6-1-2022

Surgical Tool Segmentation with Pose-Informed Morphological Polar Transform of Endoscopic Images

Kevin Huang
Trinity College Hartford, kevinhuang@smith.edu

Digesh Chitrakar
Trinity College Hartford

Wenfan Jiang
Mount Holyoke College

Isabella Yung
Trinity College Hartford

Yun Hsuan Su
Mount Holyoke College

OPEN ACCESS
Journal of Medical Robotics Research, Vol. 7, Nos. 2 & 3 (2022) 2241003 (16 pages)

DOI: 10.1142/S2424905X22410033

Journal of Medical Robotics Research
https://www.worldscientific.com/worldscinet/jmrr

Journal of Medical Robotics Research

Surgical Tool Segmentation with Pose-Informed Morphological Polar Transform of Endoscopic Images

Kevin Huang[*,‡], Digesh Chitrakar[*,§], Wenfan Jiang[†,¶], Isabella Yung[*,‖], Yun-Hsuan Su[†,**]

[*]*Department of Engineering, Trinity College*
300 Summit St, Hartford, CT 06106, USA

[†]*Department of Computer Science*
Mount Holyoke College, 50 College St
South Hadley, MA 01075, USA

This paper presents a tool-pose-informed variable center morphological polar transform to enhance segmentation of endoscopic images. The representation, while not loss-less, transforms rigid tool shapes into morphologies consistently more rectangular that may be more amenable to image segmentation networks. The proposed method was evaluated using the U-Net convolutional neural network, and the input images from endoscopy were represented in one of the four different coordinate formats (1) the original rectangular image representation, (2) the morphological polar coordinate transform, (3) the proposed variable center transform about the tool-tip pixel and (4) the proposed variable center transform about the tool vanishing point pixel. Previous work relied on the observations that endoscopic images typically exhibit unused border regions with content in the shape of a circle (since the image sensor is designed to be larger than the image circle to maximize available visual information in the constrained environment) and that the region of interest (ROI) was most ideally near the endoscopic image center. That work sought an intelligent method for, given an input image, carefully selecting between methods (1) and (2) for best image segmentation prediction. In this extension, the image center reference constraint for polar transformation in method (2) is relaxed via the development of a variable center morphological transformation. Transform center selection leads to different spatial distributions of image loss, and the transform-center location can be informed by robot kinematic model and endoscopic image data. In particular, this work is examined using the tool-tip and tool vanishing point on the image plane as candidate centers. The experiments were conducted for each of the four image representations using a data set of 8360 endoscopic images from real sinus surgery. The segmentation performance was evaluated with standard metrics, and some insight about loss and tool location effects on performance are provided. Overall, the results are promising, showing that selecting a transform center based on tool shape features using the proposed method can improve segmentation performance.

Keywords: Robot-assisted minimally invasive surgery; surgical tool segmentation; telesurgery; U-Net.

JMRR

1. Introduction

In robot-assisted minimally invasive surgery (RMIS), the primary form of feedback to the surgical operator is vision, typically supplied via intraoperative endoscopy. With the absence of proprioceptive and tactile cues, it takes expert skill to safely operate in the human body [1,2]. There are a handful of potential methods to improve operator awareness, including multi-camera systems [3–8], magnetically anchored sensors [9], sensorizing surgical end effectors [10,11], and inferring tool–tissue physical interactions through vision. To

Received 2 December 2021; Revised 22 March 2022; Accepted 8 April 2022; Published 9 June 2022. Published in JMRR Special Issue on ISMR 2021. Guest Editor: Iulian Ioan Iordachita.
Email Addresses: ‡*kevin.huang@trincoll.edu*, §*digesh.chitrakar@trincoll.edu*, ¶*jiang24w@mtholyoke.edu*, ‖*isabella.yung@trincoll.edu*, ***msu@mtholyoke.edu*

achieve the latter, one primary goal is to segment tool pixels from tissue in endoscopic images [12,13].

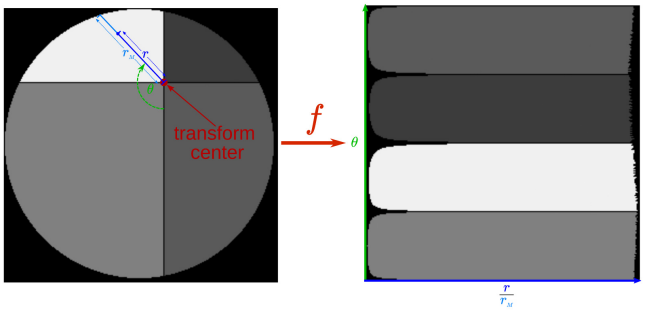
Towards that end, in previous work [14], polar morphological transforms (about the image circle center) were investigated as a means towards improved tool segmentation in RMIS sinus procedures. The method was proposed following three key observations:

(i) the image content from endoscopy is limited to the image circle — endoscopic image sensors are typically designed to be larger than the image circle [15];
(ii) ideally, the surgical tool tip and tissue interaction occur in the center of the field of view;
(iii) the sinus surgical procedure uses only one surgical instrument.

While results were promising for some images, success relied on the naive assumption of the tool tip remaining near the center of the image circle. In other cases, the original endoscopic image proved superior. As a compromise, a pair of selector neural networks were trained to predict, for any input image, whether the rectangular or polar coordinate represented mask would result in better segmentation in the Dice Coefficient sense. Analysis of the spatial features of masks suggested that, indeed, images with tool tip near the center of the image circle tended towards better performance with the preprocessing step of a morphological polar transformation.

The tool tip being near the image center cannot reliably be assumed, as during a typical procedure the operator is not restricted to using surgical tools only near the field of view center. A method for polar transformation of endoscopic images about the tool tip location (flexible transform center) is thus desired, yet requires spatial information about tool features with respect to the endoscope. Using robot forward kinematics or image data, tool shapes and tool tips may be projected on the image plane (assuming the endoscope configuration is also known) [16–18].

The work here extends the previous study by proposing and subsequently investigating a variable center morphological transform, which extends the benefits of the polar transform used in prior work while relaxing the tool tip location constraint. While incorporation of pose estimation to inform tool-tip is an intuitive augmentation, the implementation of a variable center morphological transformation on a digital endoscopic image is nontrivial and presents some technical difficulties. Namely, the spatial sampling of pixels in the polar transformation is nonuniform, and a consistent method for this transformation and its reverse operation must be established. Furthermore, losses associated with the variable-center transformation may affect segmentation performance and should be characterized. The morphological operation, f, should transform the circular image content around the transform center, as depicted in Fig. 1.



(a) (b)

Fig. 1. The polar transform is performed with reference to an arbitrary transform center within the image circle. (a) depicts an image circle with four distinct quadrants relative to a non-image-center transform center and (b) shows the variable center transformation result.

The reverse operation, g, should transform an image back to its rectangular form given the transform center coordinates. Various choices of transform center relative to spatial tool features are also investigated. Finally, different coordinate representations and transform center locations are quantitatively evaluated. The experimental procedure is depicted in Fig. 2.

1.1. *Contributions*

In this study, the authors present augmentations and improvements to previous work [14] by

(i) designing and implementing a tool-pose-informed variable center morphological polar transformation (with reverse operation) for endoscopic images;
(ii) analytically investigating the spatial loss as a function of the variable center location (forward and with back transformation);
(iii) experimentally comparing the tool segmentation performances between various approaches, namely using:
   — original rectangular baseline, $\mathcal{A}_{\mathcal{C}}$
   — the image-centered polar, $\mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}$
   — the tool tip transform center, $\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$
   — the vanishing point transform center, $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$

Compared to previous work [14], the methods presented here relax the constraint of fixed polar reference (image center) and allow for a variable transform center. To the best of the authors' knowledge, this work is the first to incorporate tool pose information as a guide to regularize the perceived tool shape through tool-pose-aware polar transformation. The results are promising, and suggest that polar morphological transformations with variable transform centers can generate image content in representations more amenable for image segmentation. The methods described are suitable for procedures similar to the sinus surgery task in this work — i.e.

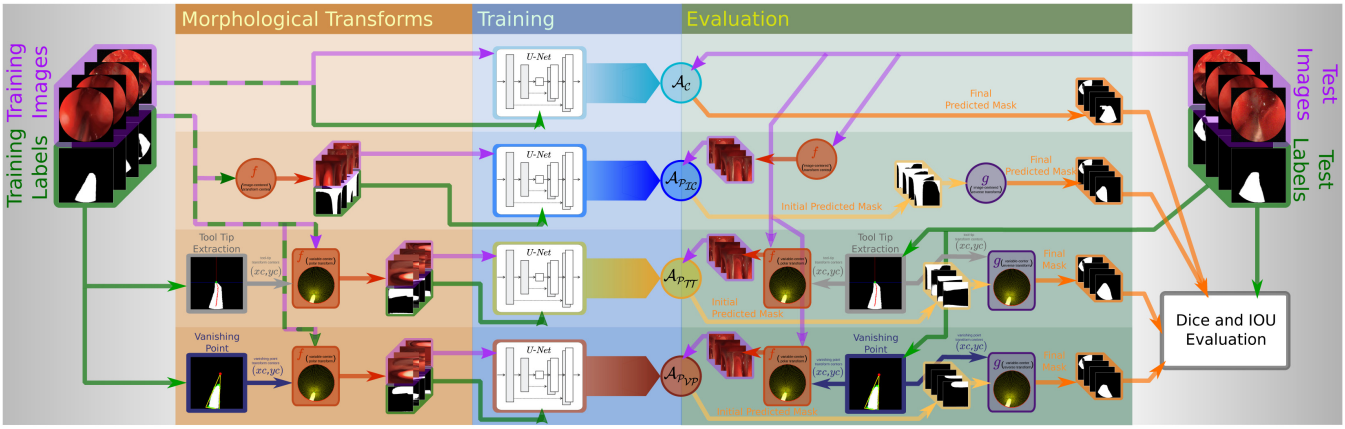


Fig. 2. The experimental workflow to evaluate various endoscopic image representations. For $\mathcal{A}_{\mathcal{C}}$, the original rectangular image format is preserved, while $\mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}$ utilized the transformation described in previous work [14]. $\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$ used the tool-tip as the transform center, and $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$ used the tool vanishing point as the transform center. The tool-tip and vanishing point centers could be inferred by kinematics, yet were designated via image processing approaches in this work.

single-tool robot-assisted endoscopic procedures — and the transform center is envisioned to be informed by robot kinematic state or image-based estimation.

1.2. *Related work*

1.2.1. *Endoscopic tool segmentation*

Endoscopes provide the primary forms of task awareness and visual feedback in RMIS [19,20]. In particular, endoscopic tools are often used in conjunction with laparoscopic procedures to reduce complications [21–23]. With that said, a key step in incorporating intelligence and enabling vision-based assistance is tool segmentation to separate tool pixels from tissue. Previous machine learning applications have attempted to increase the accuracy of tool segmentation via several approaches [24–27], including unsupervised methods like support vector machines [12] and Bayesian approaches [28]. The interest in this field is bolstered by more sophisticated methods such as deep learning via convolutional neural networks [29,30], including modifications to the popular U-Net architecture [13,31–33]. Improving tool segmentation has potential to increase surgical scene understanding and enable more intelligent assistant modes in RMIS.

This work aims to improve tool segmentation accuracy by implementing kinematics or tool shape projection aware morphological polar transformations as a preprocessing step prior to U-Net training.

1.2.2. *Image polar transformations*

Sensing modalities that sample radially, including some radar and ultrasonic sensors, provide polar representations of spatial data. This spatial arrangement is also found in some medical imaging systems, such as X-Ray scanning and computer-aided tomography, and must implement a reverse polar transformation to render and visualize the imaging content [34]. These resampling methods use the Fourier slice theorem and Jakowatz and O'Sullivan gridding methods [35]. In this work, the inverse problem is presented, where rectangular coordinate endoscopic images are transformed into an interpolated polar representation for segmentation purposes, and then back transformed to evaluate performance.

Log-polar transformation and adaptive polar transformations present methods of achieving image polar transforms. Log-Polar transformations can be used for affine image registration and to resolve scaling, shearing, rotation, or translation issues. Combining a log-polar transformation with a nonlinear least squares analysis in a hybrid approach demonstrated improvement with image registration [36]. Likewise, a hybrid approach using log-polar transformation and a Fourier transform demonstrated particular improvement in regards to rotation, scale, and translation [37]. Rotation and scaling have been a constant source of study using log-polar transforms and have several uses regarding optical flow and translational motions [38].

Adaptive polar transformations have also been used for image registration. The method effectively resolves a limitation of log-polar transforms, that is spatial alterations. Utilizing the fast Fourier transform allows greater bandwidth at a higher sampling frequency and with an increased radius [39]. A modified adaptive polar transform performed faster and with less errors than the adaptive polar transform [40]. However, these methods, as with the log-polar method, still incur high computational costs [41].

1.2.3. *Tool shape image plane estimation*

Tool-tip estimation has important applications in RMIS, since estimating the pose of surgical instruments may

assist in guiding the operator during surgery [42]. Utilizing sensors and encoders for instrument tracking requires extensive hardware integration and has limitations with accuracy. In contrast to this, image-based methods realize the pose of the instrument directly from the surgeon's view and has the added benefit of requiring no additional hardware [43,44]. Chen *et al.* [45] utilized a combination of visual tracking methods, convolutional neural networks (CNN) with line segment detector (LSD) for two-dimensional tool detection tracked with spatio-temporal context (STC) learning.

Tool-tip estimation can be simplified by augmenting instruments with color-based or infrared markers [46,47]. However, these methods struggle with visual obstacles like blood, motion blur, and occlusions [48]. Su *et al.* [16,18,49] explored methods that utilize both visual information and robot kinematics prior to providing accurate tool segmentation. Islam *et al.* [50] utilized an auxiliary supervised deep adversarial learning agent to segment the tool, but also provided more information such as parts of the surgical tool in an attempt to understand complex surgical scenarios.

One observation of existing work that incorporates robot kinematics with tool segmentation is that the tool pose information has only been used to provide additional cues of the perceived tool location on the image either in the form of a predicted mask of the surgical tool or as numeric inputs, but not as a means to simplify the perceived tool shape. Since the authors' prior study showed promising segmentation improvements by conducting polar transformation from the image center as a pre-processing step to simplify the tool shape, this work aims to investigate if tool shape location information can serve as a guide to select a custom polar transformation center for each image, and hence improve the segmentation results further by providing morphologically amenable shapes for segmentation networks.

2. Methods

2.1. *Data set curation and pre-processing*

The images used in this study were obtained from the University of Washington Sinus Surgery Cadaver/Live Data set [51,52]. The endoscopic videos were recorded using the Stryker 1088 HD camera and the Karl Storz Hopkins Ø4 mm 0° endoscope at 30 fps. The data set includes binary annotations of tool pixels manually labeled by experts, with a variety of visual obstacles present, including: motion blur, blood, occlusion, smoke, shadows, and specular reflections. The images in this dataset are obtained from a sinus surgery — a single surgical tool exists in the endoscopic images, and the orientation of the tool tip is rather consistent. Extending this work to a variety of surgical scenarios would require resolving multiple tools and their interactions.

In addition to the images and corresponding ground truth tool segmentation labels, the experiments in this work require numeric spatial information of the tool shape, i.e. tip and the vanishing point location. Although these could theoretically be extrapolated through robot kinematics, the Sinus Data set used in this study did not include robot state trajectories. The image pre-processing steps to estimate these tool shape features from the ground truth binary labels are described below in Secs. 2.3.1 and 2.3.2. The detailed usage of these tool shape location features is elaborated upon in Sec. 2.3.

2.2. *Variable center morphological transformation*

In previous work, an endoscopic image circle center morphological transform was used to spatially re-represent the image data. Readers are directed to that work [14] for details on the method. As previously described, this approach was based on the assumption that the surgical tool tip is near the image center. As the operator may navigate to various locations within the image frame, there is a desire to perform a similar morphological transformation but centered about an arbitrary location within the image. A method was developed that, given a transform-center pixel that represents the approximate tool-tip location within the endoscopic image, generates a spatially remapped version such that the

— rows represent equally sampled ray angles around the transform center;
— columns represent, along a given radial angle (specified by the row), equally sampled distances along that ray normalized by the distance from the transform center to the image circle perimeter.

In this work, endoscopic images are 256×256 pixel RGB images. Incorporating pose estimation to localize the tool-tip or vanishing point within the endoscopic image is an intuitive improvement of the image-centered polar transformation [14]. However, the technical implementation of this morphological transformation about an arbitrary pixel presents challenges. Sampling of the endoscopic image for an arbitrary polar reference center is spatially nonuniform, either in the original image or in the polar representation. Furthermore, once segmentation results are generated for the polar representation, the results must be adequately and consistently transformed back to the rectangular format to inform useful segmentation results for evaluation. A consistent and back-transformable mapping is proposed and investigated in this work, and inevitable data loss is characterized.

2.2.1. *Morphological mapping*

For the forward variable center transformation, f, let $\mathbb{N}_p$ be the set of natural numbers $\{1, 2, \ldots, 256\}$. Then

an endoscopic image may be represented by a set of $256^2 = 65,536$ different 3-tuples, which is denoted by $C = \{(x, y, v)\}$ where $(x, y) \in \mathbb{N}_p \times \mathbb{N}_p$ represent pixel coordinates and v the pixel content for the pixel at (x, y). Given transform center, $(x_c, y_c) \in \mathbb{N}_p \times \mathbb{N}_p$, the goal is to generate a polar representation of the original image at (x_c, y_c).

Since the variable center polar representation incorporates equally sampled ray angles about the transform center, intersections of the 256 equally spaced rays with the image circle perimeter are first determined (each row of the transformed representation contain pixels sampled along the line segment between (x_c, y_c) and a perimeter intersection point). To that end, let the 256 intersections be determined by solving simultaneous equations

$$\tan\left(i\frac{\pi}{128}\right) = \frac{p_{y_i}}{p_{x_i}}, \tag{1}$$

$$128^2 = (p_{x_i} - 128)^2 + (p_{y_i} - 128)^2, \tag{2}$$

where $i \in \mathbb{N}_p$ is the intersection point index. (1) ensures that the rays are equally sampled by radial angle, and (2) enforces the intersection with the image circle perimeter. Note that these two equations do not completely constrain the choices for p_{x_i}, p_{y_i} — the selection must be made to ensure that the solution within the correct quadrant is chosen, which is easily verified by the signs of $x_c - p_{x_i}$ and $y_c - p_{y_i}$. The set of 256 equally sampled (in terms of radial angle) perimeter points is thus chosen as the p_{x_i}, p_{y_i} where $i \in \mathbb{N}_p$. An example of these intersection points and the associated rays are depicted in Fig. 3(a).

For an original, rectangular representation for endoscopic image C, the accompanying variable center polar representation with transform center $(x_c, y_c) \in \mathbb{N}_p \times \mathbb{N}_p$ is expressed as a set of 3-tuples in $\mathbb{N}_p \times \mathbb{N}_p \times \mathcal{Z}$ where $\mathcal{Z}$ denotes the set of values that an image pixel may take. That set of 256×256 different 3-tuples is denoted P, where an element $(\chi, \psi, s) \in P$ represents the pixel of the transformed image in the χth row and ψth column, and s is the image content at that pixel location. The variable center transformation mapping, $f : (C, \mathbb{N}_p \times \mathbb{N}_p) \to P$ with transform center (x_c, y_c) is a surjective one. To define this mapping, select arbitrarily an element in P, call it (χ, ψ, s). Then

$$(\chi, \psi, s) = f((x, y, s), (x_c, y_c)),$$

where

$$x = \left\lfloor \frac{p_{x_\chi} - x_c}{\psi} + x_c \right\rfloor, \tag{3}$$

$$y = \left\lfloor \frac{p_{y_\chi} - y_c}{\psi} + y_c \right\rfloor, \tag{4}$$

where $(x, y, s) \in C$ and (p_{x_χ}, p_{y_χ}) is the χth perimeter point derived by (1) and (2). Pixel values in the transformed image P are pixels from C reparameterized by radial angle and normalized distance from the transform center, (x_c, y_c).

After a variable center transformed endoscopic image is segmented and a prediction is made, the prediction must be back-transformed into the original rectangular coordinate frame (i.e. the original image circle) to evaluate performance with rectangular coordinate endoscopic images and rectangular ground truth image labels. The back-transformation, g, is then composed of simple

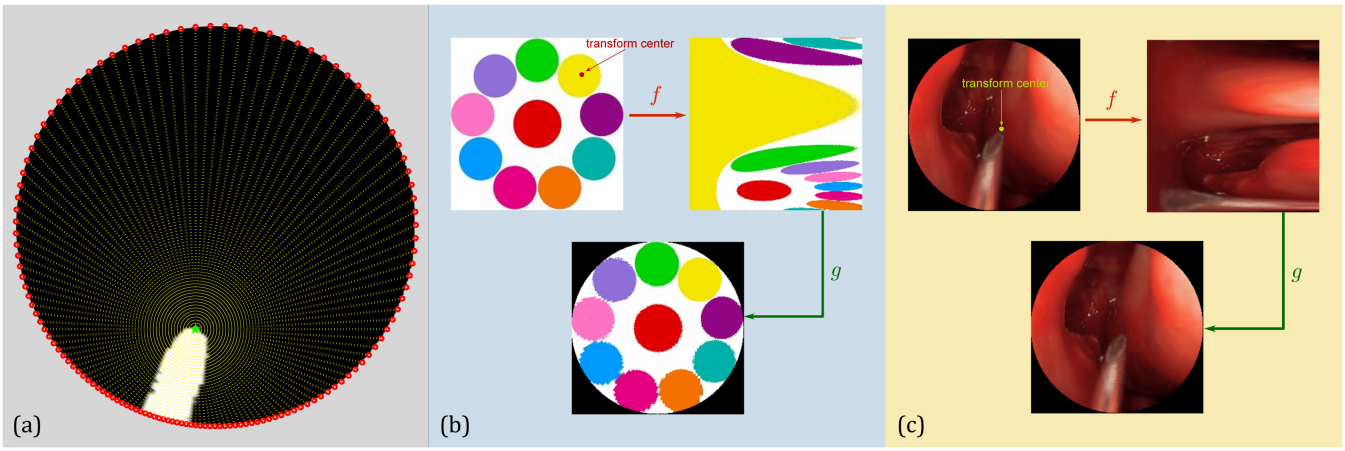


Fig. 3. Depictions of perimeter point detection, forward variable center polar transformation, f, and reverse transformation g. In (a), the transform center, green star at (x_c, y_c), is chosen at the tool tip. The perimeter points, at (p_{x_i}, p_{y_i}), are selected such that they are equally spaced in terms of angle with respect to the transform center (perimeter red circles). The polar morphological transform will contain pixel data with each row sampled from a single ray (yellow dotted lines). In (b), a color palette is the chosen input image, and the center of the yellow circle is chosen as the transform center — the resulting forward and reserve operations are shown. In (c), a sample endoscopic image from a live sinus surgery is shown, with transform center chosen at the tool tip. The forward transform results in the tool pixel representation similar to a rectangle. The back-transformed result retains the tool shape.

reverse operations of the forward operations for f. Note that this requires tracking of the transform center for the particular endoscopic image, and that g is a surjective mapping. For an arbitrary element, let it be (a, b, s), in the back-transformed image, B, the image content, s, at pixel (a, b) is obtained from the transformed image pixel (χ, ψ) by

$$(a, b, s) = g((\chi, \psi, s), (x_c, y_c)),$$

where

$$\chi = \left\lfloor 256 \sqrt{\frac{(a - x_c)^2 + (b - y_c)^2}{(p_{x_\psi} - x_c)^2 + (p_{y_\psi} - y_c)^2}} \right\rfloor, \tag{5}$$

$$\psi = \left\lfloor \text{atan2}\left(\frac{b - y_c}{a - x_c}\right) \frac{128}{\pi} \right\rceil. \tag{6}$$

An example of the forward and back transformation for variable center morphological polar mapping of a color palette and an endoscopic image are shown in Figs. 3(b) and 3(c).

2.2.2. *Pixel recovery for a given transform center*

For a given transform center, image information is hypothesized to be lost more frequently with increasing distance from transform center. This is depicted via several illustrative transform center locations in Fig. 4. Improved consistency of tool pixel content with the variable center is apparent compared to image center transformation. High recovery rate is focused near the transform center.

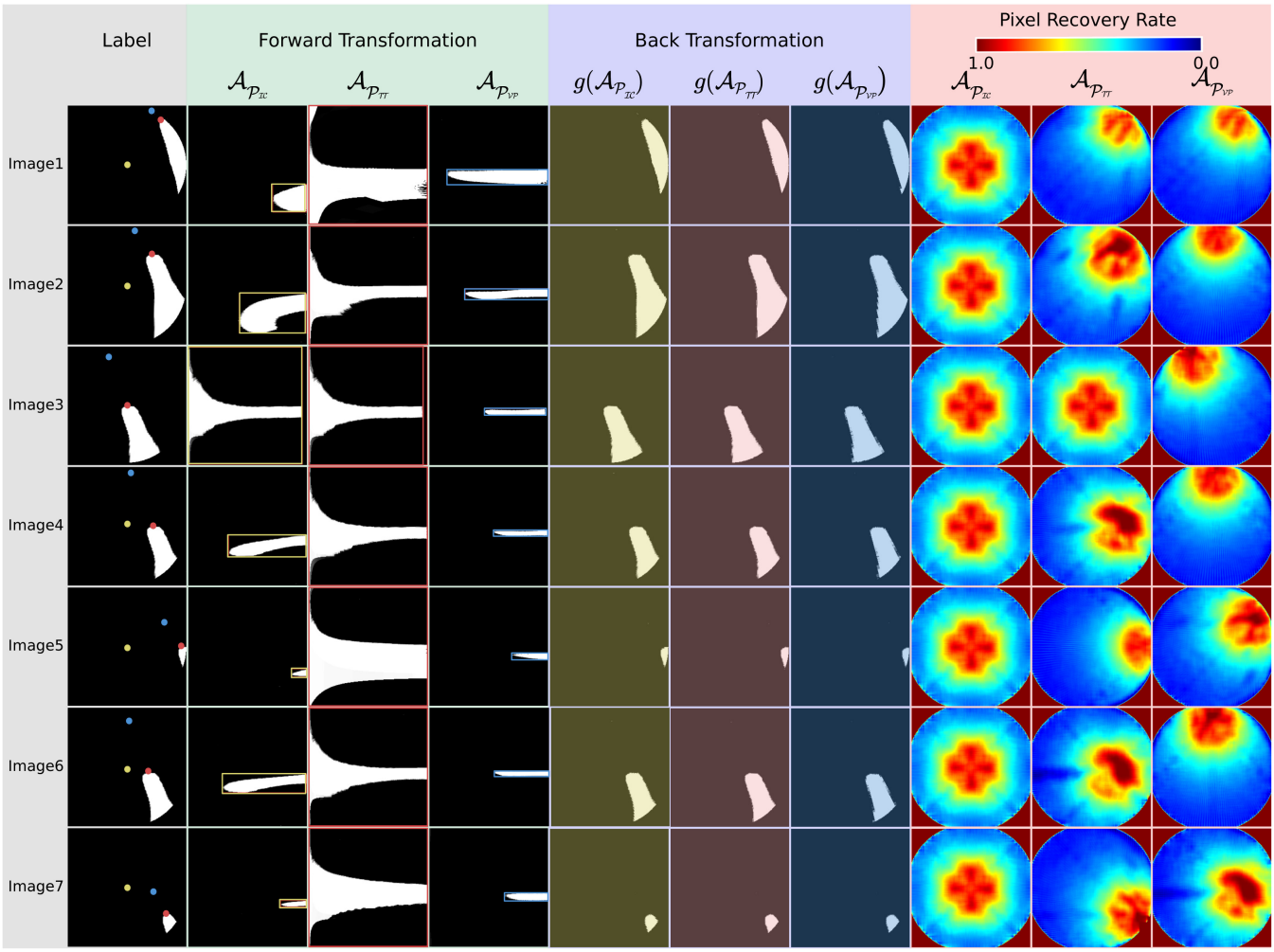


Fig. 4. Seven sample tool labels (7 rows) were selected to depict the local recoverability map as a function of the tool tip location and the selected polar approach $\mathcal{A}_{\mathcal{P}_{\mathcal{IC},\mathcal{TT},\mathcal{VP}}}$. Column one contains each raw label whose tool tip and vanishing point are marked with a red dot and a blue dot, respectively, and the image center is marked with yellow. Columns 2–4 are the polar transformation of the label sequentially using $\mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}$, $\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$, or $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$; a smallest bounding rectangle is overlaid on the polar images to illustrate the similarity of the resultant tool shape with a rectangle. The back transformation results from the three polar approaches are listed in Columns 5–7, where $\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$ most consistently preserves details surrounding the tool tip. Finally, the local recoverability maps are displayed in the last three columns to demonstrate the pixel recovery rate after the forward and backward transformations within a 9×9 neighborhood.

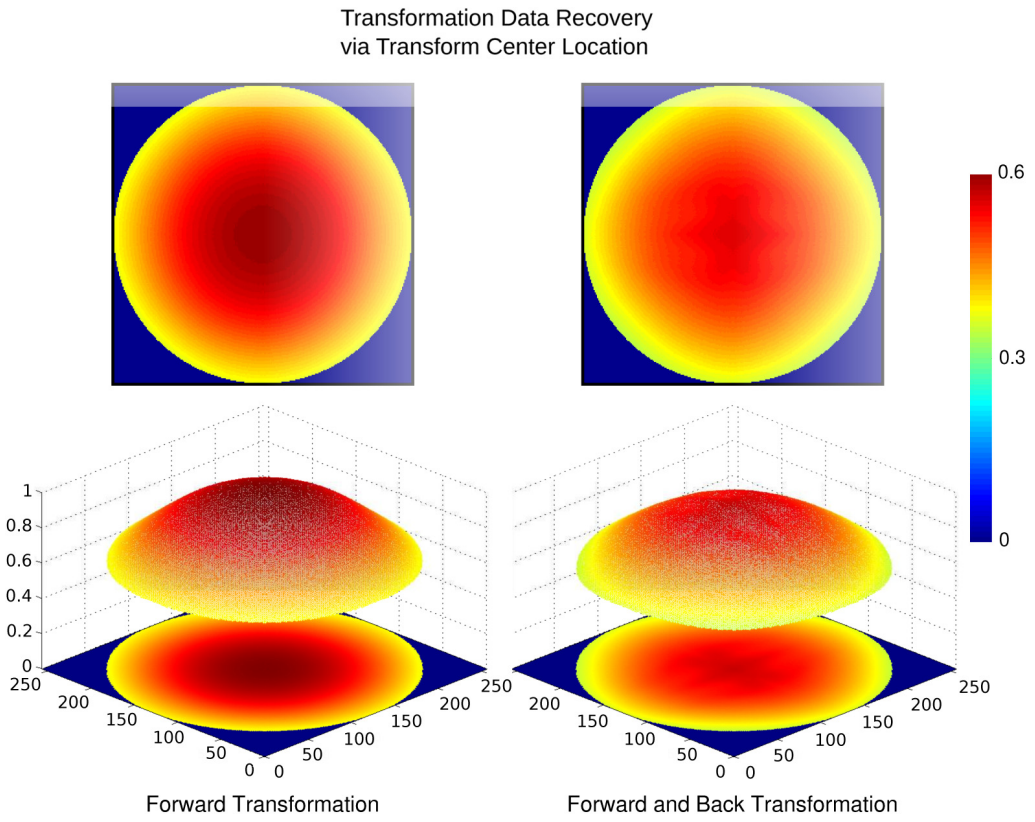


Fig. 5. The total recoverable points were tracked using the forward and back transformation for each of the possible 51,431 transform centers. The forward transformation has a mean recovery rate of 0.4706, while the total (forward and back) transformation incurs additional loss with a mean recovery rate of 0.4371. In both cases, transform centers closer to the image center result in less data loss.

2.2.3. *Recovery via transform center location*

The polar transformation does not sample a rectangular coordinate image uniformly. In particular, points are sampled more densely closer to the image center, and more sparsely near the image circle perimeter. While some methods seek to re-sample the polar mapping for more spatially uniform distribution of pixels selected, the techniques result in nonrectangular outputs [25] or are extremely computationally heavy [53]. The authors hypothesize that with straight surgical tools, data loss of pixels distant from the tool tip will not affect segmentation performance appreciably, as pertinent features are most likely near the tool tip.

With that said, the variable center transformation may exacerbate skewing of nonuniform spatial sampling towards the transform center. To evaluate the effect of transform center on total pixel loss, the transform center was applied at each pixel within the image circle. The proportion of distinct pixels recovered through the variable center polar forward and back transformation were logged at each transform center. The results are represented as a color map in Fig. 5. This spatial loss map is parameterized by transform center location, which is distinct from the spatial loss depicted in Fig. 4 where a single transform center is chosen and resultant pixel loss is tracked.

2.3. *Transform center selection*

The transform center in this method should be informed by relevant spatial features of the tool. These features can be derived by projecting the tool shape onto the image plane given known kinematics of the surgical robot and endoscopic imaging device. When choosing the transform center, two candidate locations were examined:

(1) tool tip;
(2) tool vanishing point.

The former was chosen to preserve as much information surrounding the tool–tissue interaction point, since areas near the transform center are densely sampled, as depicted in Fig. 4. The latter was considered since it would preserve, in the forward transformed image created by f, a tool region most contained within a rectangular bounding box suitable for the rectangular kernels in image segmentation networks. Unfortunately, the University of Washington Sinus Surgery Cadaver/Live Data set currently contains neither robot nor endoscope kinematic states. Instead, image-based methods were

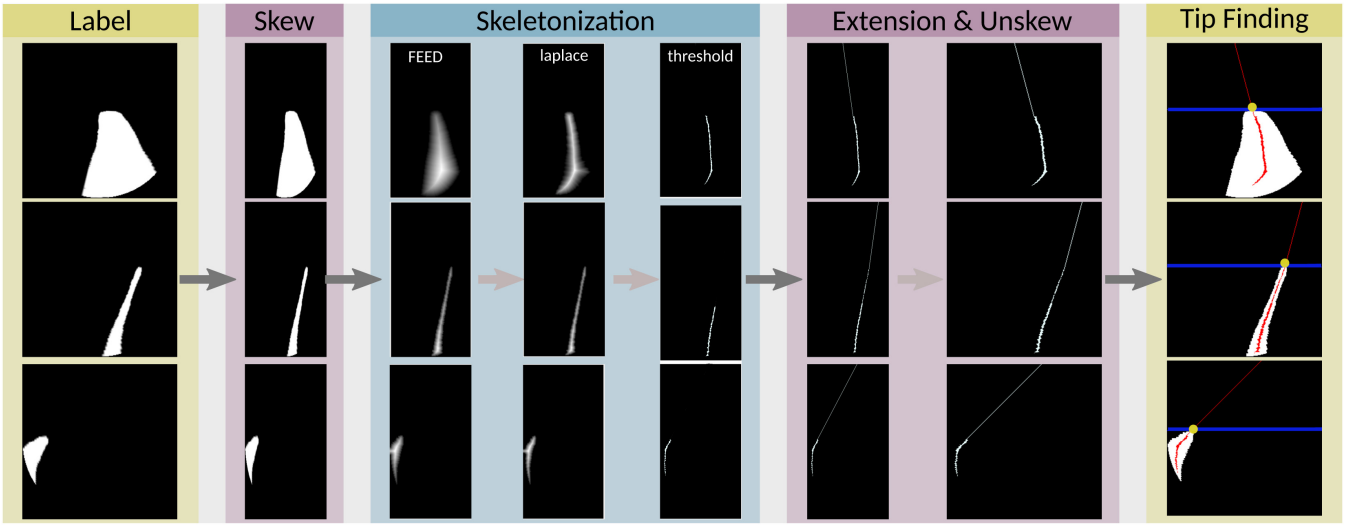


Fig. 6. Tool tip identification procedure. Pink blocks demonstrate the skew/unskew steps to amplify vertical height of the label. The blue block shows the intermediate results of the skeletonization step. The right most column denotes the mathematical derivation of the tool tip (yellow), the intersection of the skeleton (red), and the nearest nonintersecting horizontal line (blue).

used to batch identify tool-tip and tool vanishing point locations using the provided tool image labels.

2.3.1. *Tool tip identification*

The tool tips were automatically identified through a series of morphological transformations and filtering steps on the ground truth image labels. This procedure leverages one key observation of the data set: oftentimes the tool emerges into the scene from the bottom half of the image and points roughly toward the upper half.

As illustrated in Fig. 6, the image label is first vertically skewed to be twice the height to amplify the vertical appearance of the tool. Then the tool mask, which is denoted by the white pixels in the image label, is skeletonized. In this process, the pixels on object borders were removed successively until no more pixels can be removed without changing or destroying the connectivity [54]. This can be achieved through three image processing steps:

(1) apply the Fast Exact Euclidean Distance (FEED) transform [55] on the image label. This transformation outputs a distance map that has the same dimensions of the input image and each pixel encodes the Euclidean distance to the closest 0 valued pixel. This will generate a ridge in the center of the tool representing the "bones" of the skeleton;
(2) calculate the morphological Laplacian [56] edge detection of the resultant image. This process is performed to amplify the skeletal ridge;
(3) finally, take a heuristically tuned threshold to obtain a binary mask of the skeletonized tool mask.

Next, the skeletonized mask is unskewed by re-scaling the vertical height by linear extension in the tangent direction that predominantly is angled up. Finally, the tool tip is identified in the last column of Fig. 6 as the intersection between the extended skeleton and the nearest horizontal line that does not intersect the tool. Since the tool is always pointing roughly vertically in each image, vertical skewing will not significantly influence the relative position of the tool tip, and yet this extra step resolves edge cases where only a small portion of tool with greater width than height is present, which would result in a horizontal skeleton without the skewing step.

2.3.2. *Vanishing point identification*

The tool vanishing point identification was performed using simple black and white image processing, as well as a linear minimum area enclosing triangle algorithm [57]. Because the tools are rigid, mostly cylindrical objects, and the view angle from the endoscope typically views partially along the axial direction of the tool, a vanishing point is expected. That is, extending the edges of the tool shape via perspective projection should result in an intersection point — thus a minimum bounding triangle is sought.

To start the process for any endoscopic image/label pair, the perimeter of the labeled tool was first extracted in the 4-connected sense, and a minimal vertex convex hull was generated. A minimum-area bounding triangle was then fit to the minimal convex hull, resulting in three candidates for the vanishing point location. Two cases then arise:

i) at least one vertex is within the image circle;
ii) all vertices are outside the image circle.

For the former, the vertex that is closest to the image center in the 2-norm sense is selected as the tool vanishing

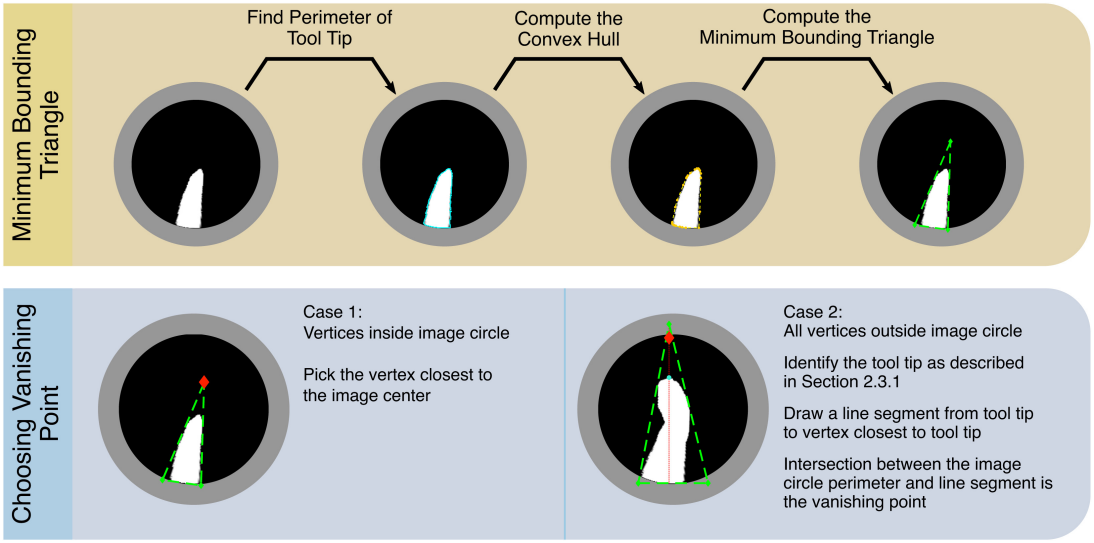


Fig. 7. Illustration of vanishing point selection from tool tip label. First the perimeter is extracted and a minimal convex hull of the tool tip is computed. A minimum bounding triangle is computed around the convex hull. For selecting the vanishing point, two cases arise, (1) at least one vertex inside the image circle, (2) no vertices inside the image circle. In the former, the vanishing point is simply the closest vertex to the image center. For the latter, the tool tip is calculated as described in Sec. 2.3.1, and a line segment is drawn from the tool tip to the triangle vertex closest to the tool tip. The vanishing point is then chosen as the intersection of the image circle perimeter and this line segment. The vanishing point is denoted as a red diamond, and the tool tip is denoted as a cyan circle.

point. Since the tool extends from outside the field of view to the tool tip, most minimum enclosing triangles will have at least two vertices outside the image circle, leaving the one closest to the tool-tip as the one inside.

In the case of the latter, the tool tip is first identified as described in Sec. 2.3.1. Then, the minimum enclosing triangle vertex closest to the tool-tip location in the 2-norm sense is identified. A line segment is then drawn between the tool-tip and the identified triangle vertex, and the point that both lies along the image circle perimeter and on the line segment is chosen as the vanishing point transform center. The entire process of vanishing point identification is depicted in Fig. 7.

3. Experiments

3.1. *Experimental conditions*

The endoscopic images were pre-processed through four different morphological representations of image–label pairs for training the U-Net image segmentation network, resulting in the experimental conditions as described below. The conditions are

$\mathcal{A}_{\mathcal{C}}$ The rectangular condition serves as the control baseline, whereby no morphological transform is applied. The original endoscopic images and labels are used as inputs and outputs to the U-Net segmentation network for training. This case represents the standard procedure.

$\mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}$ The image centered treatment transforms endoscopic images and labels via a polar transformation about the fixed image center. This operation was introduced in prior work [14], and was found to improve segmentation performance for some images. In particular, results suggested that endoscopic images with the tool tip near the center of the image tended towards better segmentation with this polar transformation.

$\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$ In this representation, for each input image, the tool tip identification as described in Sec. 2.3.1 is used on the label to identify the variable transform center for morphologically mapping the image–label pair via the proposed method in Sec. 2.2.1. The transformed images and labels are then used as inputs and outputs, respectively, for training the U-Net segmentation network.

$\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$ In the vanishing point center method, the pre-processing is identical to the preceding procedure except the tool tip identification is performed as described in Sec. 2.3.2.

3.2. *Training and evaluation*

The segmentation models in each of the $\mathcal{A}_{\mathcal{C}}, \mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}, \mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$ and $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$ methods used the U-Net architecture with dice coefficient loss function, D_L. Suppose that Y_t is the ground truth segmentation and Y_p is the generated segmentation prediction for a given input image. Then the dice loss is computed as follows:

$$D_L = 1 - \frac{2[Y_t \cap Y_p] + S}{Y_t + Y_p + S}, \tag{7}$$

where $S = 1$ as a smoothing term in order to avoid division by zero. While adding a small ϵ in the denominator also prevents division by zero, the proposed loss function additionally inhibits overfitting by limiting the penalty from Sørensen–Dice coefficient.

3.2.1. *Training specifications*

90% of the total data set was used for training, while the remaining 10% was used for testing. This resulted in 956 testing images and 7404 randomly selected training image–label pairs. For the training parameters, a batch size of two for 50 epochs was utilized. Within each epoch, 100 batches were used. A learning rate of 1×10^{-4} was used with the Adam optimizer. Additional augmentations including rotation range of 0.2, vertical and horizontal shift range of 0.05, shear range of 0.05, zoom range of 0.05, and occasional horizontal flip were implemented in each of the methods using the Keras API.

3.2.2. *System specifications and hardware*

Training and testing were performed on an Ubuntu 20.04 system with hardware specs: quad-core 64-bit Intel® 205Core™ i7-7700 running at 3.6 GHz, 16 GB of DDR4 DRAM 3000 MHz system memory, and a graphics engine consisting of an NVIDIA GeForce GTX 1070 with 8 GB of GDDR5 256 bit dedicated memory. The U-Net was realized using Keras 2.4.3 and TensorFlow 2.2.0.

4. Results and Discussion

4.1. *Rectangle similarity*

To evaluate the effectiveness of the forward polar morphological transforms, the rectangle similarity of forward transformed labels was evaluated for seven test images, as shown in Fig. 4 (bounding boxes in the "Forward Transformation" column). The rationale and procedure is described as follows.

Since the surgical tool in this data set roughly points upward and into the image from the bottom half, a larger portion of the tool shaft is visible when the tool tip extends towards the top of the image. Therefore, depending on the height of tool tip, the forward and back-transformed polar images will have different resolutions. The behavior in the left/right sense is hypothesized to be relatively symmetric. With that said, the seven points were carefully selected where the tool tips appear within the 1st and 4th quadrants of the image frame, respectively. These selections were such that the tool tip was at the image center (Image 3), within the 1st quadrant (Images 1,2), within the 4th quadrant (Images 6,7), and along the border of the two quadrants (Images 4,5).

Table 1. Rectangular similarity score.

Image	$\mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}$	$\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$	$\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$
1	69.2567	30.2714	76.0773
2	57.5502	18.8297	62.7191
3	18.1738	17.3918	72.5027
4	47.7386	19.9569	75.2137
5	54.0441	36.5521	57.8632
6	50.0000	20.0093	76.8678
7	64.2857	27.5436	75.3953
Mean	51.5784	24.3650	70.9484

It is hypothesized that more rectangular image content may be better suited for the kernels in the U-Net. To investigate this, the smallest bounding rectangle was overlaid on each forward polar label and a similarity score was assessed — note that no causal relationship is strictly demonstrated here. The score is defined as follows:

$$T_s = 100 \frac{N_t}{A_b}, \tag{8}$$

where N_t is the number of tool pixels within the bounding box, and A_b is the bounding box area. These scores are shown in Table 1. One can observe that $\mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}$ produces the most inconsistent scores; $\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$ is consistently low due to the wide angle occupancy in the small radius region (left border of the image); and the $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$ scores are consistently high. However, the back transformation results in Fig. 4 show the best recovery near the tool tip for $\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$, and noticeable noise and uncertainty near the tool base for $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$. Again, the tool region recovery varies spatially for $\mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}$, but since the transform center is fixed at the image center, the overall rate is greatest. In other words, it shows the best overall recovery rate as depicted in the "Recovery Rate" column of Fig. 4 and Table 4.

4.2. *Spatial recovery rate*

The spatial recovery rate of the three proposed transform center choices was investigated, with graphical results shown in the right columns of Fig. 4. This metric is defined as the percentage of unique points preserved after the morphological polar transformation. The numeric values of the total transformation (forward and back) recovery for various transform center approaches are shown in Table 2.

These results are consistent with graphical representations shown in Fig. 4, notably that $\mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}$ has a consistent and higher recovery rate compared to $\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$ and $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$. Furthermore, the results indicate that the vanishing point approach shows the worst recovery. This can be attributed to the propensity for the vanishing point to be further from the image center compared to the tool tip.

Table 2. Spatial recovery of transformed images [%].

Image	$\mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}$	$\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$	$\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$
1	53.8331	34.4067	34.0507
2	53.8331	43.9584	36.2786
3	53.8331	53.8331	37.9286
4	53.8331	49.5933	37.8138
5	53.8331	35.8018	40.9561
6	53.8331	49.8385	41.7695
7	53.8331	38.7652	49.0310
Mean	53.8331	43.7424	39.6898

4.3. *Segmentation performance*

4.3.1. *Forward transformation*

It was of interest to evaluate the performance of the tool segmentation at each stage of the morphological transformations. To that end, predictions from the trained models $\mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}, \mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$, and $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$ are represented in the forward transformed coordinates. Before transforming back for final evaluation, via back transformation g, the predictions were compared with the forward transformed ground truth labels. This evaluation provides insight into the accuracy of the U-Net given the morphological transform inputs and segmentation performance prior to transforming back to rectangular coordinates and incurring the losses therein. The Dice and IoU scores for the direct predictions from the polar trained U-Nets are shown in Table 3.

These results show that the morphological consistency of the vanishing point variable center transform were amenable to better segmentation, as across both Dice and IoU, $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$ attained highest scores. While the rectangle similarity scores were poor for $\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$, the tool morphology was consistent for the same tool despite shifting location, especially as compared with $\mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}$. This can be observed in the "Forward Transformation" column of Fig. 4. The transformed training label shapes were observed to be consistent across both shape and location for $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$ as compared to $\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$.

Table 3. Segmentation of transformed images.

Metric		$\mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}$	$\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$	$\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$
Dice	Mean	0.8853	0.9424	0.9482
	Median	0.9172	0.9463	0.9522
IoU	Mean	0.8077	0.8705	0.9029
	Median	0.8471	0.8753	0.9088

Table 4. Evaluation on rectangular label.

	Dice		IoU	
Algorithm	Mean	Median	Mean	Median
$\mathcal{A}_{\mathcal{C}}$	0.8987	0.9056	0.8273	0.8345
$\mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}$	0.9198	0.9268	0.8554	0.8637
$\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$	0.9368	0.9450	0.8632	0.8730
$\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$	0.9228	0.9293	0.8604	0.8680

4.3.2. *Total segmentation*

While the segmentation performance in the polar coordinate representations for various transform centers provides insight into the raw effectiveness of the trained U-Nets with respect to the training inputs and outputs, to compare with the baseline of the original endoscopic image, the back-transformed segmentation results must be analyzed. The predicted labels were thus reverted back to their rectangular coordinate representation via the appropriate back transformation, g. Subsequently, the Dice and IoU metrics for all representation methods were evaluated on the original ground truth label for the test set. The results are shown in Table 4.

As shown, the tool-tip transform center model, $\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$, outperformed all other methods when evaluating the total transformation (i.e. forward transformation image–label pairs for training, back transformed predictions for evaluation). This suggests that, while the vanishing point transform center U-Net, $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$, was best at segmenting test data represented in the respective training domain, the losses incurred by transforming back hindered overall performance. As depicted in Fig. 5, the loss observed in the forward transform is exacerbated after applying g, and the tool vanishing point propensity to be near the image circle border is greater than the tool tip. These factors combined resulted in loss of granularity near the tool base for $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$, despite consistent and superior segmentation in the forward transformed representation.

4.4. *Segmentation performance ranking by tool spatial features*

The segmentation performances of four different representation approaches were also examined based on the location of tool spatial features. Recall that in total, two such tool features were used to determine transform centers: (1) tool tip and (2) vanishing point. These shape features may be of interest in examining the performance trends of each method with respect to the feature location.

4.4.1. *Performance by tool-tip location*

For each test input image, a tool-tip location was determined via the method described in Sec. 2.3.1. The Dice

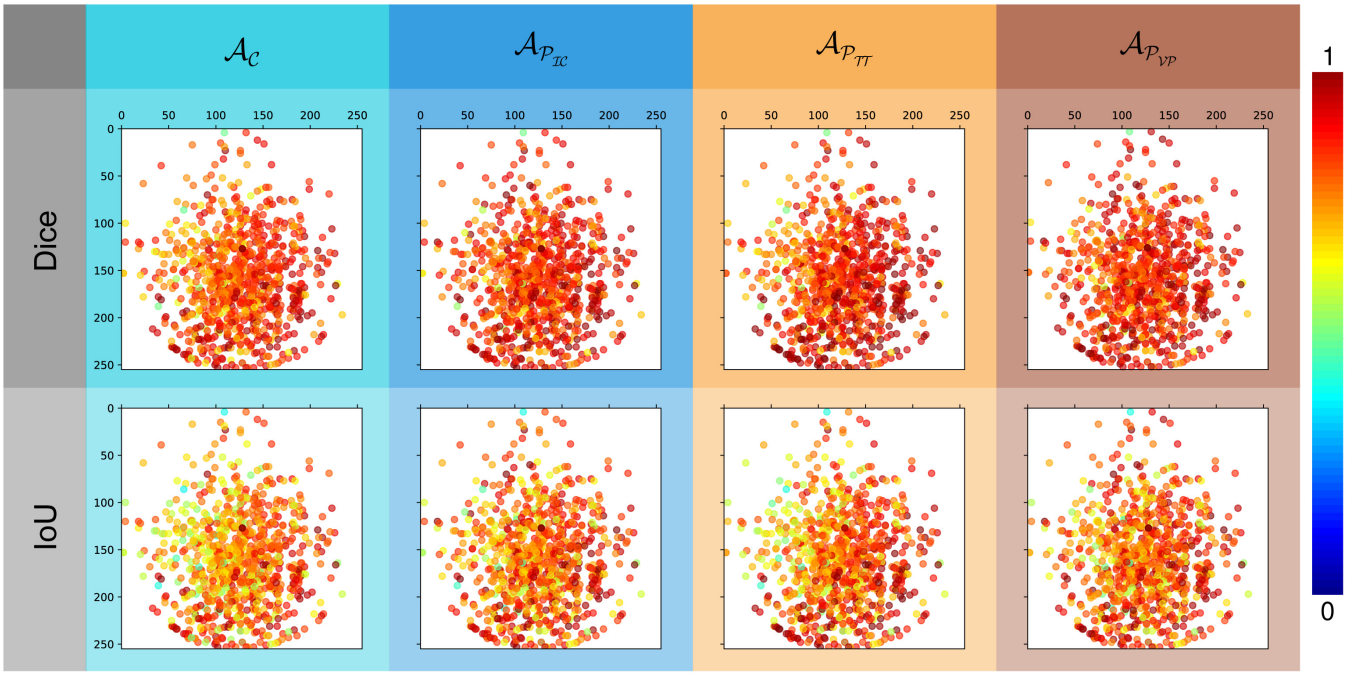


Fig. 8. Dice and IoU scores for each method, $\mathcal{A}_{\mathcal{P}_{\mathcal{C}}}, \mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}, \mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$, and $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$ with tool tip location tracked. No observable regions of particularly biased performance for any of the methods were observed, suggesting the tool tip location is not a strong indicator for selecting a transform center type for a particular image.

and IoU metrics for each metric were recorded and are plotted spatially by the determined tool tip location and via color code to show performance in Fig. 8. No distinguishable trend in performance versus identified tool tip locations is observed within any of the four methods.

4.4.2. *Performance by vanishing point location*

In Fig. 9, the best performing algorithm for each vanishing point location is color coded and a tinted spline is used to represent the spatial coverage of best performance.

The distribution of the best performing algorithm when the vanishing point is determined to be in the upper semicircle of the image circle is rather uniform. In contrast, when the vanishing point is determined to be near the bottom of the endoscopic image, the method $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$ outperforms other methods. Recall that the tools in this data set predominantly enter the field of view from the bottom of the image. This suggests that vanishing

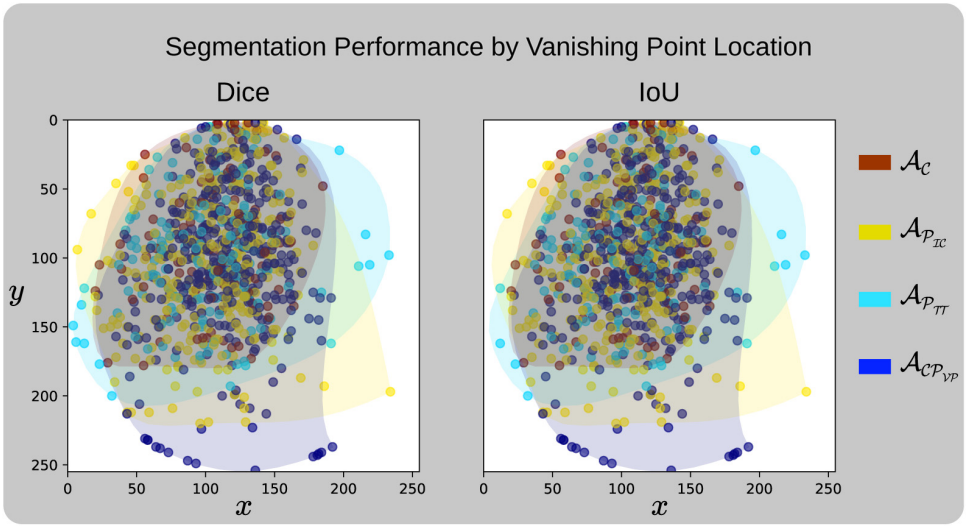


Fig. 9. For each image, the best performing method out of the four treatments, $\mathcal{A}_{\mathcal{C}}, \mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}, \mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$, and $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$, were tracked and are color coded here, spatially mapped by the vanishing point location determined for that test image.

points in the lower half of the image circle represent tools that are just beginning to enter the endoscope field of view.

With that in mind, when the tool tip enters the field of view sufficiently, the light-blue tinted spline representing $\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$ begins to perform better. This is consistent with the tool tip location approaching the center of the image.

5. Conclusion

In this work, a variable center morphological polar transformation method was introduced for re-representing endoscopic image data for the purposes of image segmentation given robot kinematic state. In previous work, a similar approach was investigated with promising results, yet with constraints on tool-tip location (kinematic state ignorant). The approach here relaxed those constraints, and is suitable for applications where kinematics and tool shape projection on the imaging plane can be estimated. This kinematics aware approach also necessitates a choice for the transform center. In this manuscript, two such choices were evaluated: (1) the tool tip and (2) the tool vanishing point. The two representation methods investigated in prior work were also examined in this paper. Thus, a total of four different methods were evaluated:

$\mathcal{A}_{\mathcal{C}}$ – original baseline
$\mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}$ – image centered transform
$\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$ – tool tip transform center
$\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$ – vanishing point transform center

In general, the transformed images will oversample pixels near the transform center. The variable center morphological transformations were shown to preserve image information proximal to the selected transform center. Method $\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$ thus will preserve details near the tool-tip location, which may be a benefit since the tool tip is the primary tool–tissue interaction point. On the other hand, method $\mathcal{A}_{\mathcal{P}_{\mathcal{IC}}}$, the original image centered polar transformation, has consistently equal or better content recovery over both $\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$ and $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$ while performing better than $\mathcal{A}_{\mathcal{C}}$ when the tool is near the center. This compromise was investigated in prior work. Finally, $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$ resulted in the most consistent input images for training the U-Net, but at the cost of lower recovery near the tool base. With that said, the method still outperformed $\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$ in certain cases.

The results of this work suggest that with robot kinematic state or otherwise tool position informed procedures, a variable center morphological polar transform can result in better tool segmentation performance. Future work will investigate which features of tool shape state may be useful in discriminating which transform center is best suited for an arbitrary input image, and on-the-fly transformation is possible with the simple transformation operations. In general, the work presented here coupled with the authors' previous work provides an encouraging framework and initial investigations into the use of the polar transform on endoscopic images for segmentation purposes.

5.1. *Future work*

Following the methods in the authors' prior work, a preliminary multi-class selection network was trained with inputs containing four layers, one for each segmentation mask type, with a convolutional neural network. The network yielded near perfect algorithm selection and resulted in mean and median dice scores of 0.945 and 0.951 and mean and median IoU scores of 0.883 and 0.892. A more interpretable feature space is sought in order to describe the characteristics of input images to inform the best usage of the kinematics informed variable center transform.

Observing Fig. 8, image samples that yielded relatively better segmentation results tended to consistently perform better across the four approaches. To further distinguish any differences, some individual segmentation results were examined. Preliminary investigations note that performance advantages may be spatially related.

Figure 10 is an example of a typical nonideal raw segmentation result. The left depicts the ground truth image label, while the center figure shows the $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$

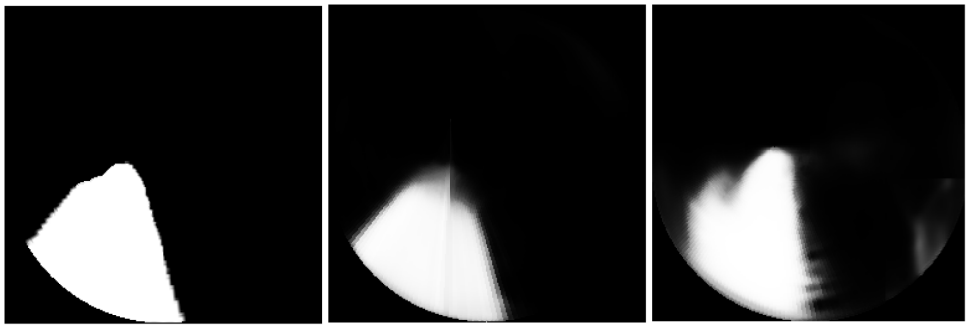


Fig. 10. From left to right: a tool tip label accompanied by raw (prior to thresholding) segmentation predictions using $\mathcal{A}_{\mathcal{P}_{\mathcal{VP}}}$ and $\mathcal{A}_{\mathcal{P}_{\mathcal{TT}}}$. This is an example of a poor segmentation, yet the results complement one another.

prediction and the right the $\mathcal{A}_{\mathcal{P}_{TT}}$ result. $\mathcal{A}_{\mathcal{P}_{VP}}$ appears to preserve the overall tool shape, whereas $\mathcal{A}_{\mathcal{P}_{TT}}$ provides great detail proximal to the tool tip. These advantages appear mutually exclusive between the two methods in this test image. Thus, instead of a selector network, a stochastic fusion-based approach, which may adopt local features from multiple predictions, might yield better results compared to choosing one prediction alone among four. As future work, kinematic information in addition to spatial image features will be incorporated with an aim to systematically predict regions or states that will yield more confident results and algorithm selection or fusion.

The methods presented in this work were developed within the context of the endoscopic images in the University of Washington Sinus Surgery Cadaver/Live Data, as described in Sec. 2.1. Of particular note, at most a single surgical tool is present in the endoscopic images, and the tool tip typically enters the frame from the lower half of the image. Many RMIS procedures involve multiple tools. Future work may need to consider multi-tool segmentation and resolving numerous predictions for a given endoscopic image. Frame-by-frame tracking of multiple tools may inform isolated ROIs for each tool and subsequent processing.

While timing considerations are beyond the interest of this work, future work may seek to gain more specific and detailed insight into timing performance. In this research, the flexible center polar transformation is a preprocessing step which operates at around 15 Hz on the hardware specified in Sec. 3.2.2. As reported in [58], the standard U-Net makes inferences at around 188 ms per 256×256 image, or roughly at 5 Hz. With hardware acceleration and parallel computing, the bottleneck in the proposed method would be in segmentation mask generation, not the polar transformation. The polar transformation pre-processing step is more computationally efficient than the U-Net inference itself, and thus is unlikely to hinder the overall segmentation frame rate.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant IIS-2101107. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author (s) and do not necessarily reflect the views of the National Science Foundation.

References

1. J. K. Koehn and K. J. Kuchenbecker, Surgeons and non-surgeons prefer haptic feedback of instrument vibrations during robotic surgery, *Surg. Endosc.* **29**(10) (2015) 2970–2983.
2. K. Huang, D. Chitrakar, R. Mitra, D. Subedi and Y.-H. Su, Characterizing limits of vision-based force feedback in simulated surgical tool-tissue interaction, *2020 42nd Annual Int. Conf. IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE, 2020), pp. 4903–4908.
3. Y.-H. Su, K. Huang and B. Hannaford, Multicamera 3d reconstruction of dynamic surgical cavities: Autonomous optimal camera viewpoint adjustment, *2020 Int. Symp. Medical Robotics (ISMR)* (IEEE, 2020), pp. 103–110.
4. H. Urey, K. V. Chellappan, E. Erden and P. Surman, State of the art in stereoscopic and autostereoscopic displays, *Proc. IEEE* **99**(4) (2011) 540–555.
5. Y.-H. Su, K. Huang and B. Hannaford, Multicamera 3d reconstruction of dynamic surgical cavities: Non-rigid registration and point classification, *2019 IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)* (IEEE, 2019), pp. 7911–7918.
6. M. Silvestri, T. Ranzani, A. Argiolas, M. Vatteroni and A. Menciassi, A multi-point of view 3d camera system for minimally invasive surgery, *Sensors Actuators A: Phys.* **202** (2013) 204–210.
7. Y.-H. Su, K. Huang and B. Hannaford, Multicamera 3d reconstruction of dynamic surgical cavities: Camera grouping and pair sequencing, *2019 Int. Symp. Medical Robotics (ISMR)* (IEEE, 2019), pp. 1–7.
8. Y.-H. Su, K. Huang and B. Hannaford, Multicamera 3d viewpoint adjustment for robotic surgery via deep reinforcement learning, *J. Med. Robot. Res.* **6**(01n02) (2021) 2140003.
9. N. Di Lorenzo, L. Cenci, M. Simi, C. Arcudi, V. Tognoni, A. L. Gaspari and P. Valdastri, A magnetic levitation robotic camera for minimally invasive surgery: Useful for notes?, *Surg. Endosc.* **31**(6) (2017) 2529–2533.
10. K. S. Shahzada, A. Yurkewich, R. Xu and R. V. Patel, Sensorization of a surgical robotic instrument for force sensing, *Optical Fibers and Sensors for Medical Diagnostics and Treatment Applications XVI*, Vol. 9702, International Society for Optics and Photonics (2016), p. 97020U.
11. D. G. Black, A. H. H. Hosseinabadi and S. E. Salcudean, 6-dof force sensing for the master tool manipulator of the da vinci surgical system, *IEEE Robot. Autom. Lett.* **5**(2) (2020) 2264–2271.
12. D. Bouget, R. Benenson, M. Omran, L. Riffaud, B. Schiele and P. Jannin, Detecting surgical tools by modelling local appearance and global shape, *IEEE Trans. Med. Imaging* **34**(12) (2015) 2603–2617.
13. Y.-H. Su, W. Jiang, D. Chitrakar, K. Huang, H. Peng and B. Hannaford, Local style preservation in improved gan-driven synthetic image generation for endoscopic tool segmentation, *Sensors* **21**(15) (2021) 5163.
14. K. Huang, D. Chitrakar, W. Jiang and Y.-H. Su, Enhanced u-net tool segmentation using hybrid coordinate representations of endoscopic images, *2021 Int. Symp. Medical Robotics (ISMR)* (IEEE, 2021), pp. 1–7.
15. B. Münzer, K. Schoeffmann and L. Böszörmenyi, Detection of circular content area in endoscopic videos, *Proc. 26th IEEE Int. Symp. Computer-Based Medical Systems* (IEEE, 2013), pp. 534–536.
16. Y.-H. Su, K. Huang and B. Hannaford, Real-time vision-based surgical tool segmentation with robot kinematics prior, *2018 Int. Symp. Medical Robotics (ISMR)* (IEEE, 2018), pp. 1–6.
17. I. Laina, N. Rieke, C. Rupprecht, J. P. Vizcaíno, A. Eslami, F. Tombari and N. Navab, Concurrent segmentation and localization for tracking of surgical instruments, *Int. Conf. Medical Image Computing and Computer-assisted Intervention* (Springer, 2017), pp. 664–672.
18. Y.-H. Su, I. Huang, K. Huang and B. Hannaford, Comparison of 3d surgical tool segmentation procedures with robot kinematics prior, *2018 IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)* (IEEE, 2018), pp. 4411–4418.
19. O. Özgüner, R. Hao, R. C. Jackson, T. Shkurti, W. Newman and M. C. Cavusoglu, Three-dimensional surgical needle localization and tracking using stereo endoscopic image streams, *2018 IEEE Int. Conf. Robotics and Automation (ICRA)* (IEEE, 2018), pp. 6617–6624.
20. A. Zia and I. Essa, Automated surgical skill assessment in rmis training, *Int. J. Comput. Assist. Radiol. Surg.* **13**(5) (2018) 731–739.

21. M. A. Minhem, B. Y. Safadi, H. Tamim, A. Mailhac and R. S. Alami, Does intraoperative endoscopy decrease complications after bariatric surgery? Analysis of american college of surgeons national surgical quality improvement program database, *Surg. Endosc.* **33** (11) (2019) 3629–3634.
22. D. Alaedeen, A. K. Madan, C. Y. Ro, K. A. Khan, J. M. Martinez and D. S. Tichansky, Intraoperative endoscopy and leaks after laparoscopic roux-en-y gastric bypass, *Am. Surgeon* **75**(6) (2009) 485–488.
23. A. Nimeri, A. Maasher, E. Salim, M. Ibrahim and M. Al Hadad, The use of intraoperative endoscopy may decrease postoperative stenosis in laparoscopic sleeve gastrectomy, *Obesity Surgery* **26**(7) (2016) 1398–1401.
24. F. Nageotte, C. Doignon, M. de Mathelin, P. Zanne and L. Soler, Circular needle and needle-holder localization for computer-aided suturing in laparoscopic surgery, *Medical Imaging 2005: Visualization, Image-Guided Procedures, and Display*, Vol. 5744, International Society for Optics and Photonics (2005), pp. 87–98.
25. R. Matungka, Y. F. Zheng and R. L. Ewing, Image registration using adaptive polar transform, *IEEE Trans. Image Process.* **18**(10) (2009) 2340–2354.
26. S. Iyer, T. Looi and J. Drake, A single arm, single camera system for automated suturing, *2013 IEEE Int. Conf. Robotics and Automation* (IEEE, 2013), pp. 239–244.
27. A. Kanakatte, A. Ramaswamy, J. Gubbi, A. Ghose and B. Purushothaman, Surgical tool segmentation and localization using spatio-temporal deep network, *2020 42nd Annual Int. Conf. IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE, 2020), pp. 1658–1661.
28. B. P. Lo, A. Darzi and G.-Z. Yang, Episode classification for the analysis of tissue/instrument interaction with multiple visual cues, *Int. Conf. Medical Image Computing and Computer-assisted Intervention* (Springer, 2003), pp. 230–237.
29. T. Kurmann, P. M. Neila, X. Du, P. Fua, D. Stoyanov, S. Wolf and R. Sznitman, Simultaneous recognition and pose estimation of instruments in minimally invasive surgery, *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* (Springer, 2017), pp. 505–513.
30. D. Pakhomov, V. Premachandran, M. Allan, M. Azizian and N. Navab, Deep residual learning for instrument segmentation in robotic surgery, *Int. Workshop on Machine Learning in Medical Imaging* (Springer, 2019), pp. 566–573.
31. D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. De Lange, P. Halvorsen and H. D. Johansen, Resunet++: An advanced architecture for medical image segmentation, *2019 IEEE Int. Symp. Multimedia (ISM)* (IEEE, 2019), pp. 225–2255.
32. D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen and H. D. Johansen, Doubleu-net: A deep convolutional neural network for medical image segmentation, *2020 IEEE 33rd Int. Symp. Computer-based Medical Systems (CBMS)* (IEEE, 2020), pp. 558–564.
33. D. Jha, S. Ali, K. Emanuelsen, S. A. Hicks, V. Thambawita, E. Garcia-Ceja, M. A. Riegler, T. de Lange, P. T. Schmidt, H. D. Johansen et al., Kvasir-instrument: Diagnostic and therapeutic tool segmentation dataset in gastrointestinal endoscopy, *Int. Conf. Multimedia Modeling* (Springer, 2021), pp. 218–229.
34. V. Naranjo, R. Lloréns, M. Alcañiz and F. López-Mir, Metal artifact reduction in dental ct images using polar mathematical morphology, *Comput. Methods Prog. Biomed.* **102**(1) (2011) 64–74.
35. L. A. Gorham, B. D. Rigling and E. G. Zelnio, A comparison between imaging radar and medical imaging polar format algorithm implementations, *Algorithms for Synthetic Aperture Radar Imagery XIV*, Vol. 6568, International Society for Optics and Photonics (2007), p. 65680K.
36. G. Wolberg and S. Zokai, Robust image registration using log-polar transform, in *Proc. 2000 Int. Conf. Image Processing (Cat. No.00CH37101)*, Vol. 1 (2000), pp. 493–496.
37. J. N. Sarvaiya, S. Patnaik and S. Bombaywala, Image registration using log-polar transform and phase correlation, *TENCON 2009 — 2009 IEEE Region 10 Conf.* (2009), pp. 1–5.
38. H. Araujo and J. Dias, An introduction to the log-polar mapping [image sampling], in *Proc. II Workshop on Cybernetic Vision* (1996), pp. 139–144.
39. R. Matungka, Y. F. Zheng and R. L. Ewing, Image registration using adaptive polar transform, *IEEE Trans. Image Process.* **18**(10) (2009) 2340–2354.
40. D. Sasikala and R. Neelaveni, Registration of multimodal brain images using modified adaptive polar transform, *Int. J. Video Image Process. Netw. Secur.* **10** (2010) 1–10.
41. D. Sasikala and R. Neelaveni, Image registration using modified adaptive polar transform, *Procedia Comput. Sci.* **2** (2010) 321–329.
42. R. Moccia, C. Iacono, B. Siciliano and F. Ficuciello, Vision-based dynamic virtual fixtures for tools collision avoidance in robotic surgery, *IEEE Robot. Autom. Lett.* **5**(2) (2020) 1650–1655.
43. M. Allan, S. Thompson, M. J. Clarkson, S. Ourselin, D. J. Hawkes, J. Kelly and D. Stoyanov, 2d-3d pose tracking of rigid instruments in minimally invasive surgery, *Int. Conf. Information Processing in Computer-assisted Interventions* (Springer, 2014), pp. 1–10.
44. M. K. Chmarra, C. A. Grimbergen and J. Dankelman, Systems for tracking minimally invasive surgical instruments, *Minim. Invasive Ther. Allied Technol.* **16**(6) (2007) 328–340.
45. Z. Chen, Z. Zhao and X. Cheng, Surgical instruments tracking based on deep learning with lines detection and spatio-temporal context, *2017 Chinese Automation Congress (CAC)* (2017), pp. 2711–2714.
46. O. Tonet, T. Ramesh, G. Megali and P. Dario, Tracking endoscopic instruments without localizer: Image analysis-based approach, *Studies in Health Technol. Informatics* **119** (2006) 544–549.
47. A. Krupa, J. Gangloff, C. Doignon, M. F. De Mathelin, G. Morel, J. Leroy, L. Soler and J. Marescaux, Autonomous 3-d positioning of surgical instruments in robotized laparoscopic surgery using visual servoing, *IEEE Trans. Robot. Autom.* **19**(5) (2003) 842–853.
48. M. Allan, S. Ourselin, S. Thompson, D. J. Hawkes, J. Kelly and D. Stoyanov, Toward detection and localization of instruments in minimally invasive surgery, *IEEE Trans. Biomed. Eng.* **60**(4) (2012) 1050–1058.
49. F. Qin, Y. Li, Y.-H. Su, D. Xu and B. Hannaford, Surgical instrument segmentation for endoscopic vision with data fusion of CNN prediction and kinematic pose, *2019 Int. Conf. Robotics and Automation (ICRA)* (IEEE2019), pp. 9821–9827.
50. M. Islam, D. A. Atputharuban, R. Ramesh and H. Ren, Real-time instrument segmentation in robotic surgery using auxiliary supervised deep adversarial learning, *IEEE Robot. Autom. Lett.* **4**(2) (2019) 2188–2195.
51. S. Lin, F. Qin, R. A. Bly, K. S. Moe and B. Hannaford, University of Washington sinus surgery cadaver/live dataset (uw-sinus-surgery-c/l) (2020).
52. F. Qin, S. Lin, Y. Li, R. A. Bly, K. S. Moe and B. Hannaford, Towards better surgical instrument segmentation in endoscopic vision: Multi-angle feature aggregation and contour supervision, *IEEE Robot. Autom. Lett.* **5**(4) (2020) 6639–6646.
53. W. Park and G. S. Chirikjian, Interconversion between truncated cartesian and polar expansions of images, *IEEE Trans. Image Process.* **16**(8) (2007) 1946–1955.
54. T. Y. Zhang and C. Y. Suen, A fast parallel algorithm for thinning digital patterns, *Commun. ACM* **27**(3) (1984) 236–239.
55. T. E. Schouten and E. L. Van den Broek, Fast exact euclidean distance (feed): A new class of adaptable distance transforms, *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(11) (2014) 2159–2172.
56. A. Anand, S. S. Tripathy and R. S. Kumar, An improved edge detection using morphological Laplacian of Gaussian operator, *2015 2nd Int. Conf. Signal Processing and Integrated Networks (SPIN)* (IEEE, 2015), pp. 532–536.
57. O. Pârvu and D. Gilbert, Implementation of linear minimum area enclosing triangle algorithm, *Comput. Appl. Math.* **35**(2) (2016) 423–438.

58. P. K. Gadosey, Y. Li, E. A. Agyekum, T. Zhang, Z. Liu, P. T. Yamak and F. Essaf, Sd-unet: Stripping down u-net for segmentation of biomedical images on platforms with low computational budgets, *sDiagnostics* **10**(2) (2020) 110.

Kevin Huang received the B.S. degree in Engineering and Mathematics from Trinity College, Hartford, CT, USA in 2012. He received his M.S. degree in Electrical Engineering and his Ph.D. degree in Electrical Engineering with a concentration in Systems, Controls and Robotics from the University of Washington, Seattle, WA in 2015 and 2017, respectively. During graduate studies, Dr. Huang received the National Science Foundation's Graduate Research Fellowship. Currently he is an Assistant Professor in the Department of Engineering at Trinity College, where he is active in robotics outreach and undergraduate engineering education. He is an Affiliate Professor at the University of Washington and a Research Associate at Mount Holyoke College, and serves as the Chair for the IEEE Connecticut Robotics and Automation Society. Dr. Huang's research interests include haptic virtual fixtures, telerobotics, surgical robotics, and human robot interaction.

Digesh Chitrakar is an undergraduate student at Trinity College in the Department of Engineering and Department of Mathematics. Digesh has been a research assistant in the Perceptual-robotics and Automation (PandA) Laboratory since February 2019. His research interests include surgical robotics, robot manipulation, and under-actuated robotics.

Wenfan Jiang received the B.A. degree from Mount Holyoke College in Mathematics and Computer Science. She has done several research projects in machine learning since June of 2020, focusing on generated adversarial networks (GANs). She has a paper about image augmentation by GANs. She presented in the 2021 Tapia conference and received the 2nd award for graduated students. Her research interests include surgical robotics, machine learning algorithms, and computer vision.

Isabella Yung is an undergraduate student at Trinity College receiving a B.S. in Engineering and Physics. She has been a research assistant in the Perceptual-Robotics and Automation Laboratory since February 2019. She recently completed an internship with ABB robotics where she assisted in developing ABB's first collaborative welding robot. Her research interests include robot manipulation and haptic feedback.

Yun-Hsuan Su received the B.S. degree with concentration in Systems and Controls in Electrical and Computer Engineering from National Chiao Tung University, Hsinchu, Taiwan in 2016. She received her graduate degrees, the M.S. in Electrical Engineering and Ph.D. in Electrical and Computer Engineering, with focus on Systems, Controls and Robotics from the University of Washington, Seattle, WA in 2017 and 2020, respectively. In 2018, Dr. Su was a research engineer and worked on visual and force servoing for industrial robots at ABB robotics, leading to two patent applications. She is passionate about outreach STEM programs, has been closely involved in IEEE TryEngineering, and has organized and led multiple summer robotics camps. Currently she is an Assistant Professor in the Department of Computer Science at Mount Holyoke College. She is a member of Tau Beta Pi, and was nominated for the Yang Award for Outstanding Doctoral Student. Dr. Su's research interests span surgical robotics, vision-based force estimation, computer/machine vision, and haptic feedback.