Deep Learning Assisted Robotic Magnetic Anchored and Guided Endoscope for Real-Time Instrument Tracking

Truman Cheng, Weibing Li, *Member, IEEE*, Wing Yin Ng, Yisen Huang, Jixiu Li, Calvin Sze Hang Ng, Philip Wai Yan Chiu, and Zheng Li, *Senior Member, IEEE*

***Abstract*—This letter presents the first case of implementing deep-learning based instrument tracking on a magnetic anchored surgical endoscope. The compact magnetic actuated endoscope has a unique structure that allows operations near the anchor surface, ideal for video assisted thoracoscopic surgery (VATS). Autonomous tool tracking alleviates surgeon's burden and prevents human errors from muscle fatigues or miscommunication. However, conventional methods rely on color labels or require modification to instrument, and has risk of failure due to occlusion of marker. In this letter, we combine deep-learning instrument detection with visual servoing control. This allows the magnetic endoscope to track surgical tools automatically, without color markers or instrument modification. We used a modified TernausNet-16 network that can detect surgical instrument in real time, with a small training dataset of 1846 images. Experiments show that the magnetic endoscope can effectively track a marker-less instrument. It can also track continuous motions of a target traveling at 40 mm/s. The performance was also verified by completing mock-up surgical task in a simulated thoracic cavity.**

***Index Terms*—Medical robotics, machine learning, visual servoing.**

I. Introduction

IN THE past two decades, minimally invasive surgery (MIS) has become the standard practice in many surgical fields. MIS offers many patient benefits over conventional open surgery. These include reduced trauma and blood loss, less post-operative pain, faster recovery and better cosmetics. The standard practice for gaining visuals in MIS is to use a laparoscope. The laparoscope is a long, rigid optical instrument that provides visual feedback from inside the patient. To view different surgical targets, surgeons would pivot the laparoscope. Therefore, the perspective is limited by the fulcrum at abdomen wall or chest wall.

The majority of MIS nowadays are performed through multiple incisions for inserting the laparoscope and other instruments. There are emerging techniques that use a single incision to reduce invasiveness. While single port surgery benefits patients with faster recovery, there are unique challenges posed by passing multiple instruments through one port. A major drawback is the diminished triangulation, causing counter intuitive controls. The configuration also causes device clashing and interference. These factors further limit dexterity of the laparoscope, giving less than ideal perspective of surgical site. In some cases, the shift to multi-port or even open surgery becomes necessary.

These problems become even more pronounced in video assisted thoracoscopic surgery (VATS), where insufflation is impossible due to the ribcage, resulting in highly limited workspace. The ribs can also hinder the motions of the laparoscope and other instruments.

One solution to overcome these challenges is to replace the physical body of laparoscopes with a magnetic linkage. Magnetic anchored endoscopes typically involve two units. The internal unit includes the imaging sensor and internal permanent magnets (IPMs), whereas the external unit consists of the external permanent magnet (EPM) that is usually manually controlled. The endoscope can be steered by magnetic guidance along the inside of abdominal wall. This offers more freedom in endoscope placing, allowing surgeons to gain preferred perspectives without opening new incisions for the laparoscopes.

An early example is the MAGS endoscope reported by Cadeddu *et al.* Although this endoscope requires surgeons to manually press on the abdominal wall for changing the view direction, the device has shown success in guiding the completion of single port nephrectomy and appendectomy in humans [1]. Adjusting view by pressing on abdomen is less precise and affects insufflation. To overcome this, multiple research groups

Manuscript received October 15, 2020; accepted February 16, 2021. Date of publication March 17, 2021; date of current version April 6, 2021. This letter was recommended for publication by Associate Editor H. Zha and Editor C. Cadena Lerma upon evaluation of the reviewers' comments. This work was supported in part by Research Grant Council General Research Fund under Projects 14203019 and 14202820, and in part by Early Career Scheme under Project 24204818. *(T. Cheng and W. Li contributed equally to this work.) (Corresponding author: Zheng Li.)*

Truman Cheng, Yisen Huang, Jixiu Li, and Calvin Sze Hang Ng are with the Department of Surgery, The Chinese University of Hong Kong, Hong Kong (e-mail: chengtruman@gmail.com; yisenhuang@link.cuhk.edu.hk; jxli@surgery.cuhk.edu.hk; calvinng@surgery.cuhk.edu.hk).

Weibing Li is with the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China (e-mail: wellbeinglwb@gmail.com).

Wing Yin Ng is with the Chow Yuk Ho Technology Centre for Innovative Medicine, The Chinese University of Hong Kong, Hong Kong (e-mail: wingyin1997@gmail.com).

Philip Wai Yan Chiu is with the Department of Surgery and the Chow Yuk Ho Technology Centre for Innovative Medicine, The Chinese University of Hong Kong, Hong Kong (e-mail: philipchiu@surgery.cuhk.edu.hk).

Zheng Li is with the Department of Surgery, Chow Yuk Ho Technology Centre of Innovative Medicine, Li Ka Shing Institute of Health Science, and Multi-scale Medical Robotics Ltd., The Chinese University of Hong Kong, Hong Kong (e-mail: lizheng@cuhk.edu.hk).

Digital Object Identifier 10.1109/LRA.2021.3066834

have proposed motorized magnetic endoscopes [2], [3]. While these systems offer more dexterities, the use of DC motors can add to the weight and complexity of design, and introduces electrical safety concerns. Recently, magnetic actuation has been proposed as a solution for realizing internal dexterity without motors. For example, using a spherical IPM for pivoting the endoscope within a hemispheric workspace [4]. Tan *et al.* reported an untethered robotic laparoscopic surgical camera. Using two IPMs for anchoring and panning control, and another IPM for tilting, the 68 mm camera (excluding the IPMs) is promising for single port abdominal surgery [5].

In typical VATS, the surgeon deflates one side of the lungs to create space for instruments including the laparoscope. This space can also be used to introduce and operate a magnetic anchored endoscope. Unfortunately, many magnetic anchored endoscopes reported in the literature are not intended for VATS. The classic MAGS endoscope requires deflation of abdominal wall for view adjustment, which is impossible for the thorax reinforced by ribcage. Other devices with internal dexterities usually involve moving the camera through a conical or hemispheric workspace, which can cause poor view and contact with underlying organs in the thorax.

Motivated by this, we have explored designing a magnetic anchored endoscope dedicated to VATS, with a compact body and operation near the chest wall [6]. We also observed many of the magnetic anchored surgical endoscopes rely on manually controlled EPM, or button prompted motor actuations. These require a dedicated endoscope operator, causing bed side crowding, and introduce risks from muscle fatigues and miscommunications. A robotic controller can overcome these problems.

We previously reported the first case of autonomous instrument tracking with color targets, using a novel magnetic endoscope dedicated to VATS [6], and demonstrated the feasibility of combining visual servoing control and magnetic surgery. However, the previous system lacked some functions crucial for clinical applications, such as on-board illumination and lens cleaning function, which has significant impact on endoscopic image quality [7]. The previous robot controller has low payload (500 g), which limits the EPM's size.

Considering the above, we design a new magnetic endoscope with self-reliant illumination using on-board LEDs and rinsing features for cleaning blood contamination. The new system uses the UR5 robot arm (Universal Robots, Odense, Denmark) for controlling the external magnet. With 5 kg payload, the external magnets were upsized for better anchoring range. This involves designing a customized end effector to hold and control rotation of the EPM, and kinematic analysis for visual servoing control.

Color detection also has risk of failure due to blood occlusion of marker. A detection method that does not rely on color labels can also make instrument tracking more robust in clinical settings. In recent years, using deep learning to assist in patient screening and diagnosis has become an emerging field in medical imaging. Deep learning can also be applied to surgical instrument segmentation. Unlike diagnostic applications, instrument detection has lower requirements on accuracy, but demands real time feedback. Although a relatively new field, there are examples in the literature for deep-learning based surgical instrument methods [8], [9], [10].

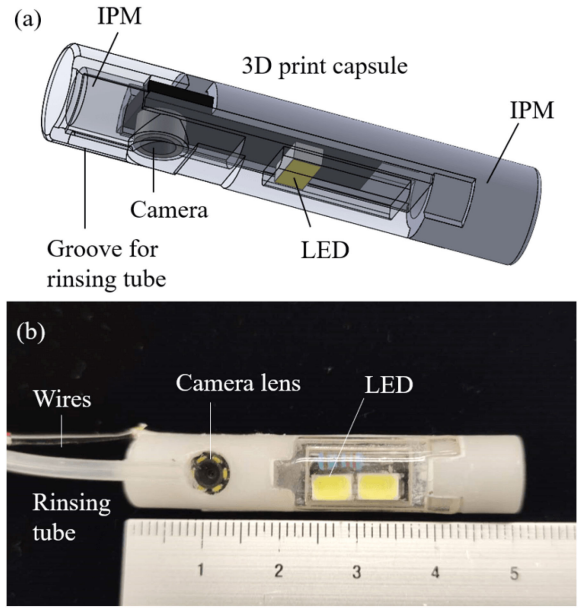

Fig. 1. Magnetic anchored and actuated endoscope. (a) Design schematics. (b) Prototype.

In this letter, we report the first case of deep learning based instrument tracking on a magnetic anchored endoscope. we selected a TernausNet method after comparing with MobileNet, and ShuffleNet. This method is robust against blood contamination on lens or instrument, and can bring magnetic anchored endoscope closer towards clinical applications.

The rest of the letter is organized as follows. Section II describes the design and working principle of the magnetic anchored endoscope. Section III details the kinematics modeling and visual servoing control. Section IV discusses the deep neural network method for instrument segmentation. Section V shows the experiments on instrument detection and automatic tracking. Section VI summarizes and concludes this work.

II. Design and Working Principle

A. Design Overview

The system consists of a magnetic anchored endoscope and an external robot controller.

1) Magnetic Anchored and Actuated Endoscope: Fig. 1 shows the endoscope design schematics and the prototype. It consists of a 5 cm $\times$ 1 cm cylinder capsule. Inside the capsule there are two internal permanent magnets (IPMs), a camera module and two LEDs. The structural design is shown in Fig. 1(a). The capsule is custom designed, then 3D printed with ABS resin. It is a three parts assembly, with two parts printed in opaque white, and one part transparent. The IPMs are identical 6 mm $\times$ 6 mm Neodymium-Boron-Iron magnets

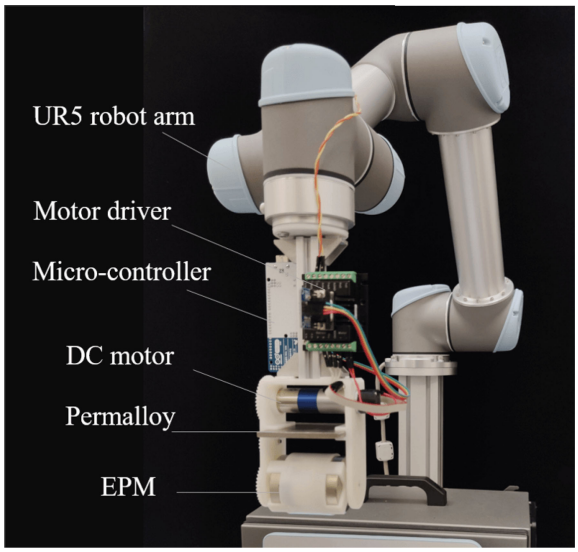

Fig. 2. The external robot controller.

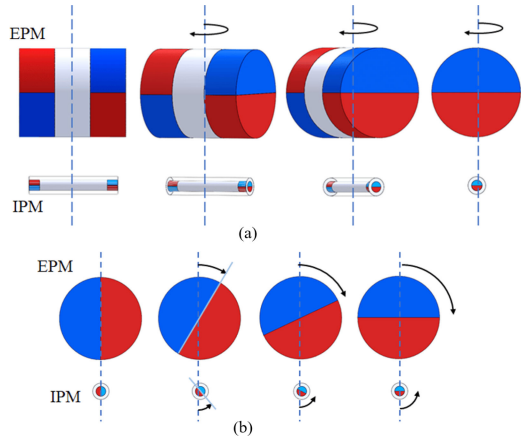

Fig. 3. Actuation principle of the magnetic endoscope.

(Grade N42) embedded at the two ends of the capsule. They are diametrically magnetized, and fixed in the capsule with opposite polar direction. The camera board is a 22 mm $\times$ 7 mm circuit (BL31107, MISUMI). The image sensor is soldered parallel to the circuit board, so the camera view is normal to the cylinder length. When the capsule rolls about its long axis, the view direction can be adjusted. There is a circular opening in the capsule, where a glass lens is installed to protect the camera. Two LEDs, each 5.6 mm $\times$ 3 mm, are installed in parallel with the camera circuit. They are located behind the transparent part of the capsule, protected from blood and other contaminants, and their illumination points in a same direction as the camera view. The assembled capsule encloses all components, with a small opening for cables connecting to the camera and LEDs. All gaps between the capsule parts are sealed with silicone epoxy. The completed endoscope prototype is shown in Fig. 1(b).

2) External Robot Controller: The external robot controller consists of a robot arm, an end effector, and the EPMs. The EPMs are two identical Neodymium-Boron-Iron cylindrical magnets. Each is 2 cm thick and 5 cm in diameter. The EPMs are diametrically magnetized, and arranged at opposite polar directions. Part of the end effector is a magnet holder, with a buffer creating a 2 cm gap between the EPMs. The end effector also includes a U shaped bracket frame, a DC motor (2237 CXR, Faulhaber) with an optical encoder (IER3-1000), a motor driver unit (MCDC 3006), an Arduino microcontroller, a gear set connecting the motor to the magnet holder, and a 5 mm thick permalloy plate for magnetic shielding. The layout is shown in Fig. 2. The bracket frame, magnet holder and gears are non-ferromagnetic 3D printed material. The end effector is connected to the robot arm via a 15 cm aluminum extrusion. This protects the robot joint encoders from magnetic interference of the EPMs. The robot arm is the UR5 collaborative robot (Universal Robots) with 6 rotational joints, 5 kg payload. The high payload makes it possible to increase the size of EPMs in this design compared to our previous version.

B. Working Principle

The cylindrical endoscope capsule is inserted via the MIS incision, into the patient's chest cavity. The capsule diameter (1 cm) fits the size of MIS trocar. Once the endoscope is inside the patient, the EPMs can be placed on the other side of the chest wall, near the patient's skin. The magnetic coupling attracts the endoscope capsule against gravity, and anchors it under the chest wall. The robot arm can change the location and panning orientation of the EPMs. This guides the endoscope to slide along the inside of the chest wall, and pans to alter camera view left and right [6].

The end effector controls the rolling orientation of the EPMs. The change of magnetic field causes the endoscope capsule to roll in opposite direction, as shown in Fig. 3. Since the camera view is normal to capsule length, rolling of capsule adjusts the camera view up and down. These combine to control two linear degrees-of-freedoms (DOFs) and two rotational DOFs of the endoscope. The endoscope can be guided to many locations, without the limit of abdominal or thoracic fulcrum. These allow perspectives not possible with conventional laparoscopy.

III. Kinematic Modeling and Control of Magnetic Endoscope

To automate the endoscope using visual servoing techniques, this section details the kinematic modeling of the endoscope system. Then, a visual servoing control law is described. A deep learning surgical instrument detection method is presented in the next section to complete the control loop.

A. Kinematic Modeling

The robotic system consists of two elements: a UR5 robot with EPMs and an endoscope with IPMs. For kinematic modeling, coordinate frames are assigned for the robotic system as shown in Fig. 4. Considering the working principle of the MAGS endoscope, motions of the EPMs and IPMs are synchronized to exhibit four DOFs. To better describe the motions of the EPMs and IPMs, an auxiliary fixed coordinate frame $\{O_{0'}, X_{0'}Y_{0'}Z_{0'}\}$ is established.

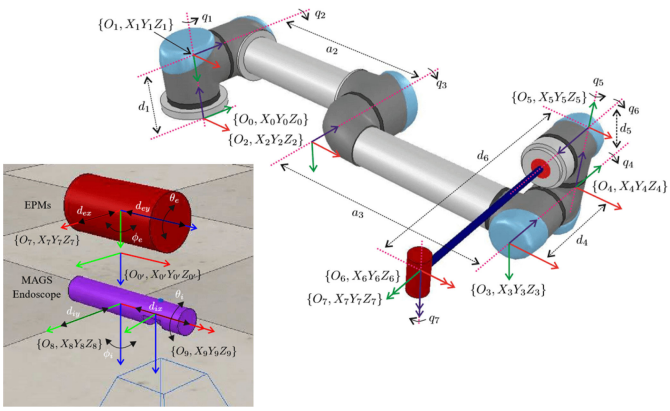

Fig. 4. Coordinate frames of the endoscope system.

1) Mapping From Task Space to Image Space: The well-known pinhole camera model leads to the following kinematic equation [11]:

$$\dot{\boldsymbol{s}} = \boldsymbol{J}_{image}\,{}^{9}\dot{\boldsymbol{\xi}}_c \tag{1}$$

where

$$\boldsymbol{J}_{image} = \begin{bmatrix} -\frac{\bar{f}}{z} & 0 & \frac{\bar{u}}{z} & \frac{\bar{u}\bar{v}}{\bar{f}} & -\frac{\bar{f}^2+\bar{u}^2}{\bar{f}} & \bar{v} \\ 0 & -\frac{\bar{f}}{z} & \frac{\bar{v}}{z} & \frac{\bar{f}^2+\bar{v}^2}{\bar{f}} & -\frac{\bar{u}\bar{v}}{\bar{f}} & -\bar{u} \end{bmatrix} \tag{2}$$

is the Jacobian that maps the instantaneous linear and angular velocities of the camera ${}^{9}\dot{\boldsymbol{\xi}}_c \in \mathbb{R}^6$ to the image feature velocities $\dot{\boldsymbol{s}} = [\dot{u}, \dot{v}]^{\mathrm{T}}$. In the image Jacobian, $\bar{u} = u - u_0$ and $\bar{v} = v - v_0$ with $\boldsymbol{s}_0 = [u_0, v_0]^{\mathrm{T}}$ indicating the principal point of the image plane. Besides, $\bar{f} = f/\rho$ where f and ρ denote the focal length and the pixel width of each square pixel, respectively.

2) Mapping From Configuration Space to Task Space: The MAGS endoscope with IPMs exhibits a total of four DOFs including two translational DOFs and two rotational DOFs with respect to the auxiliary coordinate frame $\{O_{0'}, X_{0'}Y_{0'}Z_{0'}\}$. The pose of the camera expressed in coordinate frame $\{O_{0'}, X_{0'}Y_{0'}Z_{0'}\}$ is

$${}^{0'}_{9}\boldsymbol{H} = \begin{bmatrix} c_\phi & s_\phi c_\theta & -s_\phi s_\theta & d_{Ix} + c_2 c_\phi \\ -s_\phi & c_\phi c_\theta & -c_\phi s_\theta & d_{Iy} - c_2 s_\phi \\ 0 & s_\theta & c_\theta & c_1 + c_2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

where $s_\phi := \sin(\phi)$, $c_\phi := \cos(\phi)$, $s_\theta = \sin(\theta)$ and $c_\theta = \cos(\theta)$, with $\phi := \phi_i$ and $\theta := \theta_i$ representing the panning and tilting angles of the IPMs. c_1 and c_2 are constant parameters relevant to the translations between the involved coordinate frames. Naturally, it yields the following kinematic equation:

$${}^{0'}\dot{\boldsymbol{\xi}}_c = {}^{0'}\boldsymbol{J}_{mags}\dot{\boldsymbol{\zeta}}_i \tag{4}$$

where

$${}^{0'}\dot{\boldsymbol{\xi}}_c = \begin{bmatrix} {}^{0'}_{9}\boldsymbol{R} & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & {}^{0'}_{9}\boldsymbol{R} \end{bmatrix} {}^{9}\dot{\boldsymbol{\xi}}_c \tag{5}$$

is the linear and angular velocities of the camera in the coordinate frame $\{O_{0'}, X_{0'}Y_{0'}Z_{0'}\}$ with ${}^{0'}_{9}\boldsymbol{R}$ being the rotation matrix between coordinate frames $\{O_9, X_9Y_0Z_9\}$ and $\{O_{0'}, X_{0'}Y_{0'}Z_{0'}\}$. In addition, $\dot{\boldsymbol{\zeta}}_i = [\dot{d}_{ix}, \dot{d}_{iy}, \dot{\phi}_i, \dot{\theta}_i]^{\mathrm{T}}$ stands for the translational and rotational velocities of the IPMs, and the Jacobian matrix is

$${}^{0'}\boldsymbol{J}_{mags} = \begin{bmatrix} 1 & 0 & -c_2 s_\phi & 0 \\ 0 & 1 & -c_2 c_\phi & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & c_\phi \\ 0 & 0 & 0 & -s_\phi \\ 0 & 0 & -1 & 0 \end{bmatrix}. \tag{6}$$

3) Mapping From Actuation Space to Configuration Space: Relations of velocities of the EPMs and IPMs' DOFs are quite straightforward on the basis of the working principle:

$$\dot{\boldsymbol{\zeta}}_i = \boldsymbol{J}_{actuator}\dot{\boldsymbol{\zeta}}_e \tag{7}$$

with Jacobian matrix

$$\boldsymbol{J}_{actuator} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{8}$$

where $\dot{\boldsymbol{\zeta}}_e = [\dot{d}_{ex}, \dot{d}_{ey}, \dot{\phi}_e, \dot{\theta}_e]^{\mathrm{T}}$ represents the translational and rotational velocities of the EPMs.

For the UR5 robot attached with EPMs, the pose of the EPMs with respect to coordinate frame $\{O_0, X_0Y_0Z_0\}$ is

$${}^{0}_{7}\boldsymbol{H} = \prod_{k=1}^{k=7} {}^{k-1}_{k}\boldsymbol{H} \in \mathbb{R}^{4\times4} \tag{9}$$

where

$${}^{k-1}_{k}\boldsymbol{H} = \begin{bmatrix} c_{q_k} & -s_{q_k}c_{\alpha_k} & s_{q_k}s_{\alpha_k} & a_k c_{q_k} \\ s_{q_k} & c_{q_k}c_{\alpha_k} & -c_{q_k}s_{\alpha_k} & a_k s_{q_k} \\ 0 & s_{\alpha_k} & c_{\alpha_k} & d_k \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{10}$$

represents the transformation matrix from coordinate frame $\{O_k, X_kY_kZ_k\}$ to coordinate frame $\{O_{i-1}, X_{i-1}Y_{i-1}Z_{i-1}\}$ [12]. In transformation matrix ${}^{k-1}_{k}\boldsymbol{H}$, q_k, α_k, a_k and d_k are standard Denavit-Hartenberg parameters of the robot shown in Fig. 4 with $c_{q_k} := \cos(q_k)$, $s_{q_k} := \sin(q_k)$, $c_{\alpha_k} := \cos(\alpha_k)$ and $s_{\alpha_k} := \sin(\alpha_k)$ satisfied. Based on the above kinematic relationship, the corresponding differential kinematics is obtained:

$${}^{0}\boldsymbol{J}_{robot}\dot{\boldsymbol{q}} = {}^{0}\dot{\boldsymbol{\xi}}_e \tag{11}$$

where ${}^{0}\boldsymbol{J}_{robot} \in \mathbb{R}^{6\times7}$ is the Jacobian mapping joint velocity $\dot{\boldsymbol{q}} = [\dot{q}_1, \dot{q}_2, \ldots, \dot{q}_7]^{\mathrm{T}}$ to the EPMs' spatial velocity ${}^{0}\dot{\boldsymbol{\xi}}_e$ expressed in the robot base frame $\{O_0, X_0Y_0Z_0\}$. As a result, the EPMs' spatial velocity expressed in coordinate frame $\{O_{0'}, X_{0'}Y_{0'}Z_{0'}\}$ satisfies

$${}^{0'}\boldsymbol{J}_{robot}\dot{\boldsymbol{q}} = {}^{0'}\dot{\boldsymbol{\xi}}_e \tag{12}$$

where

$$ {}^{0'}\boldsymbol{J}_{robot} = \begin{bmatrix} {}^{0}_{0'}\boldsymbol{R} & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & {}^{0}_{0'}\boldsymbol{R} \end{bmatrix}^{\mathrm{T}} {}^{0}\boldsymbol{J}_{robot} \in \mathbb{R}^{6\times7} \tag{13} $$

is the Jacobian matrix with ${}^{0}_{0'}\boldsymbol{R}$ denoting the rotation matrix between coordinate frames $\{O_{0'}, X_{0'}Y_{0'}Z_{0'}\}$ and $\{O_0, X_0Y_0Z_0\}$.

To actuate the MAGS endoscope, the UR5 robot's six joints are assigned to control the first three DOFs (i.e., d_{ex}, d_{ey} and ϕ_e) of the EPMs, while the customized rotational joint is requested to control the fourth DOF (i.e., θ_e) of the EPMs. Then, the following relationships hold true:

$$ \begin{cases} {}^{0'}\bar{\boldsymbol{J}}_{robot}\dot{\bar{\boldsymbol{q}}} = {}^{0'}\dot{\boldsymbol{\xi}}_e \\ \dot{q}_7 = \dot{\theta}_e \end{cases} \tag{14} $$

where $\bar{\boldsymbol{J}}_{robot} \in \mathbb{R}^{6\times6}$ is the Jacobian extracted from the first 6×6 block of $\boldsymbol{J}_{robot}$, $\dot{\bar{\boldsymbol{q}}} = [\dot{q}_1, \dot{q}_2, \ldots, \dot{q}_6]^{\mathrm{T}} \in \mathbb{R}^6$, and ${}^{0'}\dot{\boldsymbol{\xi}}_e := [{}^{0'}\dot{p}_{ex}, {}^{0'}\dot{p}_{ey}, {}^{0'}\dot{p}_{ez}, {}^{0'}\dot{w}_{ex}, {}^{0'}\dot{w}_{ey}, {}^{0'}\dot{w}_{ez}]^{\mathrm{T}} \in \mathbb{R}^6$ with ${}^{0'}\dot{p}_{ex} = \dot{d}_{ex}, {}^{0'}\dot{p}_{ey} = \dot{d}_{ey}, {}^{0'}\dot{p}_{ez} = 0, {}^{0'}\dot{w}_{ex} = 0, {}^{0'}\dot{w}_{ey} = 0, {}^{0'}\dot{w}_{ez} = \dot{\phi}_e$.

4) Integrated Kinematics: Based on the above analysis, the following kinematic equation is attained to link the image space and the actuation space:

$$ \dot{\boldsymbol{s}} = \boldsymbol{J}_{visual}\dot{\boldsymbol{\zeta}}_e \tag{15} $$

where $\boldsymbol{J}_{visual} = \boldsymbol{J}_{image}{}^{9}\boldsymbol{J}_{mags}\boldsymbol{J}_{actuator} \in \mathbb{R}^{2\times4}$ is the Jacobian matrix and

$$ {}^{9}\boldsymbol{J}_{mags} = \begin{bmatrix} {}^{0'}_{9}\boldsymbol{R} & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & {}^{0'}_{9}\boldsymbol{R} \end{bmatrix}^{\mathrm{T}} {}^{0'}\boldsymbol{J}_{mags}. \tag{16} $$

By solving (15), the EPMs' translational and rotational velocities $\dot{\boldsymbol{\zeta}}_e = [\dot{d}_{ex}, \dot{d}_{ey}, \dot{\phi}_e, \dot{\theta}_e]^{\mathrm{T}}$ can obtained. Then, the resultant EPM's velocities can be further substituted into (14) to resolve the joint velocities of the robot controller.

B. Visual Servoing Controller

A classical solution to kinematic equation (15) is

$$ \dot{\boldsymbol{\zeta}}_e = \boldsymbol{J}^{\dagger}_{visual}\dot{\boldsymbol{s}} \tag{17} $$

where $\boldsymbol{J}^{\dagger}_{visual}$ denotes the pseudoinverse of Jacobian $\boldsymbol{J}_{visual}$. The control law in (17) provides a least square solution to determine translational and rotational velocities of the EPMs based on the target instrument's position. In some occasions, it is desired to adjust the camera view by rotating instead of translating the MAGS endoscope. Inspired by this fact, the following control law is selected:

$$ \dot{\boldsymbol{\zeta}}_e = \boldsymbol{W}^{-1}\boldsymbol{J}^{\mathrm{T}}_{visual}(\boldsymbol{J}_{visual}\boldsymbol{W}^{-1}\boldsymbol{J}^{\mathrm{T}}_{visual} + \delta\boldsymbol{I})^{-1}\dot{\boldsymbol{r}} \tag{18} $$

to deliver a weighted least norm solution [13], where $\boldsymbol{W} = \mathrm{diag}\{w_1, w_2, w_3, w_4\}$ is a diagonal weighting matrix and $\delta > 0$ is a damping gain. This allows weighted motions of the available four DOFs to be controlled. The joint velocities of the robot controller are eventually resolved as

$$ \begin{cases} \dot{\bar{\boldsymbol{q}}} = {}^{0'}\bar{\boldsymbol{J}}^{\dagger}_{robot}{}^{0'}\dot{\boldsymbol{\xi}}_e \\ \dot{q}_7 = \dot{\theta}_e \end{cases} \tag{19} $$

by solving kinematic equation (14).

IV. Instrument Detection Using Deep Learning

The visual servoing control requires input of target pixel location. While it is possible to add color markers to instruments and use RGB filters to detect target, it require customizing surgical instruments. Detection can also fail if the color markers are covered by blood or debris. Considering these, a deep learning method for detecting un-modified instruments can improve clinical feasibility of the system.

A. Custom Deep Neural Network for Instrument Detection

In visual servoing of robotic endoscopes, retrieving timely and correct information from the camera is essential. Therefore, custom deep neural networks (DNN) should balance the segmentation accuracy and processing time required for our application. We trained and compared networks based on TernausNet-11, TernausNet-16 [14], MobileNet-V3, and ShuffleNet-V2. To optimize for processing time, we also compared half and quarter reduction of the channel width. For the decoder backbone, we tried both the U-Net expanding path, and the Light-Weight RefineNet (LWRN) decoder. From these, we trained and compared 17 different models combinations. Networks using RGB or grayscale image input are both trained and tested.

To train the network, we created our own training dataset. We recorded endoscope videos of instruments at different locations and orientations, then extracted 1846 image frames (640×480). The images are annotated. For the training, Jaccard index (Intersection over Union, IoU) is used as the evaluation metric. The Jaccard index is defined as follows [14]:

$$ \mathcal{J} = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i x_{ip}}{x_i + x_{ip} - x_i x_{ip}} \tag{20} $$

where x_i is the pixel in ground truth (i.e., the annotation mask for machine recognition), x_{ip} is the predicted pixel after the input image passing through the network, n is the total number of pixels of one image.

Based on the results, we chose the TernausNet-16 [14], which is a U-Net [15] variant using the VGG-16 network [16] as the encoder backbone, with quarter reduction in channel width and grayscale image input. The Jaccard is 85.87% and processing time is only 5.77 ms.

B. Software Implementation

The visual servoing control framework in our system consists of a kinematic controller descripted in Section III, and the deep neural network instrument detector introduced above. These are implemented in the robot operating system (ROS). The kinematic controller is a ROS node coded in C++, with open source Eigen library included for linear algebra computation. The instrument detector is trained, tested and implemented in Pytorch. The DNN receives image streams from the endoscope, then extracts the visual information and determines the location of the target. Then, the information is sent as input to the kinematics controller, which determines the robot controller's

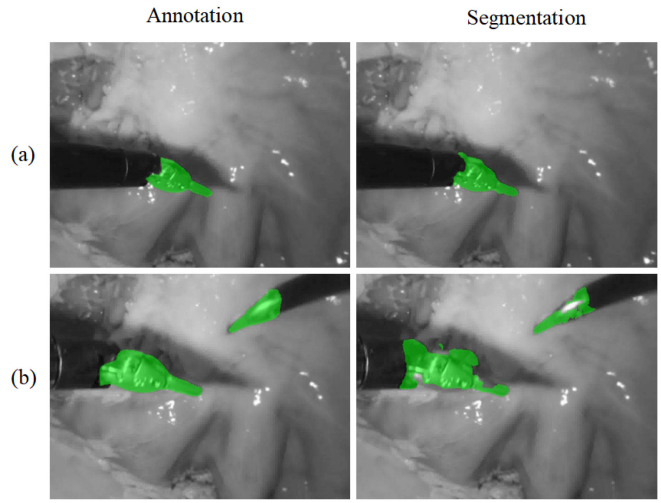

Fig. 5. Comparison of ground truths (left) and segmentation results (right) for (a) one instrument and (b) two instruments.

joint velocities. The robot then actuates and adjusts the pose of the magnetic endoscope. The updated image feed is sent to the DNN again. Repeating this cycle, the system automatically tracks the target instrument towards the center of view, reducing the burden on surgeons.

V. Experimental Validation

A. Test of Deep Neural Network Instrument Detection

To test the accuracy of the instrument detection method, we set up the magnetic endoscope in a black box, and recorded videos of instrument manipulations over porcine stomach and intestines. The LEDs inside the capsule are the only source of illumination. 50 image frames are extracted from the recorded video. These raw images are tested with the DNN to get the detected pixels. The images are then manually labeled to determine ground truths of instruments' pixels. The DNN detection results and ground truths are then compared to determine the segmentation accuracy by the Intersection over Union (eq.20). Fig. 5 shows the comparison examples of annotated ground truths (left) and segmentation results (right). Fig. 5(a) shows results of a single instrument, Fig. 5(b) shows results for two instruments. From the extracted frames, the IoU is determined as 51.17%. This accuracy is effective in determining instrument location for autonomous tracking, while also balancing for the need of real-time feedback. The training dataset and the prototype used different camera modules, which may explain the gap in performance.

B. Test of Lens Rinsing Function

To evaluate the effects of lens rinsing on instrument segmentation, we purposefully contaminated the lens to simulate blood spills in MIS. Using the same setup as above, endoscope images of instrument manipulations over porcine organs are recorded and tested using the DNN. With contaminated lens, instruments can neither be detected by human eyes or by the DNN. We then rinsed the lens by injecting water to the silicone tube attached to capsule, then clear the remaining water droplets by air suction through the same tube. A video of instrument manipulations is recorded. 50 frames are extracted and tested with the DNN, then compared to manually labeled ground truths. The IoU is determined as 50.45%. The results are comparable to those obtained before lens contamination. This shows the rinsing function is effective in cleaning the lens for DNN instrument detection. Fig. 6 shows the effect of rinsing on instrument detection. Fig. 6(a)–(c) present the camera view before rinsing, a raw image after rinsing, and the segmentation after rinsing, respectively.

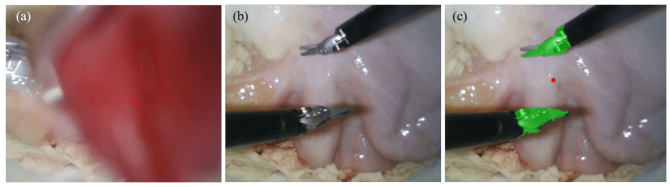

Fig. 6. Effects of rinsing on instrument detection. (a) No detection before rinsing. (b) Raw image after rinsing. (c) Segmentation after rinsing.

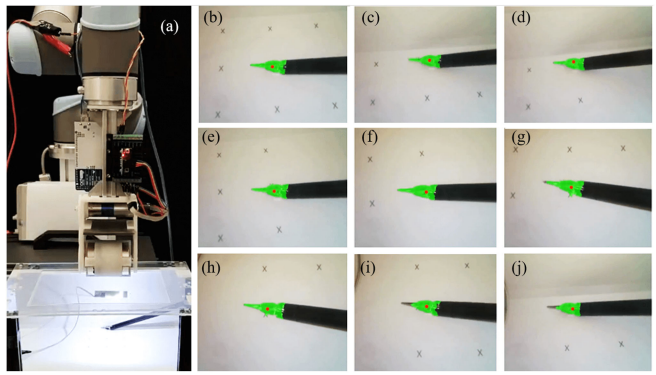

Fig. 7. Tracking instrument over an area. (a) Experimental setup. (b–j) Camera views at each step.

C. Tracking Instrument Over an Area

To test the visual servoing performance combined with the deep learning based instrument segmentation, we conducted a tracking experiment. Fig. 7 shows the experimental setup using an acrylic test platform. The setup approximates parameters in clinical settings. The capsule only travel within an area of 13.5×20 cm. The target path is located 10 cm from the camera. The coupling distance between EPM and endoscope is 3 cm. A surgical instrument (da Vinci Research Kit, Intuitive Surgical) is used as the target. The instrument is detected by deep learning method and color markers in separate tests. Both methods are also repeated with the instrument contaminated using fake blood. During the experiment, the target is moved manually to follow a pre-designed path. To better illustrate the segmentation performance, we overlaid a green mask to indicate

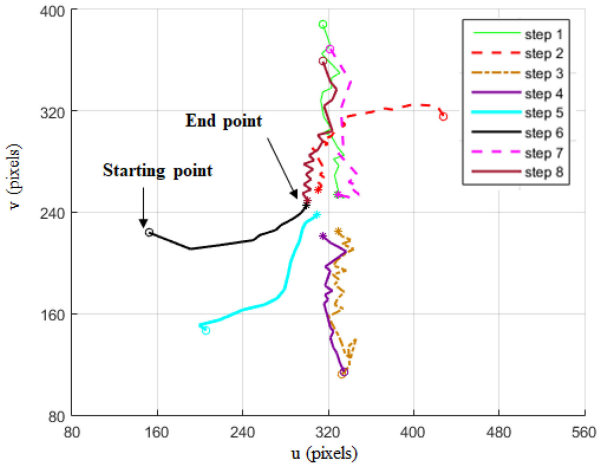

Fig. 8. Target pixel trajectory in each tracking step.

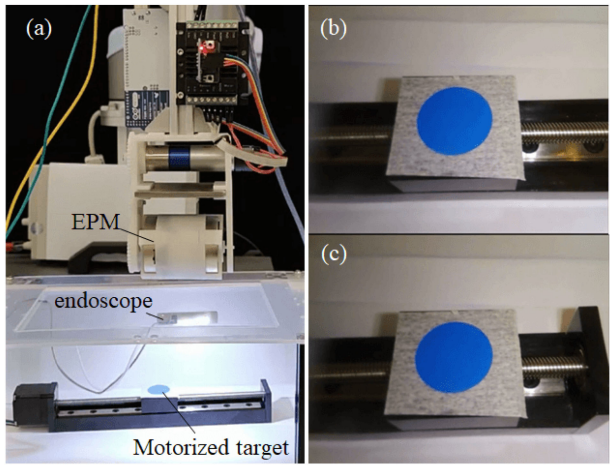

Fig. 9. Tracking a continuously moving target. (a) Experimental setup. (b) and (c) Camera views at initial step and the right boundary.

the detected pixels, and a red dot to present the target pixel used for visual servoing.

The pre-designed path is a nine-point grid. Steps are indicated by crosses, each 3 cm apart, so the target moves in a 6 cm $\times$ 6 cm area. Each test begins with the target at the center point. The target is moved one step up, then follows the path clockwise until it reaches the top left corner. For each step, we wait for the system to track target to center of view, then quickly move the target to the next step. Both the setup and endoscope view are video recorded for time measurements. For deep learning method with clean instrument, the test was repeated three times, each completed in about 1 minute. Tracking overshoot was not observed in any steps. The experiment results are shown in Figs. 7 and 8. Fig. 7(a) shows the test setup and Fig. 7(b)–(j) shows the target tracked to center of view in each step. Fig. 8 shows the target pixel trajectory throughout the test. The starting and end points of each step are labeled with round and asterisk markers, respectively. The endoscope system successfully tracks target to center from various directions. These results demonstrate the feasibility of both the visual servoing control and the deep learning instrument detection.

Tracking task is also completed successfully for color detection with clean instrument, showing the kinematic model can control the endoscope even without deep learning component. For blood contaminated instrument, only deep learning method succeed in detection and tracking, while color detection fails. This demonstrates deep learning method is more reliable in surgery.

D. Tracking Speed Test

To evaluate the system's ability in tracking a target in continuous motions, we conducted another test using a motorized target. A colored target is used here, as the test focuses only on system motion control. The custom testing platform uses a DC motor to drive linear motions of a target at specified velocity and distance. We conducted tracking tests at target velocities 1, 2, 3, 4, 5 cm/s. The target begins at the center, then moves to the right for 6 cm, and then to the left for 12 cm. No human interventions are involved during each test. The target is motorized, and the endoscope system tracks it automatically. If the camera loses sight of the target at any point, the test is classified as failure for that speed. Fig. 9(a) shows the experiment setup. Fig. 9(b) and (c) show the target at the initial position and right boundary, respectively. The system succeeds at target velocities 1 to 4 cm/s and fails at the target velocity of 5 cm/s. This shows the system is able to track a continuously moving target at velocity 4 cm/s over at least 12 cm distance, which is reasonable considering the space available inside the thorax. In comparison, a cadaveric study with expert surgeons found mean instrument velocities to be 2.2–2.6 cm/s in confined anatomical region [17].

E. Lung Resection in Simulated Thoracic Cavity

To test the feasibility of the endoscope system in settings similar to VATS, we also performed experiments inside a simulated thoracic cavity. Using only the magnetic endoscope as visual guidance, we completed a wedge resection task on a porcine lung. The simulated thoracic cavity consists of a rib cage model with silicone layer and foam padding to mimic flesh. To test generalizing of our detection method, the hand-held instruments used are not in the training dataset images. Under these settings, the deep learning based method is effective in detecting the surgical instruments, and the visual servoing control automatically keeps the target in view, allowing for an intuitive operation. Fig. 10(a) shows the test setup and Fig. 10(b) shows the inside view of the simulated thoracic cavity, where the endoscope magnetically anchored and the lung sample located beneath. Fig. 10 (c)–(f) present selected frames from the camera view captured during the procedure. The task was completed in 52 s, with successful resection of a tissue sample about 40 mm $\times$ 15 mm, shown in Fig. 10 (g). The results demonstrate preliminary success of the endoscope system in guiding VATS completion. Magnetic coupling between the EPMs and the endoscope remains stable during all tests, loss of anchoring did not occur.

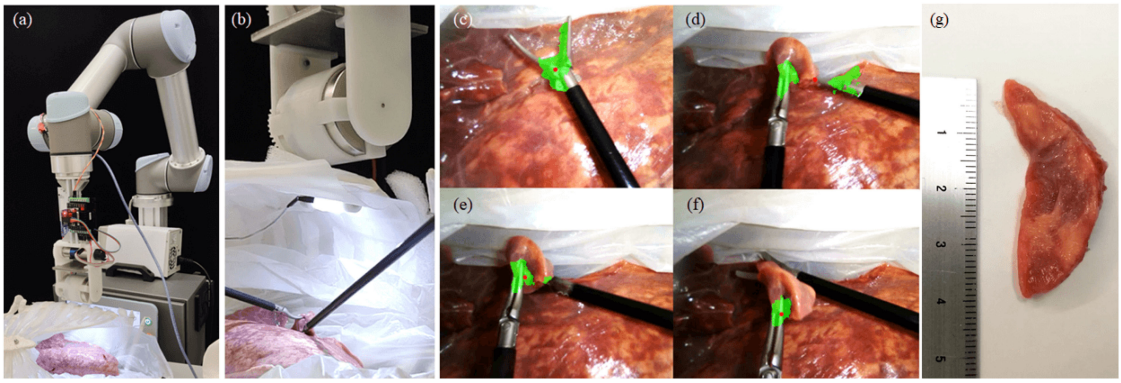

Fig. 10. Mock-up wedge resection on porcine lung. (a) Experimental setup. (b) Closeup inside the simulated thoracic cavity. (c–f) Snapshots of camera views. (g) Resection specimen.

VI. Conclusions

In this letter, we realized autonomous tracking of marker-less instrument using a novel magnetic anchored and actuated endoscope. The system consists of the magnetic anchored endoscope, and the external controller. The endoscope is a compact 5×1 cm capsule with two IPMs, two on board LEDs for independent illumination, and lens cleaning capability. The external controller is made of a UR5 robot arm and a customized end effector holding the EPM. Kinematics analysis was performed on the system. This offers the relationship between target pixels in the camera view, and the joint states of the robotic controller. Visual servoing control is implemented accordingly. A contribution of this work is the first case of employing deep learning methods for real-time instrument detection on a magnetic surgical endoscope. We accomplished this by using a modified TernausNet-16 network. In the experiments, we demonstrated this method is effective in tracking surgical instrument, even with blood contamination. We also showed the system can track a continuously moving target at 40 mm/s over at least 12 cm distance. This is compatible with VATS considering thoracic spatial constraints. We successfully performed a wedge resection on porcine lung inside a simulated thoracic cavity. This serves as preliminary evidence of the system's feasibility in clinical settings. Considering lighting conditions can affect performance, we also test the system in black box with porcine organs to mimic realistic surgical scene. The instrument segmentation has IoU 51.17%, and 50.45% after rinsing of contaminated lens. These demonstrate effectiveness of the illumination and lens cleaning functions of the endoscope. The segmentation IoU may be higher in deeper network, at the cost of more processing time. With real-time tracking of un-modified instruments, and initial evidence of feasibility, we believe that magnetic anchored endoscope will become a valuable tool in single-port thoracic surgery. Future improvements of the system may include reducing the effects of wire tether, and implementation of sensors for spatial feedback to reduce risk of position errors.

References

[1] J. Cadeddu, R. Fernandez, M. Desai, R. Bergs, C. Tracy, and S. J. Tang, "Novel magnetically guided intra-abdominal camera to facilitate laparoendoscopic single-site surgery: Initial human experience," *Surg. Endosc.*, vol. 23, no. 8, pp. 1894–1899, 2009.

[2] M. Simi, R. Pickens, A. Menciassi, S. D. Herrell, and P. Valdastri, "Fine tilt tuning of a laparoscopic camera by local magnetic actuation: Twoport nephrectomy experience on human cadavers," *Surg. Innov.*, vol. 20, pp. 385–394, 2013.

[3] G. Tortora and A. Dario, P.and Menciassi, "Array of robots augmenting the kinematics of endocavitary surgery," *IEEE/ASME Trans. Mech.*, vol. 19, no. 6, pp. 1821–1829, Dec. 2014.

[4] N. Garbin, P. R. Slawinski, G. Aiello, C. Karraz, and P. Valdastri, "Laparoscopic camera based on an orthogonal magnet arrangement," *IEEE Robot. Autom. Lett.*, vol. 1, no. 2, pp. 924–929, Jul. 2016.

[5] X. Liu, G. J. Mancini, Y. Guan, and J. Tan, "Design of a magnetic actuated fully insertable robotic camera system for single-incision laparoscopicsurgery," *IEEE/ASME Trans. Mech.*, vol. 21, no. 4, pp. 1966–1976, Aug. 2016.

[6] T. Cheng, W. Li, C. S. H. Ng, P. W. Y. Chiu, and Z. Li, "Visual servo control of a novel magnetic actuated endoscope for uniportal video-assisted thoracic surgery," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 3098–3105, Jul. 2019.

[7] X. Liu, R. Y. Abdolmalaki, T. Zuo, Y. Guan, G. J. Mancini, and J. Tan, "Transformable in vivo robotic laparoscopic camera with optimized illumination system for single-port access surgery: Initial prototype," *IEEE/ASME Trans. Mech.*, vol. 23, no. 4, pp. 1585–1596, Aug. 2018.

[8] A. Reiter, P. Allen, and T. Zhao, "Feature classification for tracking articulated surgical tools," in Proc. Int. *Conf. Med. Image. Comput. Comput. Assist. Interv.*, 2012, pp. 592–600.

[9] X. Du *et al.*, "Combined 2d and 3d tracking of surgical instruments for minimally invasive and robotic-assisted surgery," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 6, pp. 1109–1119, 2016.

[10] Z. Zhao, Y. Voros, S. Weng, F. Chang, and R. Li, "Tracking-by-detection of surgical instruments in minimally invasive surgery via the convolutional neural network deep learning-based method," *Comput. Assist. Surg.*, vol. 22, pp. 26–35, 2017.

[11] P. Corke, *Robotics, Vision and Control: Fundamental Algorithms in MATLAB*. Berlin, Heidelberg, Germany: Springer-Verlag, 2011.

[12] B. Siciliano and O. Khatib, *Springer Handbook of Robotics*. Berlin, Heidelberg, Germany: Springer, 2008.

[13] T. Hu, T. Wang, J. Li, and W. Qian, "Gradient projection of weighted jacobian matrix method for inverse kinematics of a space robot with a controlled-floating base," *J. Dyn. Syst., Meas., Control*, vol. 139, no. 5, Mar. 2017, Art. no. 051013.

[14] A. A. Shvets, A. Rakhlin, A. A. Kalinin, and V. I. Iglovikov, "Automatic instrument segmentation in robot-assisted surgery using deep learning," in *Proc. IEEE Int. Conf. Mach. Learn. Appl.*, 2018, pp. 624–628.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image. Comput. Comput. Assist. Interv.*, 2015, pp. 234–241.

[16] K. Simonyan and A. Zisserma, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Rep.*, 2014.

[17] R. A. Harbison, A. M. Berens, Y. Li, R. A. Bly, B. Hannaford, and K. S. Moe, "Region-specific objective signatures of endoscopic surgical instrument motion: A. cadaveric exploratory analysis," *J. Neuro. Surg.*, vol. 78, no. 1, pp. 99–104, 2017.