IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 23 October 2023, accepted 14 November 2023, date of publication 7 December 2023,
date of current version 4 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3340677

RESEARCH ARTICLE

An End-to-End Learning-Based Control Signal Prediction for Autonomous Robotic Colonoscopy

VAN SY NGUYEN[1,2], BOHYUN HWANG[1], BYUNGKYU KIM[3], AND JAY HOON JUNG[4]

[1]Department of Aerospace Engineering and Mechanical Engineering, Korea Aerospace University, Goyang 10540, South Korea
[2]Mechanical and Aerospace Engineering Department, University of Central Florida, Orlando, FL 32816, USA
[3]School of Aerospace and Mechanical Engineering, Korea Aerospace University, Goyang 10540, South Korea
[4]School of Artificial Intelligence, Korea Aerospace University, Goyang 10540, South Korea

Corresponding authors: Byungkyu Kim (bkim@kau.ac.kr) and Jay Hoon Jung (jhjung@kau.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) funded by the Korean Government (MSIT) under Grant 2021R1A2C2010903 and Grant 2021R1F1A106072513.

ABSTRACT We introduce a novel 3 degrees-of-freedom based robotic colonoscopy system that performs the necessary movements for colonoscopy while working within the movement range of a flexible colonoscope (FC). In addition, we have developed deep learning models to generate motor control signals directly from input images without the need for motor control signal labels. The first presented model comprises a deep learning algorithm for predicting steering points and an image-based visual servo control (IBVS) algorithm for generating the motor control signal. The experiments showed that the proposed model's cecal intubation time (CIT) and rate (CIR) are comparable to those of human operators, despite requiring a shorter training time. Furthermore, we propose a model that replaces the IBVS algorithm with a deep learning algorithm that does not rely on rotation angles. The second model showed similar CIT (165s) and CIR (92%) compared to the first model. Finally, the last model, which solely comprises a single deep learning algorithm, demonstrates a reduction in CIT (127s) and an increase in CIR (96%), resulting in reduced physical demand for operators, improved safety, and shorter patient recovery time.

INDEX TERMS Autonomous system, deep learning, robotic colonoscopy, visual servo control.

I. INTRODUCTION

The prevalence of colorectal cancers (CRC), the third leading cause of cancer death [1], is increasing among middle-aged and elderly individuals and is expected to rise in the future [2]. Early detection and monitoring of potential cancers through colonoscopy procedures significantly increase the 5-year survival rate by around 90% [3] and are essential for avoiding hospitalization. However, the growing rate of patients diagnosed with CRC places a heavy workload on medical staff, potentially leading to doctor fatigue and more unsafe incidents during procedures [4]. Manipulating a colonoscope accurately is challenging due to the combination of various movements such as pushing, pulling, rotating, and steering motions [5]. Therefore, substantial efforts have been focused on new developments which are aimed to improve the effectiveness and efficiency of CRC diagnosis while alleviating physical burdens that endoscopists may experience, such as operator fatigue and musculoskeletal injuries.

Various mechanical innovations, such as a conventional colonoscope's knob [6], and user-friendly joysticks [7], [8], have been studied to develop new devices that are easier to use than traditional FCs and help reduce surgeons' workload. Additionally, several master-slave systems [9] and haptic devices [10] have been developed to assist operators in precisely controlling the device, leading to reduced operator effort and improved safety. Furthermore, the further development of semi-autonomous or autonomous control systems has emerged as a potential strategy. Recent studies on autonomous, magnetic navigation-based capsules have shown promising progress towards enhancing and ensuring patient safety. These magnetic capsule endoscopes are designed with movements controlled by a robot arm [11], [12], [13]. By using external and internal magnetic capsules, endoscopists can manipulate the capsule with less friction,

The associate editor coordinating the review of this manuscript and approving it for publication was Yangmin Li.

reducing patient discomfort, and resulting in efficient and smooth diagnoses.

Despite remarkable progress in mechanism, navigation or autonomous operation design in the field of colonoscopy, there are still technical challenges that require attention. These challenges include inadequate mobility in assisted-robot colonoscopy [6], [7], [8], unreliable approaches for lumen detection and navigation [14], [15], [16], [17], [18], [19], [20], lengthy intubation times [13], [22], [23], low success rates [11], [13], and inadequately developed autonomous systems [11], [13], [22]. Additionally, magnetic-based systems have the drawback of requiring a more prolonged intubation time, potentially leading to increased patient discomfort or pain. Moreover, unstable or fluctuating magnetic forces and electromagnetic interference can cause control inaccuracies.

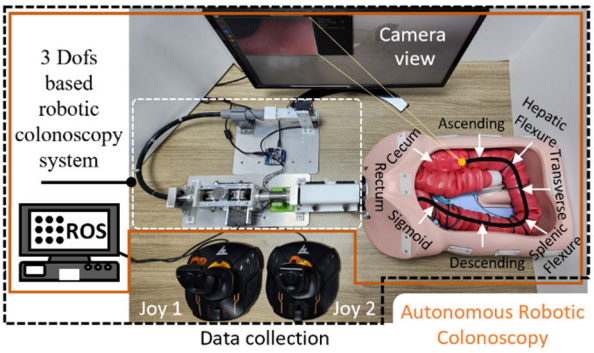

FIGURE 1. The overview of an autonomous robotic colonoscopy system.

To address these shortcomings, we proposed the Autonomous Robotic Colonoscopy System (ARCS), as illustrated in Figure 1. The ARCS has a 3 degrees-of-freedom-based robotic colonoscopy system (3)-RCS), allowing the colonoscope to steer up-down/right-left and advance forward/backward. The ARCS operates autonomously within a Colonoscopy Training Simulator (CTS) using a control algorithm and a deep learning model. A pair of joysticks is utilized to collect data for training the model. Furthermore, we proposed several deep learning-based approaches to enhance the performance of the autonomous robotic colonoscopy system in terms of intubation time and intubation rate in CTS.

A summary of our contributions contained within this paper is as follows:

- We designed the 3-RCS to perform up-down, right-left, insertion motions for colonoscopy procedures, while maintaining the range of movement of the conventional FC. Additionally, the 3-RCS can be adjusted to accommodate different FC diameters. To demonstrate the reliability of the system, we conducted a hysteresis experiment.
- We also conducted in-vitro experiments to validate the feasibility of the autonomous 3-RCS by combining the IBVS algorithm with a deep learning model.
- We recorded and compared the intubation time of human operators using joysticks with the ARCS using the IBVS algorithm.
- Finally, we suggested a deep learning model to produce all control signals simultaneously, leading to reduced computational time, lower CIT, and increased CIR.

II. ROBOT SYSTEM AND IMAGE-BASED VISUAL SERVO

The overall system is illustrated in Figure 2. The 3-RCS comprises a feeding mechanism (FM) for insertion/retraction motion, and a steering mechanism (SM) for 2-directional steering (up/down and right/left) motions.

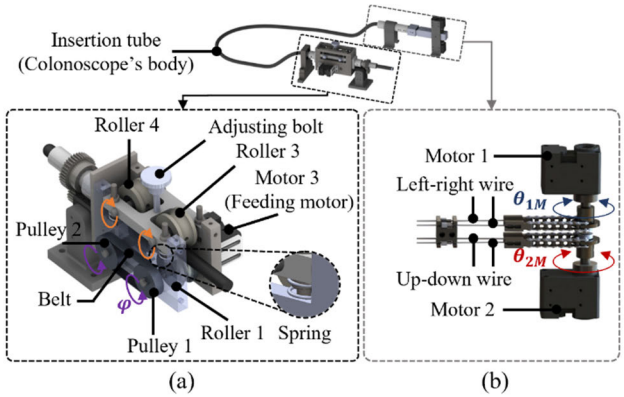

FIGURE 2. The detailed design of the robotic colonoscopy system (3)-RCS). (a) Feeding mechanism. (b) Steering mechanism.

A. THE FEEDING MECHANISM

The FM, as shown in Figure 2(a), consists of motor 3, driving rollers (1, 2), pulleys (1, 2), and driven rollers (3, 4). When motor 3 drives roller 1, since pulley 1 is connected to pulley 2 by a gear belt, rollers 1 and 2 can rotate (ϕ) at the same direction and speed. The distance between the driving rollers and driven rollers can be adjusted with a bolt. Thus, the 3-RCS not only allows the operator to modify the normal force generated between the rollers and the colonoscope's body, but also makes the system compatible with different conventional FC diameters. With 4 rollers, the feeding mechanism is considered flat, ensuring the smooth and stable movement of the colonoscope's body. A compression spring is used to adjust the contact damping between the rollers and the colonoscope. At a fixed feeding velocity (20mm/s [10]), and feeding motor torque of $T_{FeedingMotor} = 0.11$ N.m, manual adjustment of the compression spring and bolt limits the insertion force to less than 15 N, preventing damage to the colon wall [24]. It is worth noting that all motors are dynamixel motors (XM430-W350-T) which can be controlled in torque/position/speed modes with high accuracy and precision. This precise modulation allows the algorithms to easily limit force/torque values, ensuring system safety.

B. STEERING MECHANISM

As shown in Figure 2(b), the SM can steer the colonoscope's head in both the left/right (L/R) and up/down (U/D) directions

by using two motors: motor 1 and motor 2. The SM is designed to ensure that the FC (Olympus, Japan) can reach its maximum bending angle.

C. EVALUATION OF FEEDING AND STEERING MECHANISM

1) EXPERIMENTAL SETUP

Although conventional FCs have been widely used and are considered reliable, it is essential to evaluate the left/right bending angle (ψ_{LR}), up/down bending angle (ψ_{UD}), and insertion motion displacement (F). This evaluation is necessary because a reliable robotic system is a fundamental prerequisite for building a successful autonomous system using advanced deep learning algorithms.

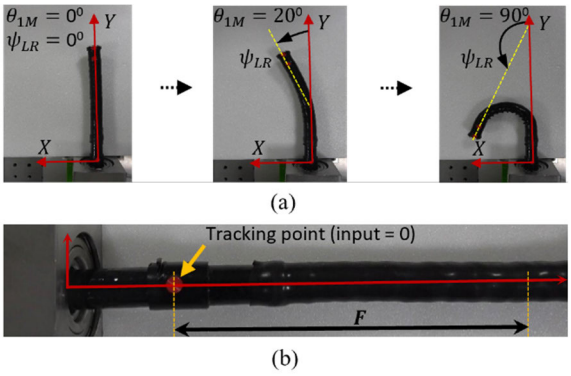

FIGURE 3. Experimental setup for validating the steering and insertion motion. (a) Experimental configuration for the steering motion. (b) Experimental configuration for the insertion motion.

Particularly, hysteresis experiments that measured the synchronized response of the steering/insertion motion with respect to motor control inputs were performed. As shown in Figure 2(b) and Figure 3(a), θ_{1M} and θ_{2M} present the motor positions used to measure the bending angle ψ_{LR} and ψ_{UD}, respectively. Motors 1 and 2 are controlled to rotate from 0^0 to 90^0 to obtain expected values of (ψ_{LR}, ψ_{UD}), which range from 0^0 to 160^0. At the same time, a change of 0^0 to -90^0 in motor input (θ_{1M}, θ_{2M}) results in a bending angle of (ψ_{LR}, ψ_{UD}), which ranges from 0^0 to -160^0.

For the insertion motion, an input of 100 mm is used to create repeated forward/backward movements of the insertion tube, and then the displacement F of the tracking point on the colonoscope's body is measured, as shown in Figure 3(b).

2) RESULTS AND DISCUSSION OF THE ROBOT SYSTEM

Figure 4(a) illustrates the hysteresis profiles of both the L/R and U/D steering angles which are relatively symmetric for the loading and unloading phases. Like the previous study [10], the hysteresis profiles are due to inherent cable slip, component deformation, friction, or nonuniform movement within the tendon-driven mechanism. However, the proposed mechanism shows a smaller backlash (around 8^0 at the maximum steering angle) compared to Huang's mechanism [10] (around 20^0) or Ott's [25] (around 10^0). This is mainly because the head of the colonoscope in the proposed mechanism has a lighter weight camera (XD-V31105LH-158). Huang's mechanism contains other surgical instruments, which can make the mechanism more susceptible to tendon twists and other tension variations. Additionally, due to the difference in angulation range between L/R and U/D motion in the conventional FC, there is a slight difference in the shape of ψ_{LR} and ψ_{UD}.

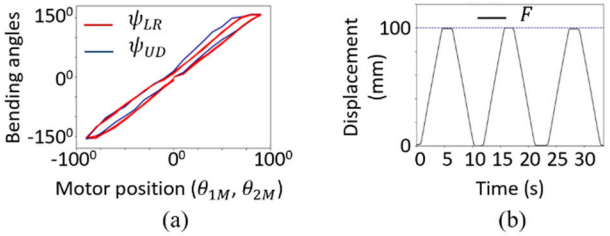

FIGURE 4. The hysteresis profiles. (a) L/R and U/D motions. (b) Insertion motion.

Figure 4(b) shows the measured displacement of the tracking point during the insertion motion. The measured displacement of the insertion motion ($F = 99.3 \pm 0.6$ mm) indicates that the feed mechanism enables the colonoscope to move forward/backward accurately since two pairs of rollers are employed. This contrasts with a previous study [26] where slip occurred because only one pair of rollers was used, which made the feeding velocity inconsistent or less predictable.

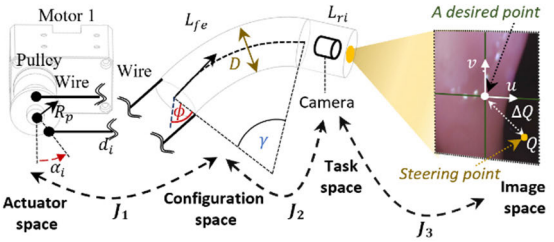

FIGURE 5. Diagrammatic representation of IBVS for 3-RCS.

D. IMAGE-BASED VISUAL VERVO CONTROL SCHEMATIC

Recently, the Image-Based Visual Servo (IBVS) [27] control algorithm, which relies on visual feedback from the colonoscope's camera, has become a favorable method for tracking and guidance in autonomous systems. In our system, the IBVS algorithm is applied to produce steering signals that control the colonoscope's head. Figure 5 displays the correlation between the rotation angle of the pulleys (actuator space), backbone (configuration space), and the bending angle of the colonoscope's head (task space). Particularly, a mapping from the actuator space ($\mathrm{P} = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}^T$) to configuration space ($C = \begin{bmatrix} \emptyset & \gamma \end{bmatrix}^T$) can be described by the transformation matrix J_1, as shown in (1). The transformation matrix J_2 is used to map the configuration space to the task space ($\mathrm{T} = \begin{bmatrix} v_T & \omega_T \end{bmatrix}^T$), as shown in (2).

$$\dot{C} = J_1 \dot{P} \tag{1}$$

$$T = J_2 \dot{C} \tag{2}$$

where the matrix J_1, as shown in (3), is the inverse kinematic of continum robotic colonoscope and actuators. J_2 describes the relationship between the configuration space and end-efector of robot [25], as shown in (4). R_p, D, and d_i represent the radius of the pulleys, the diameter of the colonoscopy's body (cross-section), and length distribution of wires, while α_1 and α_2 are the rotation angle of the pulleys 1 and 2, respectively. The angle values of $\emptyset$ and γ are caculated based on (6) and (7).

$$\mathbf{J}_1 = \begin{pmatrix} -\frac{\alpha_2}{\alpha_1^2+\alpha_2^2} & \frac{\alpha_1}{\alpha_1^2+\alpha_2^2} \\ \frac{2R_p}{D}\frac{\alpha_1}{\sqrt{\alpha_1^2+\alpha_2^2}} & \frac{2R_p}{D}\frac{\alpha_2}{\sqrt{\alpha_1^2+\alpha_2^2}} \end{pmatrix} \tag{3}$$

$$\mathbf{J}_2 = \begin{pmatrix} -\frac{L_{fe}}{\psi}s\emptyset\,(1-c\psi) - L_{ri}s\emptyset s\psi & \frac{L_{fe}}{\beta}c\emptyset\,(1-c\psi) + L_{ri}c\emptyset \\ \frac{L_{fe}}{\psi}c\emptyset\,(1-c\psi) + L_{ri}c\emptyset s\psi & \frac{L_{fe}}{\beta}s\emptyset\,(1-c\psi) + L_{ri}s\emptyset \\ 0 & \frac{L_{fe}}{\beta}\left(1-\frac{s\psi}{\psi}\right) \\ -c\emptyset s\psi & -s\emptyset \\ -s\emptyset s\psi & c\emptyset \\ -1+c\psi & 0 \end{pmatrix} \tag{4}$$

$$d_i = R_p\alpha_i, \quad i = 1, 2 \tag{5}$$

$$\emptyset = \begin{cases} 0 & if\ \alpha_1 = \alpha_2 = 0 \\ \text{atan2}(-\alpha_2, -\alpha_1) & \text{Otherwise} \end{cases} \tag{6}$$

$$\gamma = \frac{2R_p}{D}\sqrt{\alpha_1^2+\alpha_2^2} \tag{7}$$

Additionally, by employing IBVS control theory [27], the mapping between the task space and the image space can be obtained using (8), which involves the Image Jacobian matrix J_3, as described in (9).

$$\dot{Q} = J_3 T \tag{8}$$

$$J_3 = \begin{bmatrix} -\frac{f}{z} & 0 & \frac{u}{z} & \frac{uv}{f} & \frac{-f^2-u^2}{f} & v \\ 0 & -\frac{f}{z} & \frac{v}{z} & \frac{f^2+v^2}{f} & -\frac{uv}{f} & -u \end{bmatrix} \tag{9}$$

Next, we get the relationship between the motor's velocity and image feature by combining (1), (2) and (8), as shown in (10). Finally, the steering signals for the ARCS in method A can be calculated by using (11) if the steering point (Q) is given. The symbols and its desription/values of (11) are summarised in Table 1.

$$\dot{Q}^{2\times 2} = J_3^{2\times 6} J_2^{6\times 2} J_1^{2\times 2} \dot{P}^{2\times 2} \tag{10}$$

$$\begin{bmatrix} \dot{\theta}_1 & \dot{\theta}_2 \end{bmatrix}^T = \varepsilon J_1^{-1} J_2^{-1} J_3^{-1} \Delta Q \tag{11}$$

E. FEEDING SPEED, CONTROL ALGORITHM

Apart from the steering signals ($\omega = \dot{P}$), which are calculated based on the IBVS above for controlling the colonoscope's head in left/right/up/down directions, control of the insertion speed of the colonoscope's body is also required for fully autonomous operation. In this paper, we used one deep learning model (Model 1) to create a steering point Q and collision probability (col) by calculating input from a colonoscopic image. The collision probability classifies the colonoscopic image as "wall" or "non-wall" during the operation, ensuring precautions are taken to avoid perforation. The structure of model 1 will be presented in the description of our deep learning approach below.

TABLE 1. The symbols and its description of variables.

Description	Values
$R_p = 8.5$ mm	Radius of pulleys
$D = 12.85$ mm	Cross-section of the insertion tube
$L_{ri} = 10$ mm	The rigid section (distal end)
$L_{fe} = 85$ mm	Flexible section (bending part)
$\dot{P} = [\dot{\theta}_1 \quad \dot{\theta}_2]^T$	Velocity of pulley 1 and 2
$c\psi$, $s\emptyset$	cosψ, sin$\emptyset$
$Q = [Q_x \quad Q_y]^T$	Position of a steering point on an image
ΔQ	Error of pixel between steering point and the center of the lumen
$f = 485$	Focal length [28]
$z = 8$	Depth value [28]
$\varepsilon = 50$	Feedback gain

1) FEDDING SPEED

For safety, the feeding speed (V_{fe}) or insertion speed of the colonoscope's body is computed based on (12), where the collision probability (col) is given from Model 1, and the maximum velocity V_m is chosen at 20 mm/s [10]. The idea is that if the col is between 0 to 0.6, it means that the robot clearly knows the path ahead, and therefore the V_{fe} is set close to the V_m. Otherwise, V_{fe} is set to zero to avoid hitting the colon, potentially causing colon perforations.

$$V_{fe} = \begin{cases} (1-col)V_m & 0 \le col \le 0.6 \\ 0 & \text{Otherwise} \end{cases} \tag{12}$$

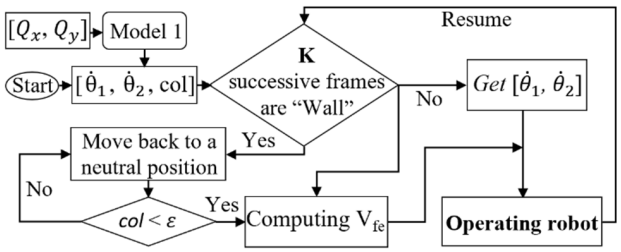

FIGURE 6. Schematic of control algorithm.

2) CONTROL ALGORITHM

When the system has steering signals (ω), collision probability (col), and insertion velocity (V_{fe}), we can control the ARCS with *a proposed control algorithm* (CA), as shown in Figure 6. We found that when the col is bigger than 0.6, V_{fe} is set to zero and the colonoscope does not advance anymore. To

enable the colonoscope to continue its autonomous operation, *a control algorithm* is employed. The working principle of CA is as follows: The system starts, and CA always obtains *col* values provided by model 1 (Hz =30). If there are K=80 successive frames classified as "wall" (a frame with $col >$ 0.6 is classified as a "wall"), that situation will be defined as "collision". In the collision state, the colonoscope must move back until the *col*becomes lower than $\varepsilon = 0.02$. After *col* is less than ε, the system can resume. By doing so, the colonoscope never hits the colon wall forcefully, ensuring safety for patients.

III. PROPOSED DEEP LEARNING APPROACHES

In this section, we will examine three deep learning-based approaches: Method A, Method B, and Method C. Each of these methods includes a set of deep learning models, specifically Model 1, Model 2, and Model End-to-End. We will describe how these models and methods are connected to each other and how we designed the models and trained them. Additionally, we will provide a description of the data generation process used to train these models and clarify our methodology for model selection.

A. OVERVIEW OF METHODS

Method A involves using a convolutional neural network (CNN) model, referred to as Model 1, to process the colonoscopic images captured by a camera on the colonoscope's head (Figure 7(a)). Model 1 generates a steering point, $Q = (Q_x, Q_y)$, and a collision probability (*col*). The IBVS algorithm utilizes the steering point along with the actual angles of the pulleys to generate steering signals (ω), allowing the colonoscope to move in the up/down or left/right directions. As mentioned above, the collision probability is used to determine the colonoscope's body speed (V_{fe} – the feeding speed). Both (ω) and V_{fe} are inserted in CA to enable the autonomous manipulation of robot system.

In Method A, the utilization of the IBVS algorithm requires feedback signals, such as the actual angles of the pulleys, to generate steering signals (ω). However, this approach using feedback signals can potentially lead to issues with accuracy and stability. To address these limitations, we introduce Method B, which involves the use of Model 2, as depicted in Figure 7(b). Model 2 is designed to predict the steering signals based on the steering points provided by Model 1, eliminating the need for real-time pulley angle calculations required in the IBVS algorithm. This helps enhance the accuracy and stability of the system.

Finally, we introduced end-to-end, learning-based control signal prediction algorithm, enabling the 3-RCS to function autonomously with just one model, which is referred to as model E2E (Figure 7(c)). Model E2E has the capability to produce all the necessary control signals for autonomous 3-RCS operation, including steering signals, collision probability, and feeding speed. This approach significantly reduces the computational load and operational time while maintaining high performance, making it a promising alternative for the ARCS. This system is denoted as 'Method C.'

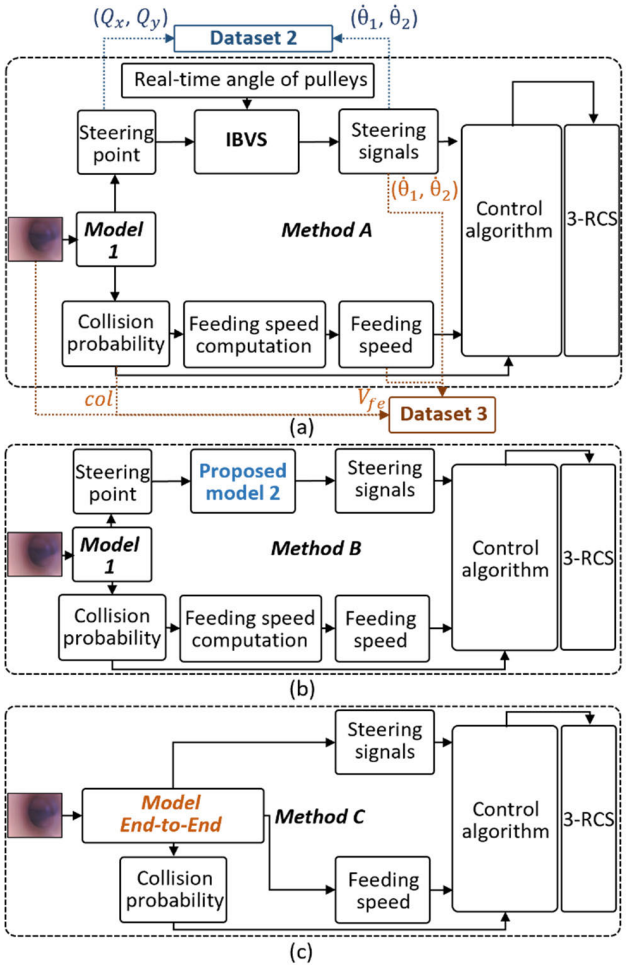

FIGURE 7. This schematic diagram illustrates three methods and provides a description of the data generation process for Dataset 2 and Dataset 3. (a) Method A - Model 1 and IBVS. (b) Method B - Model 1 and Model 2. (c) Method C - Model End-to-End (E2E).

To assess the performance of Methods A, B, and C, we conducted 50 experiments for each of the three methods, measuring both the CIT and CIR. Furthermore, we performed the CIT/CIR measurement while a human operator manipulated the colonoscope using a joystick, and these results will be compared to the three aforementioned methods.

B. DATASET GENERATION

Each method includes deep learning models that need to be trained with data. In this section we discuss how to create these datasets. We manually labeled Dataset 1, as described in the following paragraph, and subsequently created Dataset 2 and 3 based on Method A as shown in Figure 7(a) and 8.

The datasets of three models were collected as follows. **For dataset 1,** which was trained for model 1 of method A, we manually manipulated the 3-RCS using a pair of joysticks on CTS to move the colonoscope from the rectum to the

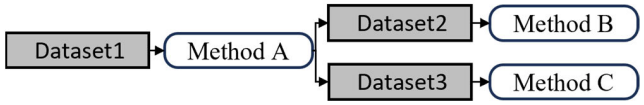

FIGURE 8. Data generation of dataset 2 and dataset 3.

cecum, as shown in Figure 1. We then collected 27 experiments consisting of a total of 52771 images. While it is possible to obtain the steering point by detecting the centroid of the darkest pixels in the images, the accuracy of determining the steering point is significantly affected by the application of lubricating oil during colonoscopy. Therefore, we chose to assign the steering point for each colonoscopic image in this study. Each image was annotated with a steering point (represented by a white dot) and assigned a class (either class 0 or class 1) based on the opinion of a colonoscopist, as shown in Figure 9. Class 1 represents a "wall" that indicates the colonoscope should not proceed further to prevent intestinal perforation, while class 0 represents "not a wall", meaning that the colonoscope can continue to move forward or steer.

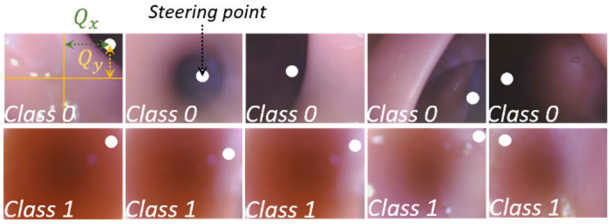

FIGURE 9. Annotated images of dataset 1.

Figure 7(a) shows the process of generating dataset 2 and dataset 3, which are used in Method B and Method C, respectively. After completing the training and selecting the best setup for model 1, we conducted 12 autonomous experiments using method A on the CTS. As a result, we generated **dataset 2**, which was used to train the proposed model 2 of Method B. This dataset was created by recording the inputs (steering points: Q_x, Q_y) and outputs (steering signals (ω): $\dot{\theta}_1$, $\dot{\theta}_2$) of the IBVS (blue round dot line).

To create **dataset 3**, which was trained for model E2E of method C, we followed a similar approach to creating dataset 2. We conducted 30 autonomous experiments using method A and saved the inputs (colonoscopic images) and outputs (steering signals, collision probability (*col*) and feeding speed (V_{fe})), as shown in Figure 7(a) (orange round dot line).

To speed up convergence, all collected images, steering points (Q_x, Q_y), and control signals ($\dot{\theta}_1$, $\dot{\theta}_2$, *col*, V_{fe}) were normalized in the range of [0:1]. Additionally, we divided all datasets into a training set (80%) and a validation set (20%) during model training.

C. MODEL ARCHITECTURES

This section presents the architectures of Model 1, Model 2, and Model E2E, which are trained on Dataset 1, Dataset 2, and Dataset 3, respectively, and implemented in Method A, Method B, and Method C.

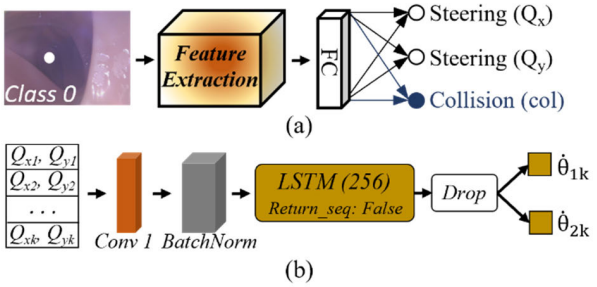

FIGURE 10. The architecture of models. (a) Model 1. (b) Model 2.

Model 1: Four pre-trained models (ResNet-18 [29], ResNet-34, MobileNet-v2 [30], DenseNet-121 [31]) are trained to determine the best version of model 1. An input image with size of 224 × 224 × 3 is passed to the feature extraction layer, followed by a flattened layer, a dropout layer (0.5), and a dense (fully-connected) layer with 128 units. The output tensors are fed to ReLU activation and to the dense layer with 3 units. Two units are responsible for a regression task of finding Q_x and Q_y, while the third unit is used for a classification task of finding "*wall*" or "*non-wall*" image, as shown in Figure 10(a).

The combination of a mean square error (MSE) loss (L_M) and binary cross-entropy (BCE) loss (L_B) was used to calculate a total loss (L_T) during training to reduce complexity and processing time. A constant loss weight (k) was also used to optimize the training performance [32]. The total loss is calculated by (13), where n represents the number of samples, and (Q_x^r, Q_y^r, col^r) and (Q_x^p, Q_y^p, col^p) refer to the real and predicted values of steering point and collision probability, respectively.

$$L_T = L_M + kL_B \tag{13}$$

$$L_M = \frac{1}{n}\sum_{i=0}^{n}\left((Qx_i^r - Qx_i^p)^2 + (Qy_i^r - Qy_i^p)^2\right) \tag{14}$$

$$L_B = -\frac{1}{n}\sum_{i=0}^{n}\left(col_i^r \log(col_i^p) + (1 - col_i^r)\log(1 - col_i^p)\right) \tag{15}$$

Model 2: This article proposed a One-Dimensional Convolutional Neural Network - Long Short-Term Memory (1DCNN-LSTM) model (Model 2) to handle the one-dimensional input and output data in the autonomous colonoscopy procedure. The proposed model was trained and compared to other architectures that have shown good performance in other tasks such as 1D-CNN [33], 1DCNN-EncoderDecoder [34], LSTM [35], DCNN-LSTM [36], Bidirectional-LSTM (Bi-LSTM) [37], and

1DCNN-BiLSTM [38]. To adapt these models to the problem at hand, all the feature extraction layers are retained, and a dense layer with two units is added at the end. The proposed Model 2 uses a 1D-convolutional layer with 16 filters, a kernel size of 7, a stride of 1, and a LeakyReLU activation function (0.1). This is followed by batch normalization, an LSTM layer with 256 units and a Tanh activation function, a Dropout layer (0.5), and a Dense layer with 2 units and a linear activation function. The overall architecture is shown in Figure 10(b).

Model E2E: The proposed model E2E follows a similar approach to model 1, using the pre-trained backbones of ResNet-18, ResNet-34, MobileNet-v2, and DenseNet-121 to generate a feature map. After the flattened data is extracted from the feature extraction layer, it passes through a dropout layer (0.5) and a dense layer (256). The output of this layer then passes through another dropout layer (0.5) before being fed to the final dense layer with 4 units. These units are responsible for predicting the values of $\dot{\theta}_1$, $\dot{\theta}_2$, *col*, and V_{fe}.

D. HYPERPARAMETER AND TRAINING

To determine the optimal model for each of the three models (1, 2, and E2E), various learning rates and batch sizes were used. Both model 1 and model E2E are similar, so they were trained using the same set of parameters: learning rates [10^{-4},10^{-5}], batch sizes [8,16,32,64], and 30 epochs. Model 2 was trained on the same learning rates, but with a different set of batch sizes (64,128,256), and epochs (100). A time step (k=40) was chosen for model 2. The cost functions for regression tasks (model 2 and E2E) were MSE, while model 1 used (15) above. Additionally, the Adam optimizer was set for all models during training. To increase dataset size, data augmentation techniques such as rotation, scale, and flipping were employed for input images.

IV. EVALUATION AND MODEL SELECTION

To quantitatively evaluate the models, the primary factor used to determine the best model is the loss value. Specifically, the model with the smallest loss value on the validation set is considered to be the best-trained model.

Model 1, the Resnet34-based model, achieved the smallest loss value ($L_T = 0.028$) when trained on the batch size of 64 and learning rate of 10^{-5}. The classification result of the best model 1 is shown in Figure 11(a), where F1-score and accuracy on the validation set are 0.9927 and 0.9866, respectively. According to Figure 11(b), when the colonoscopic image is classified as "*non-wall*", model 1 provides the good predicted (green) steering points that match the ground-truth values (white). On the other hand, when colonoscopic images are predicted as "wall", or the collision probability is higher than 0.6, we cannot determine the direction for the next advancement.

Next, the proposed **Model 2** has the same lowest loss value as the model Bi-LSTM (MSE = 0.0144). However, the average processing time (0.02732 s/image) of the proposed model B is lower than that of model Bi-LSTM (0.02814 s/image). Therefore, the best model B is selected as the proposed model 2. Figure 12(a) shows that predicted steering signals closely follow the ground truth produced by the best model 2 (without using the angle of the pulley's data).

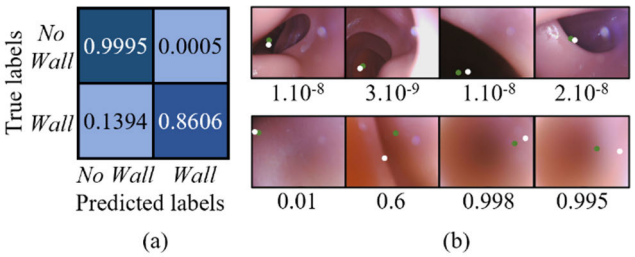

FIGURE 11. The prediction result of the best model 1. (a) Confusion matrix. (b) Predicted steering points and collision probability.

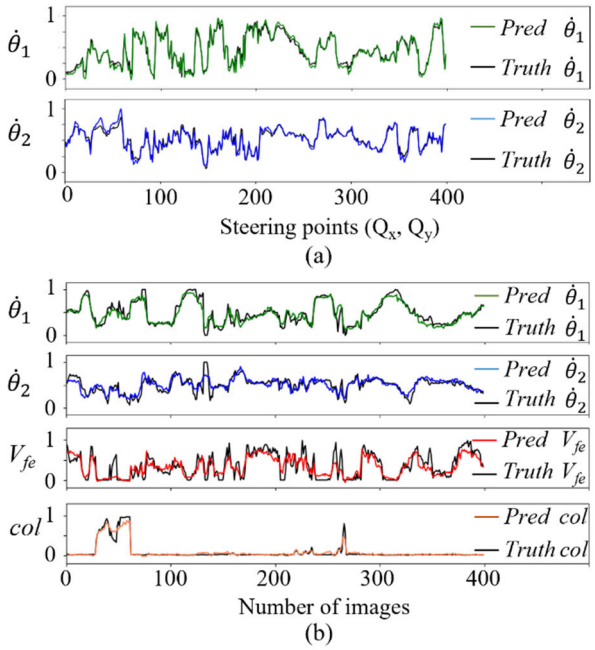

FIGURE 12. The prediction results. (a) Model 2. (b) Model E2E.

The best **Model E2E**, having the smallest loss value (MSE = 0.051), is the MobileV2-based model, trained on a batch size of 8 and learning rate of 10^{-4}. We employed the best model E2E on the validation set and see that the model E2E can produce control signals well, as shown in Figure 12(b). Obviously, the steering velocities generated by model 2 are better than those created by model E2E since model E2E must share weights to learn the behaviors of col and V_{fe}. However, using a single model reduces the complexity and computation time of the system, as compared to using 2 models (model 1 and model 2).

V. EXPERIMENTAL SETUP, RESULT AND DISCUSSION

The experiment setup, evaluation criteria, results, and discussion of human operator and deep learning-based autonomous manipulation of robotic systems are provided in this section.

A. EXPERIMENT SETUP

Fig. 1 shows an autonomous robotic colonoscopy, where the colonoscope autonomously advances from the rectum to the cecum. The experiments were conducted 50 times each, in the sequence of methods A, B, C, and back to A, for a total of 150 experiments. The experiment is considered a success if the CIT is less than 20 minutes [39], and there is no elongation of the colon. The elongation of the colon, or an '*elongation*' state, is determined when the input image is classified as 'non-wall' and the colonoscope keeps moving forward while the head of the colonoscope remains stuck in a specific position for over 5 seconds.

Figure 13 shows images of elongation cases during the experiments. The yellow dot represents the position where the colonoscope's head is stuck, and the dashed oval indicates the elongated colon position. We found that elongation of the colon can cause patient discomfort or even perforation, and therefore needs to be classified as a failure case.

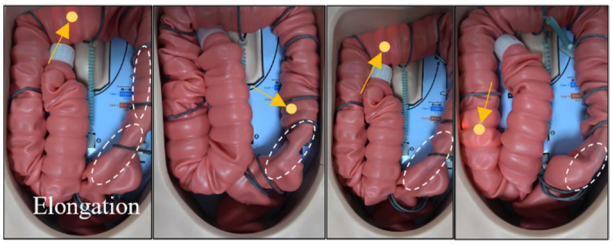

FIGURE 13. The experiment failed due to elongation.

Alongside the autonomous robotic colonoscopy experiments mentioned above, we conducted an experiment to compare the intubation time of ARCS and a human operator. The experiment involved 5 participants (aged from 25 to 32 years) who had no prior experience in manipulating a colonoscope [40]. The participants were tasked with moving the colonoscope from the rectum to the cecum 10 times using a pair of joysticks, and their intubation times/rate were recorded. Prior to the experiment, participants were given 5 minutes [7] of practice time with the assistance and advice of experts during the practice.

B. RESULTS AND DISCUSSION

The purpose of these experiments was to compare the CIT and CIR of different methods. Overall, the average CIT for the three methods and human operator was approximately 2min 46 seconds (s), and the CIR was above 92%, as shown in Table 2. There were some failed experiments due to elongated colons, especially in the transverse and ascending sections where the colon has more acute bends. As the colonoscope goes deeper into the colon, controlling it becomes more difficult due to cable slip, tension loss, and friction forces of the colonoscope. Additionally, the presence of tightly curved sections, folds, and visibility problems due to lubrication caused directional errors, resulting in potential kinematic errors and inaccurate control signals. As expected, these failed cases occurred in the same experiments for all methods, indicating that the colon's configuration also plays an important role in the success/failure of the experiments.

TABLE 2. Experimental results of Participants, methods A, B, and C.

Method	Section	Intubation time M± SD(Max, min)	Success rate (Success/trial)
Human operators	Rectosigmoid	83 ± 38 (220, 15)	100% (50/50)
	Descending	35 ± 15 (70, 12)	100% (50/50)
	Transverse	24 ± 10 (50, 13)	100% (50/50)
	Ascending	25 ± 10 (60, 13)	100% (50/50)
	Full colon	**166 ± 53 (300, 63)**	**100% (50/50)**
Method A	Rectosigmoid	76 ± 66 (252, 14)	100% (50/50)
	Descending	27 ± 27 (135,15)	100% (50/50)
	Transverse	38 ± 26 (100, 15)	94% (47/50)
	Ascending	27 ± 10 (60, 18)	92% (46/50)
	Full colon	**165 ± 73 (331, 70)**	**92% (46/50)**
Method B	Rectosigmoid	74 ± 46 (220, 22)	100% (50/50)
	Descending	27 ± 11 (75, 15)	100% (50/50)
	Transverse	34 ± 27 (176, 16)	100% (50/50)
	Ascending	29 ± 11 (67, 18)	94% (47/50)
	Full colon	**164 ± 54 (316, 88)**	**94% (47/50)**
Method C	Rectosigmoid	50 ± 40 (160, 15)	100% (50/50)
	Descending	21 ± 8 (50, 15)	100% (50/50)
	Transverse	29 ± 10 (57, 15)	98% (49/50)
	Ascending	26 ± 14 (105, 18)	96% (48/50)
	Full colon	**127 ± 46 (245, 75)**	**96% (48/50)**

We found that the rectosigmoid region poses a significant challenge during colonoscopy, despite its relatively shorter length compared to other parts of the colon. Participants find it more difficult to advance through the rectosigmoid, which has higher mean (M)± standard deviation (SD) of 83 ± 38 seconds (s), than other sections, such as the descending (35 ± 15 s), transverse (24 ± 10 s), and ascending (25 ± 10 s) segments. The reason for this difficulty is the presence of numerous folds and sharp-angled turns in the rectosigmoid section, whereas straight paths are mostly observed in the other sections.

Like human operators, the ARCS takes longer to traverse through this rectosigmoid region, with an average time of 76 s, than the descending and ascending parts which take around 27 s, and the transverse part, which takes around 38s. This finding highlights the technical difficulties associated with navigating through the rectosigmoid, despite its smaller size, and underscores the importance of developing effective control strategies to ensure accurate and efficient colonoscopy procedures.

Comparing the results of human operators with Method A, it is evident that the CIR of participants (100%) is higher than the IR of Method A (92%). This is because human

operators are aware of and can effectively handle situations where elongated colons occur by moving the colonoscope back and using the colonoscope's head to expand the colon. However, it is challenging for the ARCS to recognize elongation scenarios. Specifically, when the colonoscope's head gets trapped in a corner of the colon or folded part of the colon, the provided colonoscopic image occasionally shows a ''non-wall'' classification. In such situations, the system continues to feed the colonoscope forward since the feeding velocity is determined by the collision probability derived from the given image. This leads to the occurrence of elongation phenomena.

For the entire colon, the CIT of human operators (166 $\pm$ 53 seconds) is comparable to that of Method A (165 $\pm$ 73 seconds). However, in the case of the rectosigmoid, the CIT in human operators (83 $\pm$ 38 s) is slightly higher than in Method A (76 $\pm$ 66 s). This disparity arises because participants, who are non-expert colonoscopists, initially lack awareness of how to navigate when encountering folded colons and sharp-angled turns.

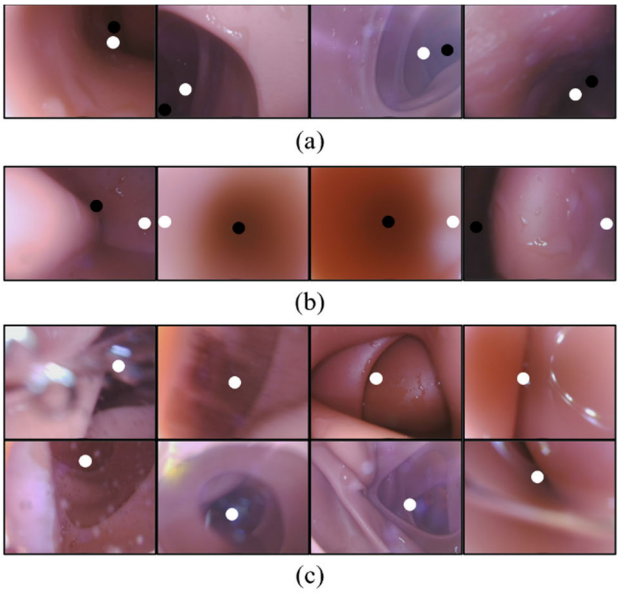

FIGURE 14. The colonoscopic images during the autonomous operation.

In method A, the robot can see various kinds of images, as shown in Fig 14. Several previous studies have achieved good results by using the centroid of the darkest pixels in the images as a steering point [13], [16]. However, this approach can be problematic for colonoscopy. In Figure 14(a), we observe that the darkest points (represented by black dots) are located near the human-labeled points (represented by white dots). However, in Figure 14(b), the darkest points are located far away from the human-labeled points. Therefore, using the darkest point can cause the colonoscope to get stuck in a corner or go the wrong direction, potentially leading to an increase in intubation and elongation. Similarly, threshold segmentation algorithms and edge-based methods [18], [19], [20] were sometimes unreliable because their accuracy is affected by image variations such as noise, uneven illumination, or varying contrast, as can be seen in Figure 14(c). Conversely, Method A, which incorporates qualitative annotations based on the endoscopist's experience, offers more stability and accuracy, making colonoscopy more reliable and safer.

Next, we observed that the CITs of methods A and B are similar for each section. However, for a complete colon examination, method B exhibited slightly better performance (CIT = 164$\pm$54 s and CIR = 94%), compared to method A (CIT = 165$\pm$73 s, CIR = 92%). The lower CIT and higher CIR of method B can be attributed to the absence of feedback signals, which are typically required for IBVS. In other words, by not relying on these feedback signals, method B experiences reduced impact from feedback signal errors, thereby resulting in improved overall performance.

The results of the experiment demonstrated that method C outperforms method B in terms of CIT and CIR for all sections of the colon, including the full colon. Specifically, the mean CIT of method C (127s) is 23% less than that of method B (164s) for the full colon. This improvement in speed (7.1 mm/s) is noteworthy and comparable to previous studies [11] [13], [22], [23], and is accompanied by a better CIR [11], [13]. Furthermore, method C shows greater stability than method B, as evidenced by a smaller max/min value (245/75s) and lower SD (46s) compared to method B which has max/min value of 316/88s and SD of 54s.

The superior performance of method C can be attributed to its use of model E2E, which is trained on 30 successful experiments of autonomous robotic colonoscopy. By leveraging the knowledge gained from previous successful experiments, method C is better equipped to complete new experiments successfully. This is reflected in its higher CIR compared to both methods B and A. Overall, the results of this experiment indicated that method C is a more effective approach to autonomous robotic colonoscopy than methods A, B and human operators, with improved CIT, CIR, and stability.

VI. CONCLUSION

In this paper, we present a reliable 3-RCS designed to provide steering and feeding motions for colonoscopy procedures. We conducted hysteresis experiments, and the results demonstrate that our steering and feeding mechanism enables the dependable operation of the colonoscope, making it suitable for autonomous colonoscopy. To achieve successful autonomous operation within the Colonoscopy Training Simulator, we employed various deep learning-based approaches. For Method A, Model 1 was utilized to detect steering points and collision probabilities using a CNN model accurately and comprehensively. This enhanced accuracy helps prevent interaction forces with the colon wall, avoiding perforations. Additionally, instead of using IBVS to control the autonomously operated colonoscope, we proposed Model 2 which is based on 1DCNN-LSTM model, which generates steering signals and achieves comparable CIT and CIR compared to the IBVS method. Finally, we employed an

End-to-End (E2E) model, which not only reduces computational processing time but also enhances CIT, CIR, and stability for autonomous operation. The results of the proposed method also outperformed novice operators. It is expected that the proposed method can support colonoscopists in reducing workload, physical demands, work-related pain, and injuries related to colonoscopy procedures.

In future work, it is crucial to develop an improved strategy or control algorithm to minimize the intubation time, specifically in the rectosigmoid section. This is important not only because it poses challenges for human operators but also for the proposed autonomous system. Furthermore, it is recommended to generate additional datasets, as we have determined that the elongation issue stems from an insufficient number of images used during training. Consequently, the robot struggles to accurately classify various situations. To enhance the network's accuracy, it would be beneficial to explore the implementation of advanced deep learning models like vision transformers. From a mechanical development standpoint, our future plans involve attaching an accelerometer-gyroscope sensor to the head of the colonoscope to monitor the occurrence of any elongation. Furthermore, it is crucial to implement a searching algorithm and incorporate additional mechanisms like an expansion mechanism to effectively overcome the elongation issue.

ACKNOWLEDGMENT

The authors would like to thank students at Korea Aerospace University and Space Mechanisms and Robotics Laboratory for participating as human operators in the colonoscopy experiments.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clinicians*, vol. 68, no. 6, pp. 394–424, Nov. 2018.

[2] H. Khil, S. M. Kim, S. Hong, H. M. Gil, E. Cheon, D. H. Lee, Y. A. Kim, and N. Keum, "Time trends of colorectal cancer incidence and associated lifestyle factors in South Korea," *Sci. Rep.*, vol. 11, no. 1, pp. 2413–2425, Jan. 2021.

[3] R. Siegel, "Cancer treatment and survivorship statistics, 2012," *CA, Cancer J. Clinicians*, vol. 62, no. 4, pp. 220–241, 2012.

[4] E. Ofori, "Occupation-associated health hazards for the gastroenterologist/endoscopist," *Ann. Gastroenterol.*, vol. 31, no. 4, p. 448, 2018.

[5] S.-H. Lee, "Colonoscopy procedural skills and training for new beginners," *World J. Gastroenterol.*, vol. 20, no. 45, 2014, Art. no. 16984.

[6] H.-S. Yoon and B.-J. Yi, "Semi-automatic knob system for assisting flexible endoscope steering," *Int. J. Control, Autom. Syst.*, vol. 18, no. 10, pp. 2650–2657, Oct. 2020.

[7] J. Ruiter, E. Rozeboom, M. van der Voort, M. Bonnema, and I. Broeders, "Design and evaluation of robotic steering of a flexible endoscope," in *Proc. 4th IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechatronics (BioRob)*, Jun. 2012, pp. 761–767.

[8] E. D. Rozeboom, J. G. Ruiter, M. Franken, M. P. Schwartz, S. Stramigioli, and I. A. M. J. Broeders, "Single-handed controller reduces the workload of flexible endoscopy," *J. Robotic Surg.*, vol. 8, no. 4, pp. 319–324, Dec. 2014.

[9] T. Iwasa, R. Nakadate, S. Onogi, Y. Okamoto, J. Arata, S. Oguri, H. Ogino, E. Ihara, K. Ohuchida, T. Akahoshi, T. Ikeda, Y. Ogawa, and M. Hashizume, "A new robotic-assisted flexible endoscope with single-hand control: Endoscopic submucosal dissection in the ex vivo porcine stomach," *Surgical Endoscopy*, vol. 32, no. 7, pp. 3386–3392, Jul. 2018.

[10] Y. Huang, W. Lai, L. Cao, J. Liu, X. Li, E. Burdet, and S. J. Phee, "A three-limb teleoperated robotic system with foot control for flexible endoscopic surgery," *Ann. Biomed. Eng.*, vol. 49, no. 9, pp. 2282–2296, Sep. 2021.

[11] H.-E. Huang, S.-Y. Yen, C.-F. Chu, F.-M. Suk, G.-S. Lien, and C.-W. Liu, "Autonomous navigation of a magnetic colonoscope using force sensing and a heuristic search algorithm," *Sci. Rep.*, vol. 11, no. 1, pp. 1–15, Aug. 2021.

[12] Y. Xu, K. Li, Z. Zhao, and M. Q.-H. Meng, "Autonomous magnetic navigation framework for active wireless capsule endoscopy inspired by conventional colonoscopy procedures," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 1729–1736, Apr. 2022.

[13] J. W. Martin, B. Scaglioni, J. C. Norton, V. Subramanian, A. Arezzo, K. L. Obstein, and P. Valdastri, "Enabling the future of colonoscopy with intelligent and autonomous magnetic manipulation," *Nature Mach. Intell.*, vol. 2, no. 10, pp. 595–606, Oct. 2020.

[14] T. Yang, Y. Yang, P. Wang, Y. Cao, Z. Yang, and H. Liu, "A lumen-adapted navigation scheme with spatial awareness from monocular vision for autonomous robotic endoscopy," *Robot. Auto. Syst.*, vol. 165, Jul. 2023, Art. no. 104444.

[15] K. B. Ozyoruk, G. I. Gokceler, T. L. Bobrow, G. Coskun, K. Incetan, Y. Almalioglu, F. Mahmood, E. Curto, L. Perdigoto, M. Oliveira, H. Sahin, H. Araujo, H. Alexandrino, N. J. Durr, H. B. Gilbert, and M. Turan, "EndoSLAM dataset and an unsupervised monocular visual odometry and depth estimation approach for endoscopic videos," *Med. Image Anal.*, vol. 71, Jul. 2021, Art. no. 102058.

[16] R. Reilink, S. Stramigioli, and S. Misra, "Image-based flexible endoscope steering," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2010, pp. 2339–2344.

[17] Q. Zhang, J. M. Prendergast, G. A. Formosa, M. J. Fulton, and M. E. Rentschler, "Enabling autonomous colonoscopy intervention using a robotic endoscope platform," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 6, pp. 1957–1968, Jun. 2021.

[18] N. van der Stap, F. van der Heijden, and I. A. M. J. Broeders, "Towards automated visual flexible endoscope navigation," *Surgical Endoscopy*, vol. 27, no. 10, pp. 3539–3547, Oct. 2013.

[19] J. M. Prendergast, G. A. Formosa, M. J. Fulton, C. R. Heckman, and M. E. Rentschler, "A real-time state dependent region estimator for autonomous endoscope navigation," *IEEE Trans. Robot.*, vol. 37, no. 3, pp. 918–934, Jun. 2021.

[20] 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 *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2022, pp. 9582–9588.

[21] D. Jha, S. Ali, N. K. Tomar, H. D. Johansen, D. Johansen, J. Rittscher, M. A. Riegler, and P. Halvorsen, "Real-time polyp detection, localization and segmentation in colonoscopy using deep learning," *IEEE Access*, vol. 9, pp. 40496–40510, 2021.

[22] S.-Y. Yen, H.-E. Huang, G.-S. Lien, C.-W. Liu, C.-F. Chu, W.-M. Huang, and F.-M. Suk, "Automatic lumen detection and magnetic alignment control for magnetic-assisted capsule colonoscope system optimization," *Sci. Rep.*, vol. 11, no. 1, pp. 1–10, Mar. 2021.

[23] J. F. Lazo, C.-F. Lai, S. Moccia, B. Rosa, M. Catellani, M. de Mathelin, G. Ferrigno, P. Breedveld, J. Dankelman, and E. De Momi, "Autonomous intraluminal navigation of a soft robot using deep-learning-based visual servoing," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2022, pp. 6952–6959.

[24] L. Y. Korman, V. Egorov, S. Tsuryupa, B. Corbin, M. Anderson, N. Sarvazyan, and A. Sarvazyan, "Characterization of forces applied by endoscopists during colonoscopy by using a wireless colonoscopy force monitor," *Gastrointestinal Endoscopy*, vol. 71, no. 2, pp. 327–334, Feb. 2010.

[25] L. Ott, F. Nageotte, P. Zanne, and M. de Mathelin, "Robotic assistance to flexible endoscopy by physiological-motion tracking," *IEEE Trans. Robot.*, vol. 27, no. 2, pp. 346–359, Apr. 2011.

[26] M. Kang, S. Joe, T. An, H. Jang, and B. Kim, "A novel robotic colonoscopy system integrating feeding and steering mechanisms with self-propelled paddling locomotion: A pilot study," *Mechatronics*, vol. 73, Feb. 2021, Art. no. 102478.

[27] M. Kmich, H. Karmouni, I. Harrade, A. Daoui, and M. Sayyouri, "Image-based visual servoing techniques for robot control," in *Proc. Int. Conf. Intell. Syst. Comput. Vis. (ISCV)*, May 2022, pp. 1–6.

[28] Y. Li, W. Y. Ng, Y. Huang, Y. Jiang, Y. Xian, W. Xin, P. W. Y. Chiu, and Z. Li, "Towards semi-autonomous manipulation of an electromagnetically actuated soft-tethered colonoscope based on visual servo control," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2022, pp. 2075–2080.
[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[30] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
[31] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
[32] A. Loquercio, A. I. Maqueda, C. R. del-Blanco, and D. Scaramuzza, "DroNet: Learning to fly by driving," *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 1088–1095, Apr. 2018.
[33] F. Mattioli, C. Porcaro, and G. Baldassarre, "A 1D CNN for high accuracy classification and transfer learning in motor imagery EEG-based brain-computer interface," *J. Neural Eng.*, vol. 18, no. 6, Dec. 2021, Art. no. 066053.
[34] J. Kuester, W. Gross, and W. Middelmann, "1D-convolutional autoencoder based hyperspectral data compression," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XLIII-B1-2021, pp. 15–21, Jun. 2021.
[35] M. Moradi A., S. A. Sadrossadat, and V. Derhami, "Long short-term memory neural networks for modeling nonlinear electronic components," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 11, no. 5, pp. 840–847, May 2021.
[36] L. Ma and S. Tian, "A hybrid CNN-LSTM model for aircraft 4D trajectory prediction," *IEEE Access*, vol. 8, pp. 134668–134680, 2020.
[37] J. Wang, G. Wen, S. Yang, and Y. Liu, "Remaining useful life estimation in prognostics using deep bidirectional LSTM neural network," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Chongqing)*, Oct. 2018, pp. 1037–1042.
[38] D. Zhou, X. Zhuang, and H. Zuo, "A hybrid deep neural network based on multi-time window convolutional bidirectional LSTM for civil aircraft APU hazard identification," *Chin. J. Aeronaut.*, vol. 35, no. 4, pp. 344–361, Apr. 2022.
[39] A. Plooy, A. Hill, M. Horswill, A. Cresp, R. Karamatic, S. Riek, G. Wallis, R. Burgess-Limerick, D. Hewett, and M. Watson, "The efficacy of training insertion skill on a physical model colonoscopy simulator," *Endoscopy Int. Open*, vol. 4, no. 12, pp. E1252–E1260, Sep. 2016.
[40] R. Nakadate, T. Iwasa, S. Onogi, J. Arata, S. Oguri, Y. Okamoto, T. Akahoshi, M. Eto, and M. Hashizume, "Surgical robot for intraluminal access: An ex vivo feasibility study," *Cyborg Bionic Syst.*, vol. 2020, pp. 1–9, Jan. 2020.

VAN SY NGUYEN received the B.S degree from Viet Nam National University Ho Chi Minh City (VNU-HCM)-Ho Chi Minh City University of Technology, in 2017, and the M.S./Ph.D. Integrated degree from Korea Aerospace University, in 2023. He is currently a Postdoctoral Research Associate with the Department of Mechanical and Aerospace Engineering, University of Central Florida, USA. His research interests include robotics, mechanisms and control, bio/medical application robots, computer vision, and deep learning.

BOHYUN HWANG received the B.S. and M.S. degrees in mechanical engineering from Korea Aerospace University, in 2015 and 2017, respectively, where he is currently pursuing the Ph.D. degree. His research interests include bio/medical application robots and mechanism design.

BYUNGKYU KIM received the Ph.D. degree in mechanical engineering from the University of Wisconsin, Madison, in 1997. From 1997 to 2000, he was a Technical Staff Member with the Center for X-ray Lithography, University of Wisconsin, where he developed a computer code for thermal modeling of a mask membrane and wafer during beam exposure. From 2000 to 2005, he was with the Microsystem Center, KIST, as a Principal Research Scientist. He was in charge of developing a microcapsule-type robotic endoscope. He is currently a Professor with Korea Aerospace University. His research interests include space mechanism, robotics, micro/nano-manipulator, and bio/medical application robots.

JAY HOON JUNG received the Ph.D. degree in computer science from the State University of New York at Stony Brook. He is currently an Assistant Professor with the Artificial Intelligence Department, Korea Aerospace University. His primary research delves into uncovering the foundational principles of artificial neural networks. He aspires to leverage these insights to develop quantum algorithms for deep learning. Additionally, he is invested in creating efficient algorithms for daily-life applications, particularly in the realm of autonomous systems.