

# Markerless Ultrasound Probe Pose Estimation in Mini-Invasive Surgery

Mohammad Mahdi Kalantari, Erol Ozgur, Mohammad Alkhatib,  
Emmanuel Buc, Bertrand Le Roy, Richard Modrzejewski,  
Youssef Mezouar, Adrien Bartoli

Clermont Auvergne INP, Institut Pascal, Clermont-Ferrand, France.  
University Hospital of Clermont-Ferrand, France.  
University Hospital of Saint-Etienne, France.

## Abstract

In mini-invasive surgery, the laparoscopic ultrasound probe is visible in the laparoscopic image. We address the problem of estimating the probe pose with respect to the laparoscope without using markers and additional sensors. We propose the first method using a single standard laparoscopic monocular RGB image. It is robust, initialization-free and runs at **10** fps, thus forming a promising tool to improve robotic and augmented reality-based surgery.

## 1 Introduction

Mini-invasive surgery (MIS) has important advantages over open surgery. It reduces the severity of postoperative complications, the length of hospital stay, and healthcare costs [1]. MIS however does not allow the surgeon to palpate the organ. Consequently, the localization of tumors internal to the organs forms a surgical challenge. Laparoscopic ultrasound (LUS) is one of the important intraoperative tools to help localize such tumors. In particular, LUS forms the gold standard for liver and kidney surgery. However, LUS has drawbacks, including a long learning curve and operator dependence. Two applications were proposed to mitigate these drawbacks: robotically-controlled ultrasound (RobUS) and augmented reality (AR) based guidance [2, 3]. RobUS relieves the surgeon from manually maneuvering the LUS probe to find and track the tumor; AR-based guidance automatically fuses the LUS images into the

laparoscopic images. Both applications are very promising in terms of surgical assistance; however, they crucially depend on the ability to quickly and accurately solve the **LUS pose problem**, which is to estimate the position and orientation of the LUS probe in the laparoscope’s coordinate frame, or equivalently the rigid body motion between the LUS probe and the laparoscope. Specifically, most LUS probes include an articulated joint connecting a **shaft** to the transducer **head** and the ultrasound probe pose refers to the head’s.

The LUS pose problem is challenging, owing to the operating room (OR) conditions, with four main reasons. First, the laparoscope is the only sensor and it provides monocular RGB images. Second, the LUS probes are usually textureless, shiny, and partially occluded. This makes it difficult to find features in the laparoscopic images that strongly constrain the pose. Third, an approach using markers, optical or electromagnetic tracking would require one to introduce new elements or sensors in the abdominal cavity, which is highly impractical and hardly feasible for obvious sterilization and safety reasons. They would also increase the OR clutter and the risk of occlusions by the staff and other devices. Fourth, surgical robots with cable-driven actuation do not allow one to estimate LUS pose reliably via their internal encoders. This is because of the inherent inaccuracy of the readings [4] and the error accumulation of encoder values over time [5]. Even with an accurate kinematic model of the robot and tools, drop-in US probes are grasped differently each time by the robot’s gripper.

We set four requirements for a practical LUS pose method: it should be free of *(i)* markers, *(ii)* extra sensors, *(iii)* (re)initialization, and should *(iv)* run in real-time in any OR. As for the vast majority of pose estimation problems, existing LUS pose estimation methods can be broadly split into two categories: *(C1)* methods requiring new elements or sensors to be integrated into the OR, and *(C2)* methods using only the standard laparoscopic monocular RGB images. Methods in category *C1* use either an off-the-shelf tracking device requiring special equipment such as electromagnetic trackers and optical motion capture systems [6, 7], or marker-based vision techniques [8]. Methods in this category thus fail requirements *(i)* or *(ii)*; yet, they form the state-of-the-art in LUS pose estimation. On the other hand, methods in category *C2* are promising, as they generally fulfill the four requirements for a practical solution by construction. Yet, this category is still void of existing methods for LUS pose.

We propose the first method in category *C2* for LUS pose estimation complying with the four requirements. Our method uses the observation that most existing commercial LUS probes have cylindrical shaft and head shapes and hemispherical tips. These include Fujifilm L44LA, Siemens Acuson P300 LP323, Philips L10-4lap, SonoScape LAP7, and Aloka UST-5418. We model the LUS probe mathematically as a straight homogeneous spherical cylinder (SHSC) segment, which is also known as a **spherocylinder** [16]. We define the LUS pose as the position of the tip’s hemispherical center and as the direction vector of the head’s cylinder axis. It is obvious that, owing to the symmetry of the spherocylinder, a 1 DoF rotational ambiguity exists on the pose, as an unrecoverable angle about the cylinder axis. From this model, we propose a vision-based solution method named **LUP** (Laparoscopic Ultrasound Pose). Our method takes as input the laparoscopic image, the laparoscope calibration and the

**Table 1** Related work in image-based surgical tool pose estimation.

Reference	Tool	#DoF	Initialization-free	Single image	FPS	Marker-free	Additional sensors
Sarikaya et al.[9]	Surgical tools*	2**	✓	✓	10	✓	None
Laina et al.[10]	Surgical tools*	2**	✓	✓	18	✓	None
Kurmann et al.[11]	Surgical tools*	2**	✓	✓	9	✓	None
Ye et al.[12]	Surgical tools*	6	×	×	29	✓	Robot encoders
Allan et al.[5]	Surgical tools*	6	×	×	0.9	✓	None
Allan et al.[13]	Surgical tools*	6	×	×	3	✓	None
Doignon et al.[14]	General cylinder	4	✓	✓	×	✓	None
Zenteno et al.[15]	Fiberscope	5	×	×	†	✓	None
LUP (proposed)	LUS	5	✓	✓	10	✓	None

\* Passive tools not comprising sensors

\*\* In 2D image space

† Just mention “real-time” but do not report FPS

spherocylinder radius and returns the LUS pose automatically. It relies on two main technical contributions:

- A well-engineered algorithmic pipeline, including image segmentation and robust model estimation.
- A novel minimal formulation for the reconstruction of a spherocylinder from its occluding contour.

We provide numerous experiments showing that LUP respects the four requirements of a practical solution and thus has the potential to be directly deployed to standard ORs.

## 2 Related Work

We review related work in vision-based surgical tool pose estimation and sphere and cylinder reconstruction.

### 2.1 Vision-based Surgical Tool Pose Estimation

Tracking the surgical tools of the da Vinci robot in 2D images is well-studied [9–11]. For tool pose estimation, [12] incorporates robot joint values and achieves a fast run-time. [5, 13] use optical flow to achieve 3D tool tracking. These works are for specific tools and inapplicable to LUS. For tools closer to an LUS shape, [14] includes a markerless method but only for cylinders, lacking the tip position. [15] finds the pose of a cylindrical fiberscope by maximizing the similarity of binary masks based on [17]. This method is applicable to LUS, but is overly sensitive to the quality of the masks. These works are summarized in Table.1. Finally, [18–21] register LUS and laparoscopic images to the preoperative CT model. While beneficial for AR purposes and indirectly providing the LUS pose, they are complex and unlikely to run in real-time. In contrast, the proposed LUP method simply requires an RGB image as input

and makes simple, yet general, assumptions on the LUS probe shape. It is thus widely applicable in standard ORs.

## 2.2 Sphere and Cylinder Reconstruction

Both the cylinder and sphere are quadric surfaces, which can be reconstructed from images. There exist methods exploiting multiple images such as [22], but we focus on single image methods. Existing methods follow two steps for both the cylinder and sphere: 1) occluding contour detection and 2) shape reconstruction from the occluding contours. For the cylinder, the occluding contour is a pair of lines [23–25], from which the cylinder axis can be reconstructed and also its depth, should the radius be given. For the sphere, the occluding contour is in general an ellipse, which can be estimated following [26], from which the sphere can be reconstructed given the radius [27]. The main shortcoming of these methods is their computational requirements. In particular, at least 5 points are needed to reconstruct the sphere center for a known radius, making it impractical to use in a random sampling robust method such as RANSAC, as the number of required samples will be prohibitively large. A very recent work [28] estimates the sphere directly without ellipse fitting. The minimal case for this method requires only 3 points and thus speeds up RANSAC substantially. In contrast, we model the LUS pose using a hemisphere-cylinder pair forming a spherocylinder shape. We propose a new minimal formulation for the hemisphere reconstruction from a single view, which only requires 2 points from the hemisphere silhouette, constraining the solution by the 2 silhouette lines of the cylindrical part. Our minimal formulation thus brings a further speed boost to RANSAC.

## 3 Methodology

### 3.1 Problem Statement

Given a monocular RGB image  $\mathcal{I}_{LAP}$  of a calibrated laparoscope and the spherocylinder LUS head radius  $r$ , we estimate the pose  $\xi_{LUS}$  of the LUS probe seen in  $\mathcal{I}_{LAP}$ .

#### Notation and modelling

We define  $\mathbf{s}_i \in \mathbb{R}^2$ ,  $\mathbf{c}_j \in \mathbb{R}^2$  and  $\mathbf{o}_k \in \mathbb{R}^2$  as the occluding contour points of the hemisphere and cylinder, and other points, respectively. The point indices count up  $i = 1 \dots n \in \mathbb{N}$ ,  $j = 1 \dots m \in \mathbb{N}$  and  $k = 1 \dots t \in \mathbb{N}$ . The initial set of unclassified image contour points is thus  $\mathcal{P} = \{\mathbf{s}_i, \mathbf{c}_j, \mathbf{o}_k \mid \forall i, j, k\}$ . We denote the laparoscope’s intrinsic parameter matrix as  $\mathbf{K} \in \mathbb{R}^{3 \times 3}$  and the lens distortion parameters with set  $\mathcal{D}$ . These parameters are to project a 3D point in the camera coordinate frame  ${}^c\mathbf{X} \in \mathbb{R}^3$  to the image point  $\mathbf{x} \in \mathbb{R}^2$ . Equivalently,  $\bar{\mathbf{x}} \propto \mathbf{K} {}^c\mathbf{X}$  where  $\bar{\mathbf{x}} = [\mathbf{x}^\top, 1]^\top$  is the image point in homogeneous coordinates. The left superscript  $c$  indicates the camera coordinate frame. A line  $\ell \in \mathbb{R}^3$  in Hesse normal form on the image plane holds  $\ell^\top \bar{\mathbf{x}} = 0$  for any of its points  $\bar{\mathbf{x}}$ . The backprojection of line  $\ell \in \mathbb{R}^3$  is a plane whose unit normal vector is  ${}^c\mathbf{m} = \mathbf{K}^\top \ell / \|\mathbf{K}^\top \ell\|$ . The occluding contour of the LUS head’s cylinder forms a pair of image lines denoted  $\ell_+$  and  $\ell_-$ , oriented outwards from the

cylinder. We denote the normals of the backprojection planes of these lines as  ${}^c\mathbf{m}_+$  and  ${}^c\mathbf{m}_-$ . The backprojection of an image point  $\mathbf{x} \in \mathbb{R}^2$  is a ray that starts from the optical center and passes through this image point. We denote the direction of the ray with a unit vector  ${}^c\mathbf{x} = {}^c\mathbf{x} / \|{}^c\mathbf{x}\|$  where  ${}^c\mathbf{x} = \mathbf{K}^{-1} \bar{\mathbf{x}}$ . We denote the hemisphere center of the spherocylinder as  ${}^c\mathbf{H}$  and the direction of the ray which passes through it as  ${}^c\mathbf{h}$ . We denote the cylinder axis direction of the spherocylinder as  ${}^c\mathbf{u}$ . Finally, we assemble the LUS pose into a vector  $\xi_{LUS} = [{}^c\mathbf{H}^\top, {}^c\mathbf{u}^\top]^\top \in \mathbb{R}^{6 \times 1}$  with  $\|{}^c\mathbf{u}\| = 1$ . Figure 1 shows the spherocylindrical LUS model.

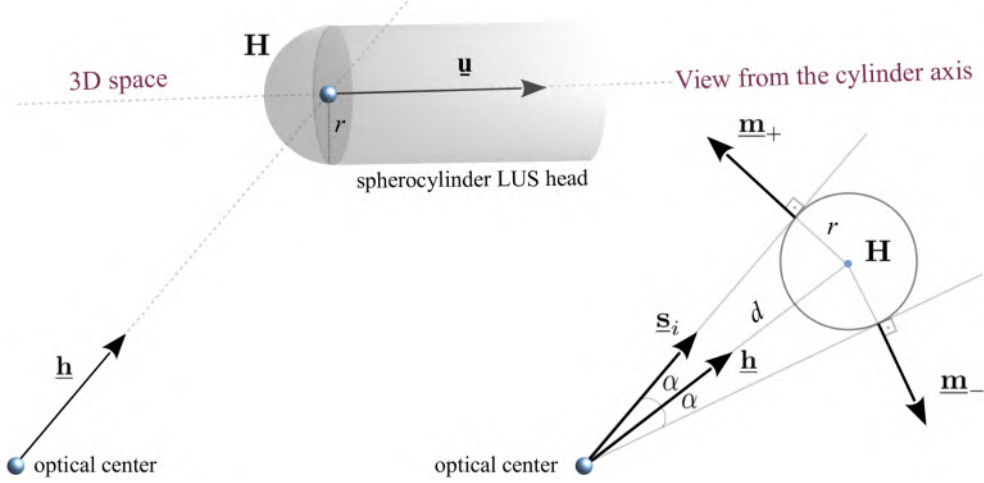


Fig. 1 Proposed LUS head geometry from two viewpoints.

## Solution pipeline

Our pipeline has two main steps: (i) LUS probe segmentation and contour extraction, and (ii) pose estimation from unclassified contour points, which we explain in the next two sections. Both steps represent challenging problems.

### 3.2 LUS Probe Segmentation and Contour Extraction

This step is a semantic segmentation problem, which we address with deep learning (DL) [29]. We trained and compared the state-of-the-art neural networks, including UNet [30, 31], TeraNet [32] and LinkNet [33]. LinkNet-34 used in [34] provided the best results. First, we sampled laparoscopic images from laparoscopic liver surgery collected in our hospital with prior ethical approval. Second, we manually segmented the LUS probes in these images to generate the ground-truth masks, with segmentation validation by two medical experts. Intentionally, we selected a low quantity of images (132) from Surgery 1 for training to demonstrate the robustness of our algorithm toward improper segmentation. Then we applied data augmentation using the Albumentations library. Third, we trained the network and evaluated their accuracy

with a different surgery (144 images from Surgery 5) with the Intersection over Union (IoU) metric in test images. LinkNet-34 achieved the highest accuracy, at 87%. The bottom-left image in figure 2 shows a qualitative result. We extract the LUS probe’s contour points from the segmentation mask using [35]. Our experiments in section 4.2 use laparoscopic images from 5 other surgeries.

### 3.3 Pose Estimation from Unclassified Contour Points

We use sampling-based robust estimation from a minimal solution.

#### 3.3.1 Minimal solution for hemisphere reconstruction

Our solution method requires 2 points  $\{\mathbf{s}_1, \mathbf{s}_2\}$  on the hemisphere silhouette and 2 silhouette lines  $\{\ell_+, \ell_-\}$  on the cylinder. The points backprojections are the rays with directions  $\{{}^c\mathbf{s}_1, {}^c\mathbf{s}_2\}$ . The lines backprojections are the planes with normals  $\{{}^c\mathbf{m}_+, {}^c\mathbf{m}_-\}$ . The two planes are tangent to the hemisphere and to the cylinder of the spherocylinder. Given the spherocylinder radius  $r$ , the tangent planes constrain the hemisphere center and the cylinder axis to lie on a 3D line  $\mathcal{L}$  parallel to the 3D intersection line of the two tangent planes. While line  $\mathcal{L}$  fully constrains the cylinder axis, it leaves 1 translational DoF to the hemisphere center. Fortunately, a ray from the hemisphere silhouette is also tangent to the hemisphere, which further constrains the hemisphere center on the line  $\mathcal{L}$ . Specifically, while a single tangent ray leaves two possible solutions, adding the second tangent ray fully resolves the hemisphere center.

Concretely, any image point’s ray direction  ${}^c\mathbf{s}_i$  from the hemisphere silhouette holds the following relation due to the rotational symmetry:

$${}^c\mathbf{s}_i^\top {}^c\mathbf{h} = \cos(\alpha) \quad (1)$$

where  ${}^c\mathbf{h}$  is the direction of the ray passing through the hemisphere center, and  $\alpha$  is an unknown angle encoding the hemisphere center depth as  $d = r/\sin(\alpha)$ . Similarly, any image line  $\ell$  tangent to the hemisphere silhouette defines a plane whose outward normal  ${}^c\mathbf{m}$  holds the following relation due the rotational symmetry:

$${}^c\mathbf{m}^\top {}^c\mathbf{h} = \cos\left(\frac{\pi}{2} + \alpha\right) \quad (2)$$

Using equations (1) and (2), we write the following minimal linear system:

$$\underbrace{[{}^c\mathbf{s}_1 + {}^c\mathbf{m}_+, {}^c\mathbf{s}_1 + {}^c\mathbf{m}_-, {}^c\mathbf{s}_2 + {}^c\mathbf{m}_+]}_{\mathbf{Q}_{3 \times 3}}^\top {}^c\mathbf{h} = \lambda \mathbf{1}_{3 \times 1} \quad (3)$$

where  $\lambda = \cos(\alpha) + \cos(\frac{\pi}{2} + \alpha)$  and  $\mathbf{Q}_{3 \times 3}$  is the minimal design matrix. Solving (3) for  ${}^c\mathbf{h}$  and  $\alpha$  leads to the hemisphere center  ${}^c\mathbf{H}$ , as shown in Algorithm 1. An  $n$ -point overdetermined case with  $n \geq 3$  has the design matrix:

$$\mathbf{Q}_{3 \times 2n} = [{}^c\mathbf{s}_1 + {}^c\mathbf{m}_+, {}^c\mathbf{s}_1 + {}^c\mathbf{m}_-, \dots, {}^c\mathbf{s}_n + {}^c\mathbf{m}_-] \quad (4)$$

---

**Algorithm 1: 2-point hemisphere fit**

---

**Inputs:** ray directions  $\{^c\mathbf{s}_1, ^c\mathbf{s}_2\}$  on hemisphere silhouette, plane normals  $\{^c\mathbf{m}_+, ^c\mathbf{m}_-\}$  on cylinder silhouette, spherocylinder radius  $r$ .  
**Output:** Hemisphere center  $^c\mathbf{H} \in \mathbb{R}^3$ .

```
/* Design matrix */
1  $\mathbf{Q} = [^c\mathbf{s}_1 + ^c\mathbf{m}_+, ^c\mathbf{s}_1 + ^c\mathbf{m}_-, ^c\mathbf{s}_2 + ^c\mathbf{m}_+]_{3 \times 3}$ 
/* Ray direction to hemisphere center */
2  $^c\mathbf{h} = ^c\mathbf{h} / \|\mathbf{h}\|$  where  $^c\mathbf{h} = \mathbf{Q}^{-\top} \mathbf{1}_{3 \times 1}$ 
/* Depth */
3  $d = r / \sin(0.5 \arcsin(1 - 1/\|\mathbf{h}\|^2))$ 
/* Hemisphere center */
4  $^c\mathbf{H} = d ^c\mathbf{h}$ 
```

---

and solved in the least-squares sense using the pseudo-inverse  $(\mathbf{Q}^\top)^\dagger$  for the scaled ray direction vector as:

$$^c\mathbf{h} = (\mathbf{Q}^\top)^\dagger \mathbf{1}_{2n \times 1} \quad (5)$$

The  $n$ -point hemisphere reconstruction is obtained by replacing the expressions in lines 1 and 2 of Algorithm 1 with equations (4) and (5), respectively.

### 3.3.2 Principal algorithm

We give the main LUP method in Algorithm 2, with a complete pipeline including *LUS probe segmentation and contour extraction* in line 1. It continues with distortion correction of the unclassified contour points  $\mathcal{P}$  in line 2 and then reaches the main part, in lines 3 to 10. We estimate the pose and classify the contours simultaneously through sequential RANSACs. Each RANSAC internally iterates with a minimal solution method on random sample sets and terminates with a best-fit solution on the largest consensus set. Once a RANSAC ends, the contour points associated with its best-fit solution are removed from  $\mathcal{P}$ . The following RANSACs thus operate on fewer points, which accelerates computation. More explicitly, line 3 fits an image line to set  $\mathcal{P}$  and classifies the line's consensus set as one of the cylinder edges. Line 4 removes the line's points from  $\mathcal{P}$ . Line 5 fits a second image line to the updated set  $\mathcal{P}$  and classifies the line's consensus set as the second cylinder edge. Line 6 removes the line's points from  $\mathcal{P}$ . Line 7 orients the fitted image lines outward from the cylinder silhouette. Line 8 removes the contour points outside of the cylinder's silhouette. Line 9 finds a consensus set of the hemisphere silhouette using the 2-point solution from Algorithm 1 and terminates with the least-squares  $n$ -point solution on the consensus set. Line 10 computes the cylinder's direction vector  $^c\mathbf{u}$  from the cylinder's silhouette lines, choosing its orientation as shown in figure 1. Line 11 assembles the pose.

## 4 Experimental Results

We report a quantitative sensitivity analysis of LUP on simulated data and then qualitative results on five surgeries.

---

**Algorithm 2:** LUP – Laparoscopic Ultrasound Pose

---

**Inputs:** *Laparoscopic image*  $\mathcal{I}_{LAP}$ , *spherocylinder radius*  $r$ , *camera intrinsics*  $\{\mathbf{K}, \mathcal{D}\}$

**Output:** *LUS pose*  $\xi_{LUS}$

- 1  $\mathcal{P} = \text{ExtractContour}(\text{SegmentLUS}(\mathcal{I}_{LAP}))$
  - 2  $\mathcal{P} = \text{UndistortContour}(\mathcal{P}, \mathcal{D})$
  - 3  $\{\ell_1, \mathcal{P}_{\ell_1}\} = \text{RANSACLine}(\mathcal{P})$  where  $\mathcal{P}_{\ell_1} \subset \mathcal{P}$
  - 4  $\mathcal{P} = \text{RemovePoints}(\mathcal{P}, \mathcal{P}_{\ell_1})$
  - 5  $\{\ell_2, \mathcal{P}_{\ell_2}\} = \text{RANSACLine}(\mathcal{P})$  where  $\mathcal{P}_{\ell_2} \subset \mathcal{P}$
  - 6  $\mathcal{P} = \text{RemovePoints}(\mathcal{P}, \mathcal{P}_{\ell_2})$
  - 7  $\{\ell_+, \ell_-\} = \text{OutwardFromCylinder}(\ell_1, \ell_2)$
  - 8  $\mathcal{P} = \text{RemovePointsOutOfCylinder}(\mathcal{P}, \ell_+, \ell_-)$
  - 9  $\{^c\mathbf{H}, \mathcal{P}_h\} = \text{RANSACHemisphere}(\mathcal{P}, \ell_+, \ell_-, \mathbf{K}, r)$
  - 10  $^c\mathbf{u} = \text{ComputeCylinderDirection}(\ell_+, \ell_-, \mathbf{K})$
  - 11  $\xi_{LUS} = [^c\mathbf{H}^\top, ^c\mathbf{u}^\top]^\top$
- 

## 4.1 Sensitivity Analysis

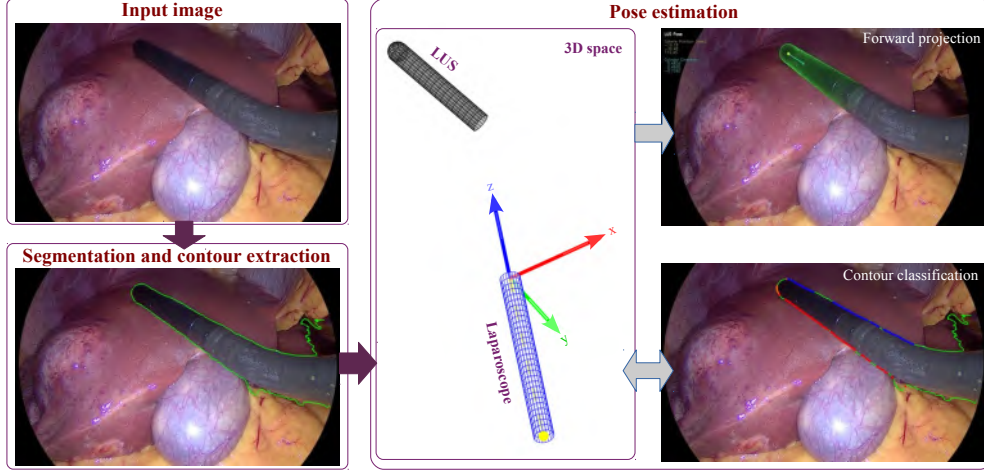
We use Monte Carlo simulation. We first create a set of ground-truth LUS poses and generate their occluding contours, which we perturb with several pre-generated noise schemes. We then estimate LUS probe pose from the perturbed data. Finally, we measure the difference between the estimated and ground-truth poses.

### 4.1.1 Perturbation

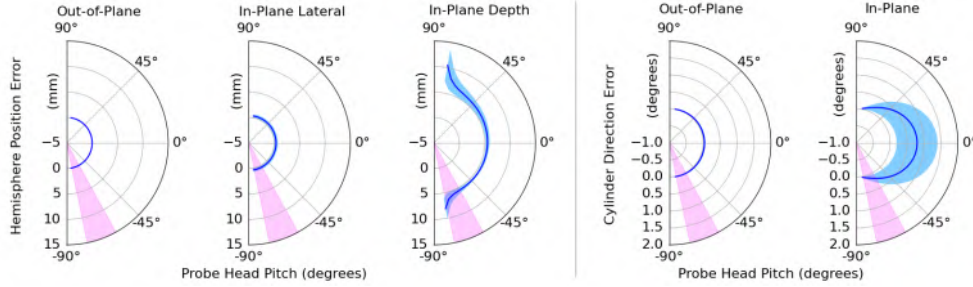
The inputs to Algorithm 1 are two lines and tip pixels. We perturb these inputs by adding noise drawn from a distribution formed from the manual and neural network segmentations of real laparoscopic images.

For the line noise distribution, we fit lines to the manual segmentation and compare them to lines fitted on the neural network segmentation. We first represent these lines by their backprojected plane normal vector, which we express in spherical coordinates under the ISO 80000-2:2019 convention. The error thus appears on the spherical coordinates  $\theta$  and  $\phi$ . Finally, we observe that the noise PDF is well approximated by zero-mean Gaussian distributions with  $\sigma = 1.5 \times 10^{-4}$  radians for  $\theta$  and  $\sigma = 1.2 \times 10^{-3}$  radians for  $\phi$ . For tip pixel noise distribution, we take a similar approach. First, we find the inlier pixels of the hemisphere for the manual and the neural network segmentations. From the inliers of the manual segmentation, we project the reconstructed hemisphere onto the image, forming an ellipse. Finally, we compute the Euclidean distance between the inliers of the neural network segmentation and the ellipse to form the error set. However, even though the first two moments of the error distribution are  $\mu = 3$  pixels and  $\sigma = 2.5$  pixels, the distribution is not Gaussian. Therefore, we keep the samples to form a non-parametric noise distribution. As the generated noise is obtained from improper segmentation results, it is generally overestimated and we expect the real noise to have a lower magnitude.





**Fig. 2** LUP estimated the probe’s pose successfully from a single laparoscopic liver surgery image. LUP also classified the contour points as the hemisphere (orange), the cylinder (blue and red), and the outliers (green). The forward projection of the LUS head model onto the image plane yields the  $\text{rmse}$  of 1.18 pixels with respect to the classified contour points. This  $\text{rmse}$  conforms to the chosen standard deviation  $\sigma = 1.25$  pixels of a zero mean Gaussian measurement error model considered in the formation of the consensus sets of the robust pose estimation algorithm.



**Fig. 3** Mean absolute errors and standard deviations versus LUS head’s pitch in the laparoscope coordinate frame. The rotation axis indicates the angle between the LUS and the  $xy$  plane of the laparoscope. The magenta shaded regions show the LUS head pitch observed in the accompanying video.

#### 4.1.2 Error Metrics

We measure the estimation error with 5 metrics for the estimated 5DoF LUS pose. First, we define a reference *plane* formed by the camera center and the axis of the LUS head cylinder. The *plane* normal is denoted by  $\underline{\mathbf{m}}$ . We then define an orthonormal coordinate frame with axes  $(\underline{\mathbf{m}}, \underline{\mathbf{h}}, \underline{\mathbf{v}})$  where  $\underline{\mathbf{v}} = \underline{\mathbf{m}} \times \underline{\mathbf{h}}$  and  $\underline{\mathbf{h}}$  as in figure 1. For the hemisphere position error, we name the projected error onto  $\underline{\mathbf{m}}$  as *Out-of-Plane*, onto  $\underline{\mathbf{h}}$  as *In-Plane Depth*, and onto  $\underline{\mathbf{v}}$  as *In-Plane Lateral*. We also quantify the LUS head cylinder’s axis direction errors in this orthonormal coordinate frame. We define *Out-of-Plane* and *In-Plane* errors as the angles between the estimated cylinder axis of the LUS head and the  $hv$ -plane and  $um$ -plane, respectively.

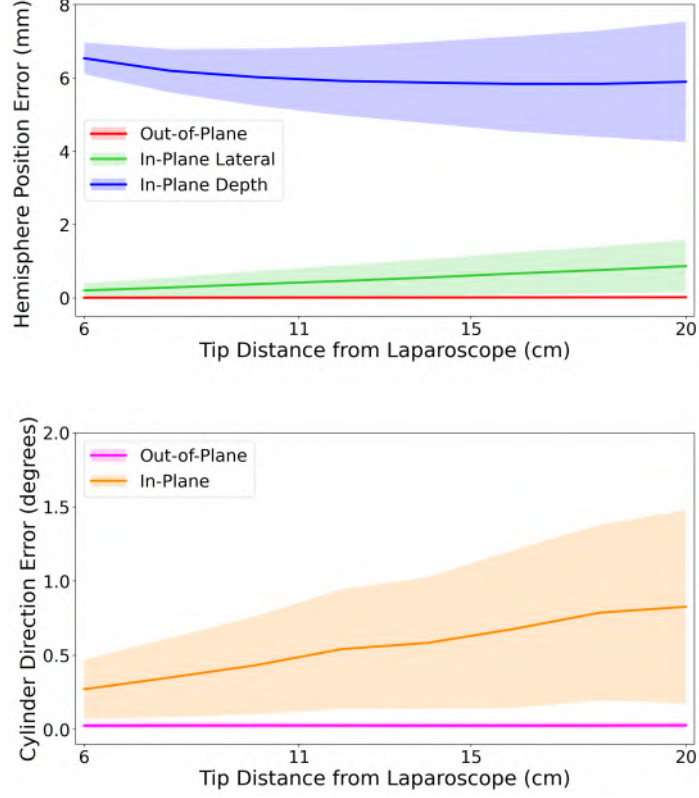
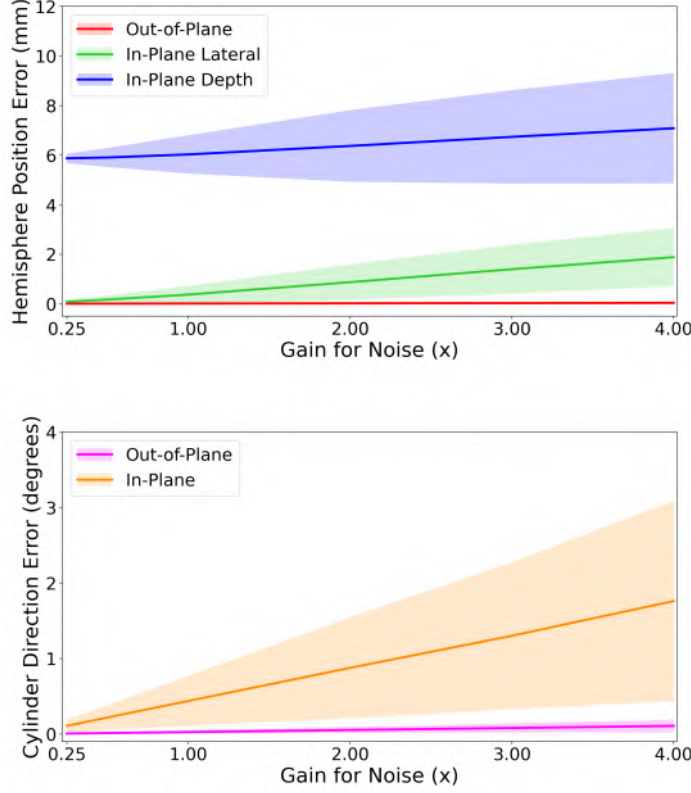


Fig. 4 LUP's accuracy versus depths.

#### 4.1.3 Results

We first investigate the sensitivity in LUS head orientation. We have simulated the LUS tip at the image center and 10 *cm* away from the laparoscope and set the LUS head to rotate around the *y*-axis of the laparoscope, with the conventional camera coordinate system shown in figure 2. For each orientation, the Monte Carlo simulation is run 10,000 times and noise is applied. In each iteration, the ellipse from the hemisphere LUS tip is uniformly divided into 50 points, and 10 of them are randomly selected. In real data, the average number of inliers is around 30 points. The results are shown in figure 3, indicating that the errors on the estimated LUS head direction and the out-of-plane errors on the estimated hemisphere position are negligible. The in-plane lateral errors on the estimated hemisphere position are sub-millimetric. The in-plane depth errors on the estimated hemisphere position are higher and increase with higher LUS head pitch. For a robotic surgery where both the laparoscope and LUS are held by the robot, figure 3 suggest relative pose constraint which could further improve the accuracy of LUP. The next analysis varies the depth of the LUS tip, as varying other orientations of the LUS does not provide new situations. From the hemisphere position parameters, depth has the strongest effect on estimation. Figure 4 shows

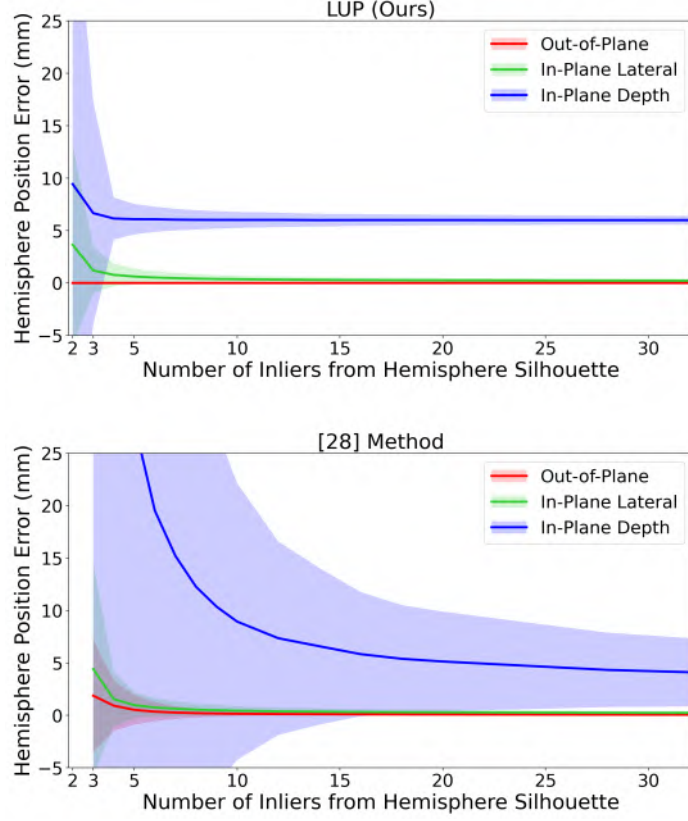


**Fig. 5** LUP's accuracy versus noise levels.

varying depths of LUS in a practical range of laparoscopic surgery. To run the relevant simulation, parameters are similar to the former test, and the error is averaged over all pitch values as a random variable. As observed, the uncertainty of in-plane depth and in-plane lateral components of the hemisphere position increases with depth. The next analysis is the same as the first analysis in figure 3, except that the noise is multiplied with a varying gain. The results are shown in figure 5 which are averaged over the different pitch values. We observe that the accuracy of LUP scales linearly. As the final analysis, we evaluate the accuracy of hemisphere position estimation with respect to the number of inliers from the hemisphere silhouette and compare it with the recent method [28]. As observed in figure 6, a smaller number of inliers provides a larger error but beyond 12 inliers, there is no strong enhancement. Compared to [28], LUP's hemisphere position reconstruction outperforms in the minimum number of required points and accuracy.

## 4.2 Experiments with Clinical Data

We applied LUP on laparoscopic images from 5 different liver surgeries. Figures 2 and 7 show qualitative LUP results. We observe that the projected LUS head model from the



**Fig. 6** LUP and [28] hemisphere reconstructions versus number of inliers in hemisphere silhouette.



**Fig. 7** Forward projections of the LUS head model onto the five different surgery images from their estimated poses by LUP.

estimated LUS poses yields visually perfect alignments with the input images. In this experiment, LUP runs at 10 fps (65 ms for image segmentation and contour extraction, and 35 ms for pose estimation). This is fast enough considering that the LUS probe moves slowly in search of the tumor.

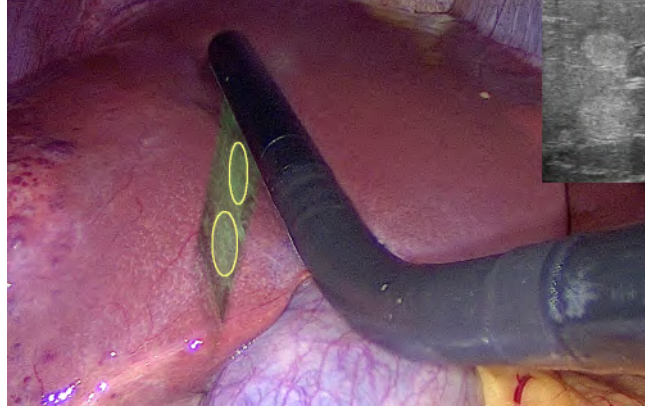
### 4.3 The 6<sup>th</sup> DoF Estimation and AR Application

Most LUS probes have 4-way (*i.e.*, left, right, up and down) articulation for the transducer head. We make the observation that *the shaft axis and the transducer head axis of the LUS probe form a plane  $\Pi_L$ , when only an up-or-down displacement is applied*

on the 4-way articulation. In such a LUS probe posture, by construction, the head transducer’s imaging plane  $\Pi_L$  overlaps the plane  $\Pi_L$ . It follows that, given the LUS pose by LUP and the LUS shaft’s cylinder axis direction in the laparoscope coordinate frame, we can estimate the pose of the LUS imaging plane  ${}^c\Pi_L$ . Any of the following scenarios allows us to recover the LUS shaft’s cylinder axis direction: scenario-(i) when the LUS shaft is also visible in the laparoscopic image; or scenario-(ii) when both the LUS and the laparoscope are controlled by robots providing their relative pose. We exploit scenario-(i) on one of the liver laparoscopy images to showcase an AR application using standard LUS calibration [8] in figure 8. We name it LARLUS (laparoscopic augmented reality from laparoscopic ultrasound). LARLUS augments the LUS image on the laparoscopy image to reveal the subsurface tumors.

## 5 Conclusion

We have proposed LUP, the first LUS pose method which works from a standard laparoscopic image, free of markers, extra sensors, and (re)initialization. LUP is robust, accurate, and fast, as shown from numerous experiments, with pose errors lower than 1 cm and an unoptimized runtime of 10 fps. It thus forms an essential tool to facilitate robotic and AR-based surgical applications. We have built LARLUS, an AR-based guidance application, using LUP, which shows the subsurface tumors on the laparoscopy images directly. As future work, we will (i) improve LUP segmentation step for faster and better performance, (ii) build a LUS pose tracking algorithm by imposing new temporal constraints on LUP’s outputs, and (iii) develop RobUS2T (robotically-controlled LUS probe for tumor tracking) for automatic tumor searching and tracking with multimodal visual feedback.



**Fig. 8** An instance of LARLUS output. The LUS image is augmented on the laparoscopy image with the manually segmented bounding ellipses of tumors. The original LUS image is also overlaid on the top-right corner.

## References

- [1] Jaffray, B.: Minimally invasive surgery. *Archives of Disease in Childhood* **90**(5), 537–542 (2005) <https://doi.org/10.1136/adc.2004.062760>
- [2] Mohareri, O., Schneider, C., Adebar, T.K., Yip, M.C., Black, P., Nguan, C.Y., Bergman, D., Seroger, J., DiMaio, S., Salcudean, S.E.: Ultrasound-based image guidance for robot-assisted laparoscopic radical prostatectomy: Initial in-vivo results. In: Barratt, D., Cotin, S., Fichtinger, G., Jannin, P., Navab, N. (eds.) *Information Processing in Computer-Assisted Interventions*, pp. 40–50. Springer, Berlin, Heidelberg (2013)
- [3] Singla, R., Edgcumbe, P., Pratt, P., Nguan, C., Rohling, R.: Intra-operative ultrasound-based augmented reality guidance for laparoscopic surgery. *Healthcare Technology Letters* **4**(5), 204–209 (2017) <https://doi.org/10.1049/htl.2017.0063>
- [4] Bouget, D., Allan, M., Stoyanov, D., Jannin, P.: Vision-based and marker-less surgical tool detection and tracking: a review of the literature. *Medical Image Analysis* **35**, 633–654 (2017) <https://doi.org/10.1016/j.media.2016.09.003>
- [5] Allan, M., Chang, P.-L., Ourselin, S., Hawkes, D.J., Sridhar, A., Kelly, J., Stoyanov, D.: Image based surgical instrument pose estimation with multi-class labelling and optical flow. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A. (eds.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 331–338. Springer, Cham (2015)
- [6] Liu, X., Plishker, W., Shekhar, R.: Hybrid electromagnetic-ArUco tracking of laparoscopic ultrasound transducer in laparoscopic video. *Journal of Medical Imaging* **8**(01) (2021) <https://doi.org/10.1117/1.jmi.8.1.015001>
- [7] Fusaglia, M., Hess, H., Schwalbe, M., Peterhans, M., Tinguely, P., Weber, S., Lu, H.: A clinically applicable laser-based image-guided system for laparoscopic liver procedures. *International Journal of Computer Assisted Radiology and Surgery* **11**(8), 1499–1513 (2015) <https://doi.org/10.1007/s11548-015-1309-8>
- [8] Rabbani, N., Calvet, L., Espinel, Y., Roy, B.L., Ribeiro, M., Buc, E., Bartoli, A.: A methodology and clinical dataset with ground-truth to evaluate registration accuracy quantitatively in computer-assisted laparoscopic liver resection. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* **10**(4), 441–450 (2022) <https://doi.org/10.1080/21681163.2021.1997642> <https://doi.org/10.1080/21681163.2021.1997642>
- [9] Sarikaya, D., Corso, J.J., Guru, K.A.: Detection and localization of robotic tools in robot-assisted surgery videos using deep neural networks for region proposal and detection. *IEEE Transactions on Medical Imaging* **36**(7), 1542–1549 (2017) <https://doi.org/10.1109/TMI.2017.2665671>



- [10] Laina, I., Rieke, N., Rupperecht, C., Vizcaíno, J.P., Eslami, A., Tombari, F., Navab, N.: Concurrent segmentation and localization for tracking of surgical instruments. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017*, pp. 664–672. Springer, Cham (2017)
- [11] Kurmann, T., Marquez Neila, P., Du, X., Fua, P., Stoyanov, D., Wolf, S., Schnitman, R.: Simultaneous recognition and pose estimation of instruments in minimally invasive surgery. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017*, pp. 505–513. Springer, Cham (2017)
- [12] Ye, M., Zhang, L., Giannarou, S., Yang, G.-Z.: Real-time 3d tracking of articulated tools for robotic surgery. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016*, pp. 386–394. Springer, Cham (2016)
- [13] Allan, M., Ourselin, S., Hawkes, D.J., Kelly, J.D., Stoyanov, D.: 3-d pose estimation of articulated instruments in robotic minimally invasive surgery. *IEEE Transactions on Medical Imaging* **37**(5), 1204–1213 (2018) <https://doi.org/10.1109/TMI.2018.2794439>
- [14] Doignon, C., Mathelin, M.: A degenerate conic-based method for a direct fitting and 3-d pose of cylinders with a single perspective view. *IEEE International Conference on Robotics and Automation*, 4220–4225 (2007)
- [15] Zenteno, O., Treuillet, S., Lucas, Y.: Pose estimation of a markerless fiber bundle for endoscopic optical biopsy. *Journal of Medical Imaging* **8**(02) (2021) <https://doi.org/10.1117/1.jmi.8.2.025001>
- [16] Dzubiella, J., Schmidt, M., Löwen, H.: Topological defects in nematic droplets of hard spherocylinders. *Physical Review E* **62**(4), 5081–5091 (2000) <https://doi.org/10.1103/physreve.62.5081>
- [17] Zenteno, O., Treuillet, S., Lucas, Y.: 3d cylinder pose estimation by maximization of binary masks similarity: A simulation study for multispectral endoscopy image registration. In: *VISIGRAPP* (2019). <https://api.semanticscholar.org/CorpusID:88499715>
- [18] Montaña-Brown, N., Ramalhinho, J., Koo, B., Allam, M., Davidson, B., Gurusamy, K., Hu, Y., Clarkson, M.J.: Towards multi-modal self-supervised video and ultrasound pose estimation for laparoscopic liver surgery. In: Aylward, S., Noble, J.A., Hu, Y., Lee, S.-L., Baum, Z., Min, Z. (eds.) *Simplifying Medical Ultrasound*, pp. 183–192. Springer, Cham (2022)
- [19] Montaña-Brown, N., Ramalhinho, J., Allam, M., Davidson, B., Hu, Y., Clarkson,

- M.J.: Vessel segmentation for automatic registration of untracked laparoscopic ultrasound to CT of the liver. *International Journal of Computer Assisted Radiology and Surgery* (2021) <https://doi.org/10.1007/s11548-021-02400-6>
- [20] Ramalhinho, J., Tregidgo, H.F.J., Gurusamy, K., Hawkes, D.J., Davidson, B., Clarkson, M.J.: Registration of untracked 2d laparoscopic ultrasound to ct images of the liver using multi-labelled content-based image retrieval. *IEEE Transactions on Medical Imaging* **40**(3), 1042–1054 (2021) <https://doi.org/10.1109/TMI.2020.3045348>
  - [21] Ramalhinho, J., Tregidgo, H., Allam, M., Travlou, N., Gurusamy, K., Davidson, B., Hawkes, D., Barratt, D., Clarkson, M.J.: Registration of untracked 2d laparoscopic ultrasound liver images to ct using content-based retrieval and kinematic priors. In: Wang, Q., Gomez, A., Hutter, J., McLeod, K., Zimmer, V., Zettinig, O., Licandro, R., Robinson, E., Christiaens, D., Turk, E.A., Melbourne, A. (eds.) *Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image Analysis*, pp. 11–19. Springer, Cham (2019)
  - [22] Cross, G., Zisserman, A.: Quadric reconstruction from dual-space geometry. In: *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pp. 25–31 (1998). <https://doi.org/10.1109/ICCV.1998.710697>
  - [23] Kallasi, F., Oleari, F., Bottioni, M., Rizzini, D.L., Caselli, S.: Object detection and pose estimation algorithms for underwater manipulation. (2014). <https://api.semanticscholar.org/CorpusID:18816775>
  - [24] Gummeson, A., Engman, J., Åström, K., Oskarsson, M.: Fast and efficient minimal solvers for quadric based camera pose estimation. In: *2022 26th International Conference on Pattern Recognition (ICPR)*, pp. 3973–3979 (2022). <https://doi.org/10.1109/ICPR56361.2022.9956135>
  - [25] Gummeson, A., Oskarsson, M.: Robust and accurate cylinder triangulation. In: Gade, R., Felsberg, M., Kämäräinen, J.-K. (eds.) *Image Analysis*, pp. 451–466. Springer, Cham (2023)
  - [26] Fitzgibbon, A., Pilu, M., Fisher, R.B.: Direct least square fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**(5), 476–480 (1999) <https://doi.org/10.1109/34.765658>
  - [27] Wong, K.-Y.K., Schnieders, D., Li, S.: Recovering light directions and camera poses from a single sphere. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *Computer Vision – ECCV 2008*, pp. 631–642. Springer, Berlin, Heidelberg (2008)
  - [28] Toth, T., Hajder, L.: A Minimal Solution for Image-Based Sphere Estimation. *International Journal of Computer Vision* **131**, 1428–1447 (2023)
  - [29] Taghanaki, S.A., Abhishek, K., Cohen, J.P., Cohen-Adad, J., Hamarneh,



- G.: Deep semantic segmentation of natural and medical images: a review. *Artificial Intelligence Review* **54**(1), 137–178 (2020) <https://doi.org/10.1007/s10462-020-09854-1>
- [30] Iglovikov, V.I., Mushinskiy, S., Osin, V.: Satellite imagery feature detection using deep convolutional neural network: A kaggle competition. *ArXiv abs/1706.06169* (2017)
  - [31] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241. Springer, Cham (2015)
  - [32] Iglovikov, V.I., Shvets, A.A.: Terausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. *ArXiv abs/1801.05746* (2018)
  - [33] Chaurasia, A., Culurciello, E.: Linknet: Exploiting encoder representations for efficient semantic segmentation. *2017 IEEE Visual Communications and Image Processing (VCIP)*, 1–4 (2017)
  - [34] Shvets, A.A., Rakhlin, A., Kalinin, A.A., Iglovikov, V.I.: Automatic instrument segmentation in robot-assisted surgery using deep learning. In: *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 624–628 (2018). <https://doi.org/10.1109/ICMLA.2018.00100>
  - [35] Suzuki, S., Abe, K.: Topological structural analysis of digitized binary images by border following. *Comput. Vis. Graph. Image Process.* **30**, 32–46 (1985)