

A Learning-Driven Framework with Spatial Optimization For Surgical Suture Thread Reconstruction and Autonomous Grasping Under Multiple Topologies and Environmental Noises

Bo Lu^{1,*}, Wei Chen¹, Yue-Ming Jin², Dandan Zhang³, Qi Dou², Member, IEEE, Henry K. Chu⁴, Member, IEEE, Pheng-Ann Heng², Senior Member, IEEE, and Yun-Hui Liu¹, Fellow, IEEE

Abstract—Surgical knot tying is one of the most fundamental and important procedures in surgery, and a high-quality knot can significantly benefit the postoperative recovery of the patient. However, a longtime operation may easily cause fatigue to surgeons, especially during the tedious wound closure task. In this paper, we present a vision-based method to automate the suture thread grasping, which is a sub-task in surgical knot tying and an intermediate step between the stitching and looping manipulations. To achieve this goal, the acquisition of a suture's three-dimensional (3D) information is critical. Towards this objective, we adopt a transfer-learning strategy first to fine-tune a pre-trained model by learning the information from large legacy surgical data and images obtained by the on-site equipment. Thus, a robust suture segmentation can be achieved regardless of inherent environment noises. We further leverage a searching strategy with termination policies for a suture's sequence inference based on the analysis of multiple topologies. Exact results of the pixel-level sequence along a suture can be obtained, and they can be further applied for a 3D shape reconstruction using our optimized shortest path approach. The grasping point considering the suturing criterion can be ultimately acquired. Experiments regarding the suture 2D segmentation and ordering sequence inference under environmental noises were extensively evaluated. Results related to the automated grasping operation were demonstrated by simulations in V-REP and by robot experiments using Universal Robot (UR) together with the da Vinci Research Kit (dVRK) adopting our learning-driven framework.

I. INTRODUCTION

R AISES of modern techniques in the last few decades prompt a rapid development and iteration of robotic system [1]. With the aids of sophisticated designs [2], advanced materials [3], and computer-integrated technologies [4], robot-assisted surgery (RAS) came into use in response to requirements of higher dexterity, accuracy and efficacy in clinical operations [5], such as in laparoscopic interventions of prostatectomy and hysterectomy in the abdominal cavity

This work is supported in part of the HK RGC under T42-409/18-R and 14202918, in part by project 4750352 of the CUHK-SJTU Joint Research Fund, and in part by the VC Fund 4930745 of the T Stone Robotics Institute.

¹B. Lu, W. Chen and Y. H. Liu are with CUHK T stone Robotics Institute, The Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong. Y. H. Liu is the Director of the T Stone Robotics Institute. *Corresponding author: Bo Lu (email: bolu@cuhk.edu.hk).

²Y. M. Jin, Q. Dou and P. A. Heng are with CUHK T stone Robotics Institute, The Department of Computer Science and Engineering, The Chinese University of Hong Kong.

³D. Zhang is with The Hamlyn Centre for Robotic Surgery, Imperial College London, UK.

⁴H. K. Chu is with The Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong.

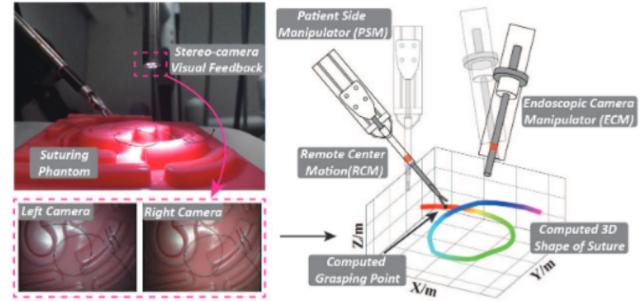


Fig. 1. With calibrated stereo visions, 3D coordinates and the grasping point of an arbitrary orientated suture can be obtained using our framework, thereby automating a vision-based suture grasping manipulation.

[6]. Nowadays, the prosperous growths of medical robots not only benefit practical operations of surgeons, but also alter their manipulating paradigms and significantly extend modern surgical capabilities.

Surgical suturing/knot tying for cuticle wound closure [7] is a fundamental and important task, yet it remains a challenging operation in minimally invasive surgery (MIS), such as in endoscopic vascular anastomosis and transanal total mesorectal excision (taTME) for low rectal cancer [8]. Rather than conducting it manually, this task can now be operated remotely and ergonomically via a 3-D visualization and EndoWrist equipment [9] using the cutting-edge master-slave surgical robot (e.g. da Vinci robotic system). However, a longtime surgery requires intensive attentions and easily brings fatigue to surgeons. This may result in low-quality wound closure or even lead to extra trauma to patient. To release burdens from surgeons, immense efforts have been leveraged to automate each standard step in surgical knot tying task [7][10], promoting the development of vision-based automated robotic surgery.

To initialize the surgical knot tying, extensive investigations have been put on automated stitching/piercing tasks, covering the research areas of collaborative hardware developments for robotic suturing [11][12][13], semantic segmentation of surgical tools [14], task planning and execution via learn-by-demonstrations [15] and supervised learning [16], interaction analysis between rigid instrument and soft tissue [17][18]. All these pioneering studies provided potential solutions to knotty challenges in automated suturing task [19]. However, the wounds were all anastomosed by continuous sutures in [11][13][16], and it may result in the loose of

the entire suturing if one stitch breaks, thereby harming the rehabilitation of patient.

To overcome this problem, we adopt the interrupted suturing technique, in which suture knots should be generated sequentially along a surgical wound. Based on pre-mentioned works, suture can be automatically stitched through soft tissues, setting a solid foundation for the coming knot tying task. In [7], an optimal suture's looping trajectory was generated and a vision-based method was proposed for the automated knot tying task. Besides, a robot can autonomously learn an online trajectory for a suture's looping task, and further, a force controller was integrated into the whole system to dynamically monitor motions of robot [20]. However, the thread should be cut off when using the interrupted suturing [21], and thereby, the grasping of a cutting thread is the key to realizing a fully automated robotic knot tying task.

To address the critical issues mentioned above, researchers put their attentions on suture segmentation, topological analysis, 3D reconstruction and relevant manipulations. Padoy et al. [22] presented a method for tracking deformable 3D suture based on discrete Markov random field optimization and a non-uniform rational B-spline (NURBS). Later on, Jackson et al. [23] adopted the NURBS curve to obtain 3D coordinates of a suture by minimizing its image matching energy in a stereo-camera system. Besides, Gu et al. [24] proposed a joint feature learning framework for suture's detection by semi-supervised and unsupervised learning domain adaptation. In addition, Lui [25] used particle filter and learning algorithm to infer and untangle a rope's knot with the data acquired from an RGB-D camera. However, these existing works only focused on 2D image semantic segmentation [24], or didn't examine the suture's reconstruction with multiple shapes under complex environments [22][23]. The relevant manipulations are only limited to macro-size cables [25][26] rather than a thin and deformable surgical suture.

Regarding these challenges, we propose a learning-driven framework for automated suture grasping task based on vision feedback from a calibrated stereo camera system, which is as shown in Fig.1. Our main contributions are: 1). To improve the segmentation performance in surgical environments, we combine images from da Vinci robotic surgery with synthetic sutures, and build up our data set to train an offline model, which learns the general information of a surgical suture. Using a few online images, our model can quickly adapt to the present condition and achieve a higher performance for a segmentation task; 2). To compute 3D coordinates of a suture with arbitrary orientations, we developed a novel cost function and termination policies based on our previous method in [21][27]. Using this upgraded approach, the ordering sequence of a suture (e.g. with self intersections or external crossings) can be completely obtained two camera frames; 3). We further construct a dense 3D vertex graph by calculating 3D positions of all pixel pairs along the suture, and optimize its spatial shape using an optimized shortest path approach. Consequently, the desired grasping point can be located based on the suturing criteria

[28], proceeding an automated grasping operation.

The rest of this article is organized as follows: Section. II introduces the learning-driven workflow for a suture's grasping point computation. Section. III presents the experimental validations concerning 2D segmentation and 3D grasping results. In Section. IV, we will conclude the entire paper.

II. METHODOLOGY

After cutting a suture during the interrupted suturing manipulation, the thread may arbitrarily drop above the surrounding tissue surface due to its nature of flexibility, and its precise 3D coordinates computation is essential for the following vision-based grasping task. To achieve this goal, we decomposed the overall procedure into 3 sub-tasks:

- 1): Suture segmentation from complex environments;
- 2): Ordering sequence analysis in stereo images;
- 3): 3D shape computation with optimization, and grasping point generation.

A. Transferred Model for Online Suture Segmentation

To segment a suture under different conditions, traditional imaging processing methods can hardly obtain satisfactory results under variant image scales, background noises, and ambient illuminations. To achieve a higher performance of segmentation to guide the grasping task, a deep learning based model using transferred knowledge is proposed.

The dominant objective of this model is to highlight a suture's curvilinear information, while removing surrounding noises. U-Net is reported as a fast and precise model for image segmentation task, and it shows outstanding outcomes of intricate neuronal structure segmentation in electron microscopic stacks [29]. Here, U-Net model is migrated to our workflow as the backbone. The main training structure is shown in Fig. 2. It is worth noticing that the model's first 11 layers are extended to 13 layers, constructing our feature extractor. By this modification, we can explicitly extract features within the image. We utilize the cross entropy as the loss function and froze the learned weights in the feature extractor to enhance the entire training efficiency.

1) Data Generation and Offline Training: To put against surgery noises, a large number of endoscopic images acquired from da Vinci robotic surgery are collected. Since limited images are obtained during surgical suturing, and to make the most of legacy surgical data, synthetic sutures are generated in these images to simulate suturing scenarios.

To ensure the diversity, sutures with different colors, scales and shapes are randomly generated, getting offline synthetic images I_{OFF} . At the meanwhile, corresponding masks M_{OFF} can be assessed, and we can establish our expert database concerning surgical suture. As shown in the left part of Fig. 2, such information is imported to the training workflow, and a model f_s , which learns the general information of a suture for surgical suturing, can be obtained.

2) Data Transfer and Online Prediction: Barely using the expert model f_s for online segmentation can merely obtain a satisfactory result due to equipment and scenario differences. Even for dVRK systems, endoscopes may output

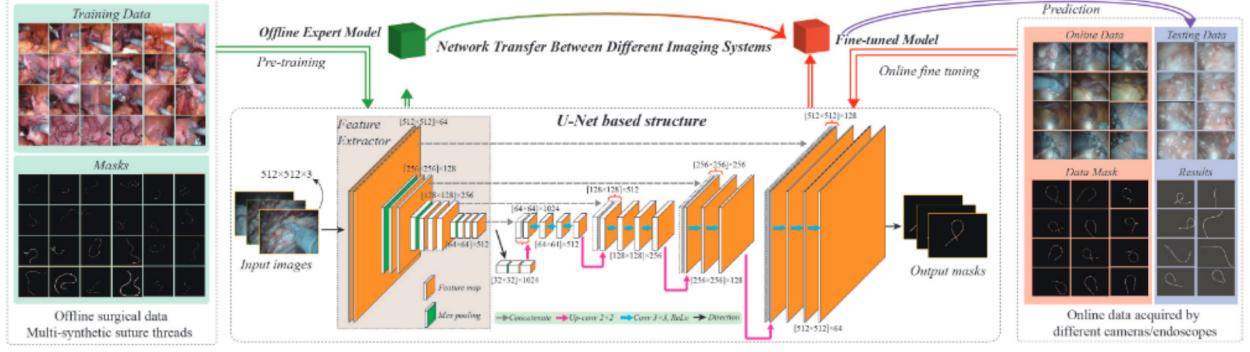


Fig. 2. The training workflow for a suture segmentation in surgical environments. An offline data set containing legacy surgical data and synthetic sutures can be established, and it can be used to pre-train an expert model f_s which contains the general information of a suture. It can be expediently transferred to a new online scenario using a small amount of target domain data.

images with variant qualities due to discrepancies caused by customized settings and long-time wear. Thus, adapting to variances between different systems is a challenging task for our pre-trained model.

To address this problem, we employ a transfer learning method in our workflow. To build a refined model \tilde{f}_s , we transfer the learned weights in f_s , then we froze the previous 13 layers in our feature extractor and fine-tune the remaining layers. By using limited online annotated samples \mathbf{I}_{ON} and \mathbf{M}_{ION} from an on-site equipment, \tilde{f}_s can be efficiently achieved for an online image prediction, which is illustrated in Fig. 2. It is defined that the online image is denoted as \mathbf{I}_v and its annotated mask is $\mathbf{M}_{\mathbf{I}_v}$, the fine-tuned model \tilde{f}_s can obtain premium segmentation results for stereo cameras, and the results are expressed by $\tilde{\mathbf{Se}}_l, \tilde{\mathbf{Se}}_r \in \mathbb{R}^2$. Validations with assessment metrics will be investigated in Section. III-A.

B. Inference of Suture's Ordering Sequence

Using $\tilde{\mathbf{Se}}_l$ and $\tilde{\mathbf{Se}}_r$, all contours in images can be picked out $\{\mathbf{S}_1^l, \mathbf{S}_2^l, \dots, \mathbf{S}_i^l, \dots\}, \{\mathbf{S}_1^r, \mathbf{S}_2^r, \dots, \mathbf{S}_j^r, \dots\}$. With manual indication in the left frame, we can implement the method in our previous paper [21] to locate the cutting tip and obtain edge elements of the desired suture. Afterwards, these elements can be moved to the shape's centreline, which can be denoted as $\hat{\mathbf{S}}_L$ and $\hat{\mathbf{S}}_R$.

To accurately align the stereo pairs, the inference of a pixel-level sequence of a suture under multi-topologies is particularly important, since a small change in suture may result in different topologies as shown in Fig. 3(a). Towards this goal, we designed a strategy which searches from the cutting tip and can be terminated automatically when the whole suture is traversed. For a 2D point $V^t \in \hat{\mathbf{S}}$ in the t^{th} searching loop¹, we categorize all elements in $\hat{\mathbf{S}}$ as *Detected Nodes* \mathbf{S}^t , *Active Nodes* \mathbf{A}^t , and *Far Nodes* \mathbf{F}^t . The connecting pixel is elected from \mathbf{A}^t , which is defined as:

$$\begin{aligned} \mathbf{A}^t &= \{A_1^t, A_2^t, \dots, A_j^t, \dots\} \in \mathbb{R}^2 \\ \text{s.t. } \forall A_j^t \in \mathbf{A}^t : |A_j^t - V^t| &\leq \Upsilon \quad \& \quad A_j^t \in \tilde{\mathbf{Se}}, A_j^t \notin \mathbf{S}^t \end{aligned} \quad (1)$$

¹For a concise expression, we will omit the right subscript and only take the left frame for example in this section. In each searching loop, we can find a connecting point V^t .

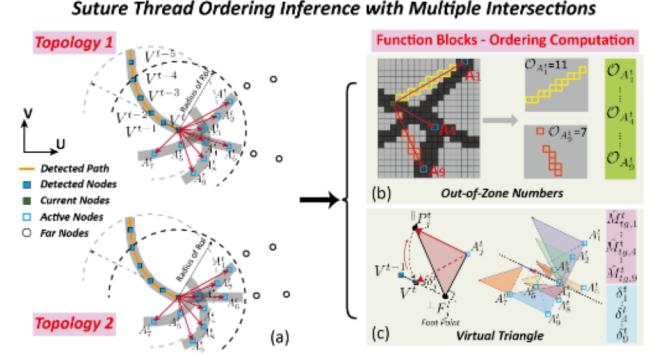


Fig. 3. The cost function of the searching strategy for the ordering sequence inference of a suture with different shapes and intersections. Using our algorithm, the connecting pixels are A_4^t and A_3^t accordingly in two topologies which have slight differences between each other.

where Υ denotes the radius of the Region of Interest (RoI), A_j^t is one active node. Principles of the searching strategy are illustrated in Fig. 3(b)(c) and described as follows:

Principle 1: We generate a line I_j^t by directly linking the candidate A_j^t and current pixel V^t as shown in Fig. 3.

Definition: For any pixel $p_{u_i, v_i} \in I_j^t$, if $p_{u_i, v_i} \notin \tilde{\mathbf{Se}}$, p_{u_i, v_i} is an out-of zone element. The total number of such pixels along I_j^t is the **Out-of-Zone Numbers** of candidate A_j^t , which is denoted as $\mathcal{O}_{A_j^t}$. For a potential candidate, its cost function regarding this term is designed following two criteria:

Criterion 1: The cost value should be dramatically improved according to the increase of $\mathcal{O}_{A_j^t}$ when it appears.

Criterion 2: If $\mathcal{O}_{A_j^t}$ exceeds a particular value, we will treat the candidate A_j^t as a false connection, thereby the cost function should be insensitive to the further increase of $\mathcal{O}_{A_j^t}$.

Taking both criteria into account, a logarithmic function is implemented to describe this principle.

Principle 2: For each candidate A_j^t , we prefer the one has shorter distance to the current point V^t . With the variation of $|A_j^t - V^t|$, the cost value changes following a linear manner.

Principle 3: For a continuous and smooth curvilinear object, the changing angle δ_j^t between two adjacent vectors forming by three successive vertices should be evaluated. For each A_j^t , we prefer the candidate generating a smaller value of δ_j^t , which ensures the smoothness along the whole shape.

Therefore, we adopt an exponential function by putting higher cost when there exists shape change, while setting it insensitive among the low value domain.

Implement: To characterize **Principle 2** and **3**, a virtual triangulate formed by V^t and each candidate is adopted. As shown in Fig. 3(c), V^{t-1} and V^t are two successive vertices. For candidate A_j^t , we have: point $\perp F_j^t \in V^{t-1}V^t$, $\perp F_j^i V^t \perp \perp F_j^i A_j^t$. Connecting A_j^t and V^t , we obtain line $A_j^t V^t$ and variation angle δ_j^t . By rotating this vector within acute angle until reaching the point $\parallel P_j^t$, which satisfies $\|P_j^t A_j^t\| = \|\perp F_j^i V^t\|$, a virtual triangle $\triangle \perp F_j^i \| P_j^t A_j^t$ can be generated, and its virtual area can be implemented as the partial term among the whole function, which is:

$$\begin{aligned} M_{tg,j}^t &= 0.5 \cdot \underbrace{|A_j^t - V^t|}_{\text{Side } A} \cdot \underbrace{(|A_j^t - V^t| \cdot \sin(\delta_j^t))}_{\text{Side } B} \\ \text{s.t. } \delta_j^t &= \arccos \frac{(A_j^t - V^t) \cdot (V^t - V^{t-1})}{|A_j^t - V^t| \cdot |V^t - V^{t-1}|}, t > 1 \subseteq \mathbb{Z} \end{aligned} \quad (2)$$

To represent a monotonically increasing property with respect to $\delta_j^t \in [0, \pi]$, the term $\sin(\delta_j^t)$ in virtual *Side B* controlling the changing direction can be optimized as: $\sin(\delta_j^t/2)$, and we further adopt an exponential function as claimed in **Principle 3** to increase costs of sharp variations. Taking these factors into account, the optimized virtual area $\hat{M}_{tg,j}^t$ can be re-designed as:

$$\hat{M}_{tg,j}^t = 0.5 \cdot |A_j^t - V^t| \cdot e^{\sin(\delta_j^t/2)} \quad (3)$$

With the term of **Principle 1**, the composite cost function for each candidate at t^{th} searching loop can be derived as:

$$\begin{aligned} C_j^t &= \log(\mathcal{O}_{A_j^t} + \mu) + \hat{M}_{tg,j}^t \xrightarrow{Tune} C_j^t(A_j^t, \epsilon_1, \epsilon_2, \epsilon_3) \\ &= \log(\epsilon_1 \cdot \mathcal{O}_{A_j^t} + \mu) + \epsilon_2 \cdot |A_j^t - V^t| \cdot e^{\epsilon_3 \cdot \sin(\frac{\delta_j^t}{2})} \end{aligned} \quad (4)$$

where μ is a small positive gain, and $\{\epsilon_1, \epsilon_2, \epsilon_3\}$ are tuning parameters that determines the influences of three principles in the function. During each searching loop, the connecting pixel can be achieved by:

$$\forall A_j^t \in \mathbb{A}^t : V^{t+1} = \arg \min_{A_j^t} \{C_j^t(A_j^t, \epsilon_1, \epsilon_2, \epsilon_3)\} \quad (5)$$

The inferred ordering sequence of a suture can be thereby obtained as $\mathbb{S} = \{V^1, V^2, \dots, V^k, \dots\}$. To enable a robust scheme for different cases, three tuning parameters can be autonomously adjusted within their given ranges to maximize the element number in \mathbb{S} . Besides, termination policies are also proposed, and taking Eq. (5) into account, tuning parameters can be selected as:

$$\begin{aligned} \forall \{\epsilon_1 \in \mathbf{R}_1, \epsilon_2 \in \mathbf{R}_2, \epsilon_3 \in \mathbf{R}_3\} : \epsilon_1, \epsilon_2, \epsilon_3 &= \arg \max \{|\mathbb{S}|\} \\ \text{s.t. } \mathbb{A}^t \neq \emptyset : \mathcal{O}_{A_j^t} &< \tau_O \text{ & } \delta_j^t < \tau_V \end{aligned} \quad (6)$$

where $\mathbf{R}_1 \in [1, 5], \mathbf{R}_2 \in [0.1, 0.5], \mathbf{R}_3 \in [0.02, 0.1]$ denote values' ranges, and $\tau_O = 2.6\Upsilon, \tau_V = 2\pi/3$ are the terminating thresholds in our research. Following the parallel optical axis assumption, we further apply this algorithm to both frames to get stereo pairs $[\mathbb{S}_l \leftrightarrow \mathbb{S}_r]$ and 3D dense

points, i.e. vertices $\mathbf{V}^{1 \leftrightarrow \omega_1} = \{\mathbf{V}^1, \mathbf{V}^2, \dots, \mathbf{V}^{\omega_1}\} \in \mathbb{R}^3$ along the suture using the method in our previous article [27].

C. Online Optimization of 3D Shape

Due to limited resolution of camera/endoscope, there may exist zigzag routes by connecting the computed dense 3D points one by one, thus forming undesired configurations. Besides, we also need to compensate for unexpected errors resulting from noises in set $\mathbf{V}^{1 \leftrightarrow \omega_1}$ in order to obtain precise 3D representative vertices $\hat{\mathbf{V}}^{1 \leftrightarrow \omega_2} \subseteq \mathbf{V}^{1 \leftrightarrow \omega_1}$ of the suture.

To realize this objective, the approach based on Dijkstra's shortest path theory [30] is developed to optimize the 3D shape between the tip and end pints. For a self-intersected suture as shown in Fig. 4(a) and (b), the traditional method gives a result of minimal physical length between two ends, which turns to the distal vertice at the crossing space. However, such results indicate a fragmentary outcome of the suture, which may jeopardize the grasping point computation.

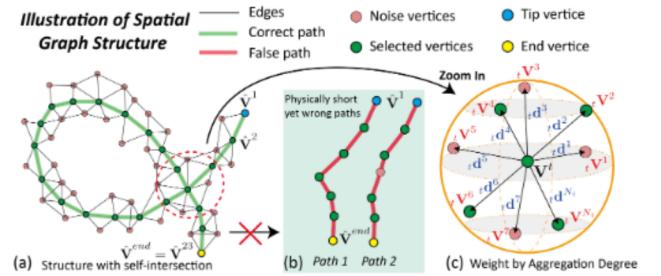


Fig. 4. Spatial shape optimization of a suture with self-intersections based on a shortest path method with optimized and composite edges' weights.

To characterize the overall information of suture, we first construct the 3D edge structure using all vertices in $\mathbf{V}^{1 \leftrightarrow \omega_1}$, and remove those edges whose lengths are larger than τ_L :

$$\forall \{\mathbf{V}^i, \mathbf{V}^t\} \in \mathbf{V}^{1 \leftrightarrow \omega_1} : D^{i \leftrightarrow t} = \begin{cases} |\mathbf{V}^i - \mathbf{V}^t|, & |\mathbf{V}^i - \mathbf{V}^t| \leq \tau_L \\ \infty, & |\mathbf{V}^i - \mathbf{V}^t| > \tau_L \end{cases} \quad (7)$$

where $D^{i \leftrightarrow t}$ denote physical length of Edge $E^{i \leftrightarrow t}$ between vertice \mathbf{V}^i and \mathbf{V}^t . By treating diagonal elements as ∞ , the complete 3D length matrix \mathbf{D}_G can be formulated as:

$$\mathbf{D}_G = \begin{bmatrix} \infty, D^{1 \leftrightarrow 2}, D^{1 \leftrightarrow 3}, \dots, D^{1 \leftrightarrow \omega_1} \\ D^{2 \leftrightarrow 1}, \infty, D^{2 \leftrightarrow 3}, \dots, D^{2 \leftrightarrow \omega_1} \\ \dots, \dots, D^{i \leftrightarrow i} = \infty, \dots, D^{i \leftrightarrow \omega_1} \\ D^{\omega_1 \leftrightarrow 1}, \dots, \dots, D^{\omega_1 \leftrightarrow \omega_1 - 1}, \infty \end{bmatrix}_{\omega_1 \times \omega_1} \quad (8)$$

To further enhance the representative capability of the graph structure, we design a composite weight matrix \mathbf{W}_G , which denotes the penalty for each $E^{i \leftrightarrow t}$ by considering the vertice density around it two ends \mathbf{V}^i and \mathbf{V}^t , as well as evaluating their sequential difference.

We find out the valid vertice $t\mathbf{V}^j$ within a virtual sphere, which is centered by the point \mathbf{V}^t with a radius Υ_3 as shown in Fig. 4(c). Counting the total number of $t\mathbf{V}^j$ within this space, we can formulate a penalty following the right part of

the Gaussian curve, which determines the clustering level of computed points around $t\mathbf{V}^j$. Apart from this, we measure all distances between surround points to \mathbf{V}^t . The total penalty for vertice \mathbf{V}^t can be designed as:

$$\begin{aligned} P_1^t &= \underbrace{\frac{1}{\sqrt{2\pi}\sigma_1} \cdot \exp\left(-\frac{N_i}{2\sigma_1}\right)}_{\text{Effect of Number}} \underbrace{\sum_j^{N_i} \frac{S_f}{N_i} (\log(|t\mathbf{V}^j - \mathbf{V}^t| + \nu_1))}_{\text{Effect of Distance}} \quad (9) \\ &= \frac{S_f}{N_i} \cdot \exp\left(-\frac{N_i}{2\sigma_1}\right) \cdot \sum_j^{N_i} (\log(|t\mathbf{V}^j - \mathbf{V}^t| + \nu_1)) \end{aligned}$$

where S_f denotes a scaling factor, σ_1 determines the steepness of the Gaussian curve, and ν_1 is a positive constant. Multiplying these values of two points \mathbf{V}^i and \mathbf{V}^j , the first term of the penalty for Edge $E^{i \leftrightarrow t}$ forming by these two vertices can be obtained as:

$$P_1^{i \leftrightarrow t} = P_1^t \cdot P_1^i \quad (10)$$

Based on the results in Section II-B, the ordering differences between two ends vertices of Edge $E^{i \leftrightarrow t}$ should be evaluated, which aims to avoid unexpected turns around crossing nodes and get a complete suture's shape. Thus, the second term of the penalty for Edge $E^{i \leftrightarrow t}$ is designed as:

$$P_2^{i \leftrightarrow t} = 1 + \exp\left(\frac{|i - t|}{20} - 2\right) \quad (11)$$

Combining $P_1^{i \leftrightarrow t}$ and $P_2^{i \leftrightarrow t}$, the composite weight for Edge $E^{i \leftrightarrow t}$ in the graph structure can be expressed by:

$$\forall \mathbf{V}^i, \mathbf{V}^t \in \mathbf{V}^{1 \leftrightarrow \omega_1} : W^{i \leftrightarrow t} = P_1^{i \leftrightarrow t} \cdot P_2^{i \leftrightarrow t} \quad (12)$$

The overall weight matrix \mathbf{W}_G is hence formulated as:

$$\mathbf{W}_G = \begin{bmatrix} 1, W^{1 \leftrightarrow 2}, W^{1 \leftrightarrow 3}, \dots, W^{1 \leftrightarrow \omega_1} \\ W^{2 \leftrightarrow 1}, 1, W^{2 \leftrightarrow 3}, \dots, W^{2 \leftrightarrow \omega_1} \\ \dots, \dots, W^{j \leftrightarrow j} = 1, \dots, \dots \\ W^{\omega_1 \leftrightarrow 1}, \dots, \dots, W^{\omega_1 \leftrightarrow \omega_1 - 1}, 1 \end{bmatrix}_{\omega_1 \times \omega_1} \quad (13)$$

We integrate \mathbf{D}_G and \mathbf{W}_G , and the optimized graph matrix can be attained using the Hadamard product, which is:

$$\mathbf{G}(\mathbf{E}, \mathbf{V}) = \mathbf{D}_G \odot \mathbf{W}_G \quad (14)$$

Putting $\mathbf{G}(\mathbf{E}, \mathbf{V})$ into Dijkstra's shortest path method, the shortest path containing the updated vertices $\hat{\mathbf{V}}^{1 \leftrightarrow \omega_2} = \{\hat{\mathbf{V}}^1, \hat{\mathbf{V}}^2, \dots, \hat{\mathbf{V}}^{\omega_2}\}$ can be obtained. The pseudocode of this optimization method is listed in Algorithm. 1.

Given the entire 3D information of a suture, the grasping point $\mathbf{GS}_C = [GX_C, GY_C, GZ_C, 1]^T$ with respect to the camera coordinate can be calculated by reserving a sufficient length for the looping manipulation, thereby proceeding the automated grasping with the Euclidean target in the task space produced by $\mathbf{GS}_R = {}_C\mathbf{T}^R \cdot \mathbf{GS}_C$, in which ${}_C\mathbf{T}^R \in \mathbf{SE}(3)$ is the transformation matrix from camera coordinate to robot frame obtained by a standard hand to eye calibration approach [31].

Algorithm 1: Spatial shape optimization of a suture

Data: 3D vertices $\mathbf{V}^{1 \leftrightarrow \omega_1}$, the radius Υ_3 of a virtual sphere, parameter τ_L in Eq. (7).

```

1 for  $\mathbf{V}^t \in \mathbf{V}^{1 \leftrightarrow \omega_1}$  find do
2    $\forall \mathbf{V}^i \in \mathbf{V}^{1 \leftrightarrow \omega_1}$  find:  $|\mathbf{V}^i - \mathbf{V}^t| \leq \Upsilon_3$  → Denoted as:  $\{\mathbf{V}^1, \mathbf{V}^2, \dots, \mathbf{V}^{N_i}\}$ ;
3   Computed  $\{\mathbf{d}^1, \mathbf{d}^2, \dots, \mathbf{d}^{N_i}\}$  w.r.t  $\mathbf{V}^t$  (Fig. 4);
4   Computed  $P_1^t \leftarrow$  Eq. (9);
5 end
6  $\forall \{\mathbf{V}^t, \mathbf{V}^i\} \in \mathbf{V}^{1 \leftrightarrow \omega_1}, t \neq i \rightarrow$  Compute their edge penalty  $P_1^{i \leftrightarrow t}, P_2^{i \leftrightarrow t}$  using Eq. (10) (11);
7 Calculate  $\mathbf{W}^{i \leftrightarrow t} \rightarrow$  Form weight matrix  $\mathbf{W}_G$ ;
8 while (True) do
9   Generate distance matrix ( $\mathbf{D}_G$ ) $_{\omega_1 \times \omega_1}$ ;
10   $\forall \mathbf{D}^{i \leftrightarrow t} \in \mathbf{D}_G \rightarrow D^{i \leftrightarrow t} = \infty$ , when  $|\mathbf{V}^t - \mathbf{V}^i| > \tau_L$  or  $i = t$ ;
11   $\mathbf{G}(\mathbf{E}, \mathbf{V}) = \mathbf{D}_G \odot \mathbf{W}_G \xrightarrow{\text{Dijkstra}}$  Shortest path  $L_{min}$  and  $\hat{\mathbf{V}}^{1 \leftrightarrow \omega_2} = \{\hat{\mathbf{V}}^1, \hat{\mathbf{V}}^2, \dots, \hat{\mathbf{V}}^{\omega_2}\}$  of the suture;
12  if  $L_{min} \neq \infty$  then
13    | break;
14  end
15   $\tau_L = \tau_L + 0.5 \leftarrow$  In case there is no path between two ends, we construct more connecting edges in the graph.
16 end
Result: Updated vertices  $\hat{\mathbf{V}}^{1 \leftrightarrow \omega_2} = \{\hat{\mathbf{V}}^1, \hat{\mathbf{V}}^2, \dots, \hat{\mathbf{V}}^{\omega_2}\}$  and minimal length  $L_{min}$  of the suture.

```

III. EXPERIMENTAL RESULTS

In this section, we extensively validated the performance of the deep learning-based suture segmentation and the feasibility of the sequence inference algorithm. Based on the stereo visual perception, experiments related to the 3D suture reconstruction and automated grasping were simulated in V-REP using the dVRK model introduced by [32]. The overall framework were tested by a UR robot and dVRK system.

A. Performance Assessment of Suture Segmentation

Our learning model was constructed based on Keras framework with a Tensorflow backend, and it was run on a Nvidia Titan 1080 GPU. Using the U-Net based structure with a transfer learning strategy, we investigated and compared the suture segmentation performance of three models, which are: Models from (1). offline synthetic images training (*M-oST*); (2). training with online target domain images (*M-TOT*); (3). transfer learning from offline synthetic data to online target domain ((*M-ToSOT*)). The metrics for the segmentation performance assessment are Intersection over Union (IoU), Precision, Recall, and F1 Score, which are defined as:

$$IoU = \frac{|\tilde{\mathbf{Se}} \cap \mathbf{I}_V|}{|\tilde{\mathbf{Se}}| + |\mathbf{I}_V| - |\tilde{\mathbf{Se}} \cap \mathbf{I}_V|}, F1 = \frac{2 \cdot Pre \cdot Rec}{Pre + Rec} \quad (15)$$

where *Pre* and *Rec* denote the segmentation precision and recall, which are accordingly expressed as $|\tilde{\mathbf{Se}} \cap \mathbf{I}_V| / |\tilde{\mathbf{Se}}|$

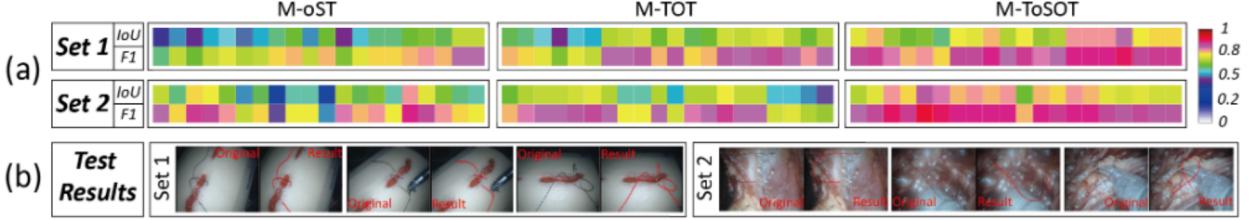


Fig. 5. (a). Segmentation results regarding IoU and F1 scores of 3 different models based on 40 test images obtained by two dVRK systems; (b). Typical segmentation results. Surgical phantoms were adopted in Set 1, and color-printed surgical scenes were used in Set 2 to simulated complex scenarios.

and $|\tilde{Se} \cap^M I_v| / |M_i_v|$.

We built up the offline model using 200 groups of data (the number can be further extended to increase the model's generality), and 40 target domain images were used to fine-tune the model. Experiments were tested on 40 images captured from 2 different dVRK endoscopic camera manipulators. To add variances to images, the target domain was a phantom-based suturing scenario in Set 1, and surgical scenes were color printed and treated as backgrounds in Set 2. The quantitative results are listed in Table. I. As noticed, *M-sOT* is unstable in the target domain testing. Its high precision with small IoU and recall tell that it only segments scattered pieces in the image. By training directly with limited online data from target domain, *M-TOT* is inferior to *M-ToSOT* with respects to all assessment metrics.

TABLE I

COMPARISONS OF THE LEARNING BASED SUTURE SEGMENTATION PERFORMANCE USING THREE MODELS.

Model		<i>M-oST</i>	<i>M-TOT</i>	<i>M-ToSOT</i>
Set 1	<i>IoU</i>	ave	55.8%	62.9%
	<i>Pre</i>	ave	84.6%	63.2%
	<i>Rec</i>	ave	61.8%	97.3%
Set 2	<i>IoU</i>	ave	59.8%	60.8%
	<i>Pre</i>	ave	79.9%	62.3%
	<i>Rec</i>	ave	71.0%	97.5%

In Fig. 5(a), we listed qualitative results of IoU and F1 Scores for each individual test. Using limited data without prior knowledge, *M-TOT* appears to be unstable across its overall test set. Regarding the numbers in Table. I, in which *M-ToSOT* achieves the best performance in all aspects, especially for the highest recall number 97.5%. Although it contains false positives in the tests, our following sequence inference algorithm can compensate for this shortcoming. Until now, the transfer learning strategy can be concluded workable for our suture segmentation task, and typical results of suture segmentation were shown in Fig. 5(b).

B. Validations of Vision-based Automated Grasping

Based on segmentation results, we will validate the feasibility of the sequence inference algorithm of suture, and further examine the performance of this vision-guided automated grasping manipulation in this section.

1). *Ordering Sequence Inference*: To begin the test, we collected image data by putting sutures on color-printed surgical scenes and orientating them with multiple shapes.

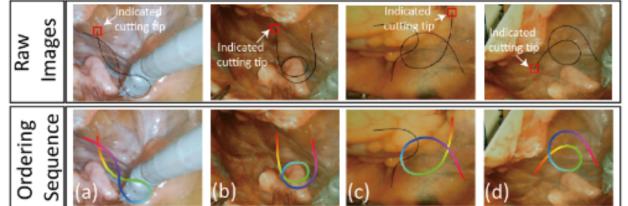


Fig. 6. Typical results of the ordering sequence inference of a suture under multiple topologies and variant conditions.

After sending the image to the fine-tuned model *M-ToSOT*, all suture-like objects *Se* will be segmented. To simplify the testing procedure, we selected suture's one end as the cutting tip, and indicated a ROI around it as illustrated in Fig. 6.

Employing the method in our previous paper [21], the precise tip in a stereo-camera could be found, and the inference can be thereby proceeded. As also shown in Fig. 6, four typical results concerning ordering inference were presented, and gradient colors represent the calculated sequence along the suture. It can be notice, the loss function based searching algorithm can efficiently figure out the correct direction with suture's self intersections. Even crossed by other sutures, which was the case in Fig. 6(c), the shape can be correctly inferred and the computation can be terminated by Eq. (6).

2). *Key Points Approaching Along the Suture*: In this part, we built up a vision-based robotic system and extensively evaluated the computational accuracy along the whole suture. As shown in Fig. 7, sutures were stitched through different surgical phantoms, possessing various orientations. Using a calibrated stereo camera, we figured out its spatial coordinates in robot base using the information from an offline hand-to-eye calibration [31]. We attached a shaft with touching tip to the end of a UR5 robot, and the transformation matrix cT^R was reformulated by adding a translational vector in Z direction, while the orientation of the end effector was set vertical to the surface.

Key points were evenly selected from calculated results, and the touching tip was repeatedly driven to these positions to simulate a grasping operation and validate computational results concerning the suture's spatial coordinates. Key frames in 5 experiments were shown in Fig. 7, and errors between the tip and the suture were recorded in Table. II. As noticed that average errors in each group were all below 2.3mm, and the maximum one happened in the 5th set, which was 4.5mm. Considering errors induced from calibrations, such accuracy can be tolerated with a 10mm-long dVRK forceps when grasping a suture.

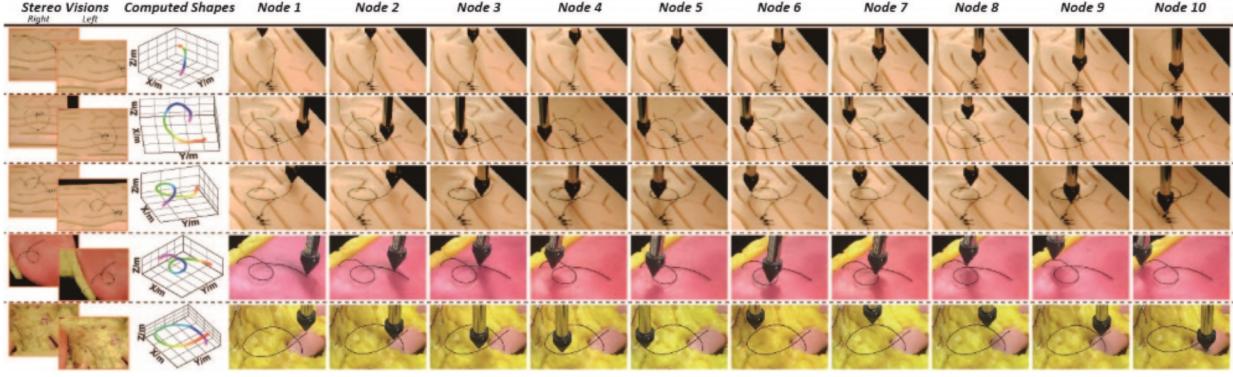


Fig. 7. The 3D reconstruction of a suture and a vision-based "grasping". We attached an end effector to UR5 to simulate the grasping task by sending the instrument's tip to follow the selected key points along the suture. Different artificial tissues were adopted as background.

TABLE II

EXPERIMENTAL RESULTS OF SUTURE "GRASPING" BY APPROACHING KEY NODES ALONG THE SUTURE - ERROR AMOUNT, UNIT: mm

Group	Selected Key Nodes 1 - 10										Ave	Max
Straight Suture	3.2	0.6	1.2	1.9	0.7	1.8	3.6	3.5	3	2.6	2.21	3.6
Curved Suture	0.8	0.3	2.3	1.9	2.4	1.6	2.1	0.8	1.3	0.8	1.43	2.4
Intersected Suture	1.8	4.1	1.4	1.1	0.8	2	1.5	2.8	1.2	1.8	1.85	4.1
Artificial Liver as background	1.6	1.4	2.8	2.7	2.2	1.7	0.9	0.5	1.3	1.1	1.62	2.8
Surgical Phantom as background	1.5	1.6	1.5	1.2	1.3	3.1	1.0	2.6	4.5	1.3	1.96	4.5

3). Feasibility Study by Simulation in V-REP and dVRK System: To evaluate the entire framework, we implemented it to V-REP simulator for validation. In the simulator, we built up sutures with different shapes by connecting various pieces of cylinders, whose length is 5mm with a radius of 0.5mm on the cross section. To characterize a smooth and slender suture, every two pieces were connected by a spherical joint, which was set invisible in the stereo vision sensors. Besides, 3D point clouds computed based on surgical videos were added to act as the background in space, and one typical case was shown in Fig. 8(a).

In the experiment, the grasping point reserved a 50mm length to the cutting tip as shown in Fig. 8(b), thus obtaining the grasping node GS_R in the camera coordinate. With a dVRK model introduced in [32], we can conveniently get the transformation relationship between PSM and ECM, and the grasping point in the robot space GS_R can be acquired.

We conducted 8 experiments in 2 test sets by orientating a suture with different shapes and lengths. Based on the computational results, a moving trajectory using the linear interpolation can be generated. Only with stereo visions and

TABLE III

EXPERIMENTAL RESULTS OF SUTURE "GRASPING" BY APPROACHING KEY NODES ALONG THE SUTURE - ERROR AMOUNT, UNIT: mm

Test Topology	Experimental Group 1 - 8							
Curved	1.4	2.1	2.4	1.2	3.2	3.2	2.0	3.1
Intersected	1.1	2.2	1.5	0.4	1.8	2.5	1.6	2.7

system's calibrations, the errors between GS_R (3D point of the end effector) and the desired grasping point (obtained by reading the 3D coordinates of the correct node) were listed in Table. III. It is worth noticing that the average errors in two testing sets were only 2.33mm and 1.73mm, and even the largest error was below 3.5mm, which is quite precise for a dVRK grasper to pick up the suture.

The proposed framework was further validated on a dVRK, and two successful trials of this vision-based automated suture's grasping were shown in Fig. 9. Summarizing all results above, we can demonstrate the feasibility of our proposed framework to automate the intermediate suture thread grasping step among the suturing, which contributes to a full automation of a robot-assisted surgical knot tying.

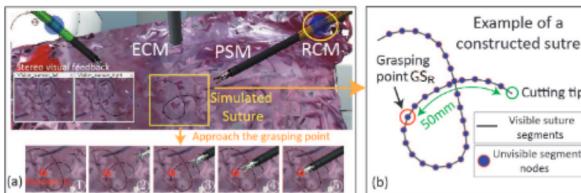


Fig. 8. Simulated suture grasping in V-REP. (a). Illustrations of a simulation environment using point cloud meshes and one simulated grasping process based on a dVRK model; (b). Example of a constructed suture in V-REP.

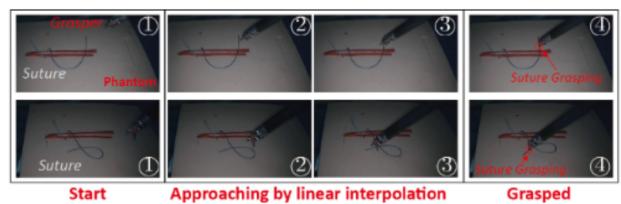


Fig. 9. Snapshots of automated grasping of a suture in a dVRK system based on the deep learning-driven framework.

IV. CONCLUSIONS AND DISCUSSIONS

This paper aims to tackle the technical challenges of automated suture grasping during an interrupted suturing task. A deep learning model, which transfers the information of large legacy surgical data to online acquired images, is leveraged in our framework for suture segment under on-site environmental noises. Based on this result, we further propose an algorithm for ordering sequence inference within stereo images, and the 3D coordinates of a designated suture can be thereby obtained by our spatial shape optimizer. Utilizing this novel framework, the grasping point can be achieved for an automated grasping manipulation.

We extensively evaluated the segmentation performance comparing our fine-tune model to models that were trained by legacy surgical data and limited online images. Our proposed approach obtained the highest IoU and F1 Scores, which were 71.1% and 97.5%. Besides, typical tests of suture's ordering sequence inference under multiple topologies were listed. By integrating a UR5 robot with a dual-camera system, experiments concerning approaching key points along the constructed suture were carried out to exploit the overall computational accuracy, and the average errors in 5 testing set were all below 2.3mm. We also demonstrated the feasibility of the entire framework by simulations in V-REP and by robot experiments in a dVRK system.

By tackling the critical issues in suture's perception, it can be concluded that our framework can provide a reliable guidance to achieve a higher level of assisted task autonomy in a suturing manipulation.

REFERENCES

- [1] R. H. Taylor and D. Stoianovici, "Medical robotics and computer-integrated surgery," Springer handbook of robotics, Springer, Cham, pp. 1657-1684, 2016.
- [2] J. Dargahi and S. Najarian, "Advances in tactile sensors design/manufacturing and its impact on robotics applications - a review" *Industrial Robot: An International Journal*, 32(3): 268-281, 2005.
- [3] C. W. Lee, et al., "Soft robot review," *International Journal of Control, Automation and Systems*, 15(1): 3-15, 2017.
- [4] R. H. Taylor, et al., "Medical robotics and computer-integrated interventional medicine," *Biomedical Information Technology*, Academic Press, pp. 617-672, 2020.
- [5] D. B. Camarillo, M. K. Krummel and J. K. Salisbury Jr, "Robotic technology in surgery: past, present, and future," *The American Journal of Surgery*, 188(4): 2-15, 2004.
- [6] H. Reich, J. DeCAPRIO and F. McGLYNN, "Laparoscopic hysterectomy," *Journal of Gynecologic Surgery*, 5(2): 213-216, 1989.
- [7] B. Lu, H. K. Chu, K. C. Huang and L. Cheng, "Vision-Based Surgical Suture Looping Through Trajectory Planning for Wound Suturing" *IEEE Transactions on Automation Science and Engineering*, vol. 16(2), pp. 542-556, 2019.
- [8] E. C. Mclemore, C. R. Harnsberger, R. C. Broderick, et al., "Transanal total mesorectal excision (taTME) for rectal cancer: a training pathway," *Surgical Endoscopy*, 30(9): 4130-4135, 2015.
- [9] H. P. Jaydeep, "Robotic assisted minimally invasive surgery," *Journal of Minimal Access Surgery*, vol. 5(1), 2009.
- [10] B. Lu, H. K. Chu and L. Cheng, "Robotic knot tying through a spatial trajectory with a visual servoing system," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2017.
- [11] S. Sen, A. Garg, D. V. Gealy, S. McKinley, Y. Jen and K. Goldberg, "Automating multi-throw multilateral surgical suturing with a mechanical needle guide and sequential convex optimization," *IEEE International Conference on Robotics and Automation*, 2016.
- [12] Y. Hu, L. Zhang, C. A. Seneci, W. Li, M. E. Abdelaziz and G-Z. Yang, "Design, Fabrication, and Testing a Semiautomatic Sewing Device for Personalized Stent Graft Manufacturing," *IEEE/ASME Transactions on Mechatronics*, vol. 24(2), pp. 517-526, 2019.
- [13] S. Leonard, et al., "Smart tissue anastomosis robot (STAR): A vision-guided robotics system for laparoscopic suturing," *IEEE Transactions on Biomedical Engineering*, vol. 61(4), pp. 1305-1317, 2014.
- [14] I. Laina, et al., "Concurrent Segmentation and Localization for Tracking of Surgical Instruments," *International conference on medical image computing and computer-assisted intervention*, 2017.
- [15] B. Huang, M. Ye, Y. Hu, A. Vandini, S. L. Lee and G-Z. Yang, "A Multirobot Cooperation Framework for Sewing Personalized Stent Grafts," *IEEE Transactions on Industrial Informatics*, vol. 14(4), pp. 1776-1785, 2017.
- [16] A. Shademan, et al., "Supervised autonomous robotic soft tissue surgery," *Science translational medicine*, vol. 8(337), pp. 337ra64-337ra64, 2016.
- [17] F. Alambeigi, Z. Wang, R. Hegeman, Y-H. Liu and M. Armand, "Autonomous Data-Driven Manipulation of Unknown Anisotropic Deformable Tissues Using Unmodelled Continuum Manipulators," *IEEE Robotics and Automation Letters*, vol. 4(2), pp. 254-261, 2019.
- [18] D. Navarro-Alarcon, H.M. Yip, Z. Wang, Y-H. Liu, F. Zhong, T. Zhang and P. Li, "Automatic 3D Manipulation of Soft Objects by Robotic Arms with Adaptive Deformation Model," *IEEE Transactions on Robotics*, vol. 32, no. 2, pp. 429-441, 2016.
- [19] G-Z. Yang, et al., "Medical robotics - Regulatory, ethical, and legal considerations for increasing levels of autonomy," *Science Robotics*, vol. 2(4), pp. 8638, 2017.
- [20] T. Osa, N. Sugita, M. Mitsuishi, "Online Trajectory Planning and Force Control for Automation of Surgical Tasks," *IEEE Transactions on Automation Science and Engineering*, 15(2): 675-691, 2017.
- [21] B. Lu, H. K. Chu, K. C. Huang, J. W. Lai, "Surgical Suture Thread Detection and 3D Reconstruction Using a Model-free Approach in a Calibrated Stereo-visual System," *IEEE/ASME Transactions on Mechatronics*, 2019.
- [22] N. Padoy, G. D. Hager, "3D thread tracking for robotic assistance in tele-surgery," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011.
- [23] R. C. Jackson, R. Yuan, D. L. Chow, et al., "Real-Time Visual Tracking of Dynamic Surgical Suture Threads," *IEEE Transactions on Automation Science and Engineering*, 15(3): 1078 - 1090, 2018.
- [24] Y. Gu, Y. Hu, L. Zhang, J. Yang and G-Z. Yang, "Cross-Scene Suture Thread Parsing for Robot Assisted Anastomosis based on Joint Feature Learning," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2018.
- [25] W. H. Lui and A. Saxena, "Tangled: Learning to Untangle Ropes with RGB-D Perception," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013.
- [26] J. Zhu, et al., "Robotic Manipulation Planning for Shaping Deformable Linear Objects WithEnvironmental Contacts," *IEEE Robotics and Automation Letters*, 5:1: 16-23, 2019.
- [27] B. Lu, X. B. Yu, J. W. Lai, et al., "A learning approach for suture thread detection with feature enhancement and segmentation for 3D shape reconstruction," *IEEE Transactions on Automation Science and Engineering*, 2019.
- [28] E. K. Batra, D. A. Franz, M. A. Towler, G. T. Rodeheaver, J. G. Thacker, C. A. Zimmer, R. F. Edlich, "Influence of emergency physician's tying technique on knot security." *The Journal of Emergency Medicine*, 10(3): 309-316, 1992.
- [29] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention*, pp. 234-241, 2015, Springer, Cham.
- [30] N. Jasika, "Dijkstra's shortest path algorithm serial and parallel execution performance analysis," *IEEE proceedings of the 35th international convention MIPRO*, 2012.
- [31] C. Wengert, et al., "Fully automatic endoscope calibration for intraoperative use," *Bildverarbeitung für die Medizin*, pp. 419-423, 2006.
- [32] G. A. Fontanelli, M. Selvaggio, M. Ferro, F. Ficuciello, M. Vendittelli and B. Siciliano, "A V-REP Simulator for the da Vinci Research Kit Robotic Platform," *IEEE International Conference on Biomedical Robotics and Biomechatronics*, pp. 1056-1061, 2018.