 2021.

[16] G. Yu, Q. Chang, W. Lv, C. Xu, C. Cui, W. Ji, Q. Dang, K. Deng, G. Wang, Y. Du *et al.*, "Pp-picodet: A better real-time object detector on mobile devices," *arXiv preprint arXiv:2111.00902*, 2021.

[17] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 7132–7141.

[18] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "Ghostnet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 1577–1586.

[19] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2980–2988.

[20] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, 2019.

[21] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," *arXiv preprint arXiv:2107.08430*, 2021.

[22] H. Zhang, Y. Wang, F. Dayoub, and N. Sunderhauf, "Varifocalnet: An iou-aware dense object detector," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 8514–8523.