

STITCH 2.0: Extending Augmented Suturing with EKF Needle Estimation and Thread Management

Kush Hari¹, Ziyang Chen¹, Hansoul Kim¹, and Ken Goldberg¹

<https://stitch-2.github.io>

Abstract—Surgical suturing is a high-precision task that impacts patient healing and scarring. Suturing skill varies widely between surgeons, highlighting the need for robot assistance. Previous robot suturing works, such as STITCH 1.0 [1], struggle to fully close wounds due to inaccurate needle tracking and poor thread management. To address these challenges, we present STITCH 2.0, an elevated augmented dexterity pipeline with seven improvements including: improved EKF needle pose estimation, new thread untangling methods, and an automated 3D suture alignment algorithm. Experimental results over 15 trials find that STITCH 2.0 on average achieves 74.4% wound closure with 4.87 sutures per trial, representing 66% more sutures in 38% less time compared to the previous baseline. When two human interventions are allowed, STITCH 2.0 averages six sutures with 100% wound closure rate.

Index Terms—Medical Robots and Systems, Surgical Robotics: Laparoscopy, Computer Vision for Medical Robotics

I. INTRODUCTION

SURGICAL robots have revolutionized minimally invasive surgery, with Intuitive Surgical’s da Vinci system performing over 2.6 million procedures in 2024 [2]. While these procedures require complete human control, recent advances in artificial intelligence (AI) present opportunities for surgical robot autonomy. However, the high-risk nature of surgery raises safety concerns for fully autonomous AI systems.

An alternative approach lies in augmenting surgeon capabilities by automating specific, well-defined subtasks. “Augmented Dexterity” proposed by Goldberg and Guthart in 2024 [3], describes systems where surgical subtasks are performed by robots under close supervision of a human surgeon who is ready to take over at a moment’s notice. Augmented Dexterity has potential to increase surgical consistency and reduce tedium amid a global surgeon shortage [4].

One such repetitive subtask is suturing, where a surgeon uses a needle and thread to close a wound. Despite being common at the end of almost every surgical procedure, suturing quality varies widely between surgeons, especially when

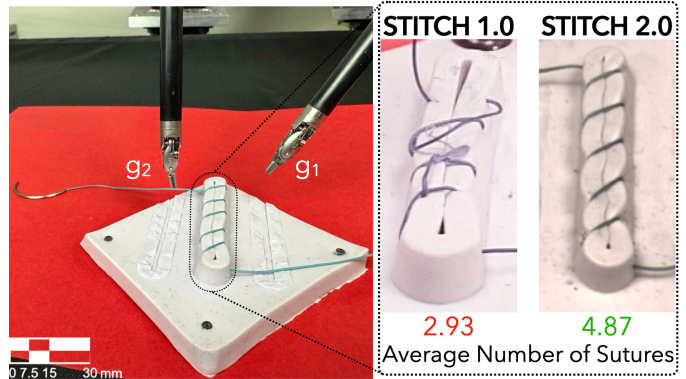


Fig. 1. **Augmented Suturing:** We use the da Vinci Research Kit, equipped with two grippers g_1 and g_2 . STITCH 1.0 averages 2.93 sutures that were untightened and did not fully close the wound. STITCH 2.0 improved to 4.87 sutures, delivering tight, even sutures that fully close the wound.

fatigue increases errors [5]. Uniform, consistent suturing can reduce infection and improve tissue healing [3].

While existing work has made great progress on components such as needle handover [6], thread manipulation [7], and suture placement [8] [9], these systems do not perform consecutive sutures. Previous multi-suture systems rely on either industrial robots using 3D tracking markers [10] or custom gripper attachments [11] not found in hospitals.

In contrast to a previous paper, STITCH 1.0 (Suture Throws Including Thread Coordination and Handoffs) [1] uses the dVRK [12] in an augmented dexterity suturing pipeline including motion primitives for suture throws and needle handover based on a 6D needle pose estimation pipeline. However, STITCH 1.0 consistently faced challenges with needle dropping, thread tangling, and cross-stitching. These failures stemmed from noisy needle pose estimates, poor thread management, and uneven suture placement—all critical issues that must be addressed for clinical applications.

To address these limitations, STITCH 2.0 introduces the following 7 components:

- 1) Extended Kalman Filter integrated with needle pose pipeline to improve estimation accuracy.
- 2) Adaptive depth threshold to filter pointcloud noise around the tip point for accurate needle positioning before insertion.
- 3) 3D suture alignment to determine evenly-spaced insertion and extraction points, eliminating the need for manual human input required by the STITCH 1.0 pipeline.
- 4) Thread management methods that prevent thread tangling and cross-stitching to improve suture quality.
- 5) Needle-to-wound alignment system to improve suture placement accuracy.

Manuscript received: July, 10, 2025; Accepted October, 3, 2025.

This paper was recommended for publication by Editor Jessica Burgner-Kahrs upon evaluation of the Associate Editor and Reviewers’ comments. Kush Hari was supported by the NSF GRFP. The da Vinci Research Kit is supported by the National Science Foundation, via the National Robotics Initiative (NRI), as part of the collaborative research project “Software Framework for Research in SemiAutonomous Teleoperation” between The Johns Hopkins University (IIS1637789), Worcester Polytechnic Institute (IIS1637759), and the University of Washington (IIS 1637444). (Corresponding Author: Kush Hari)

¹Authors are with the Department of Electrical Engineering and Computer Science, UC Berkeley, California, USA {kush_hari, ziyang.chen, robotgksthf, goldberg}@berkeley.edu

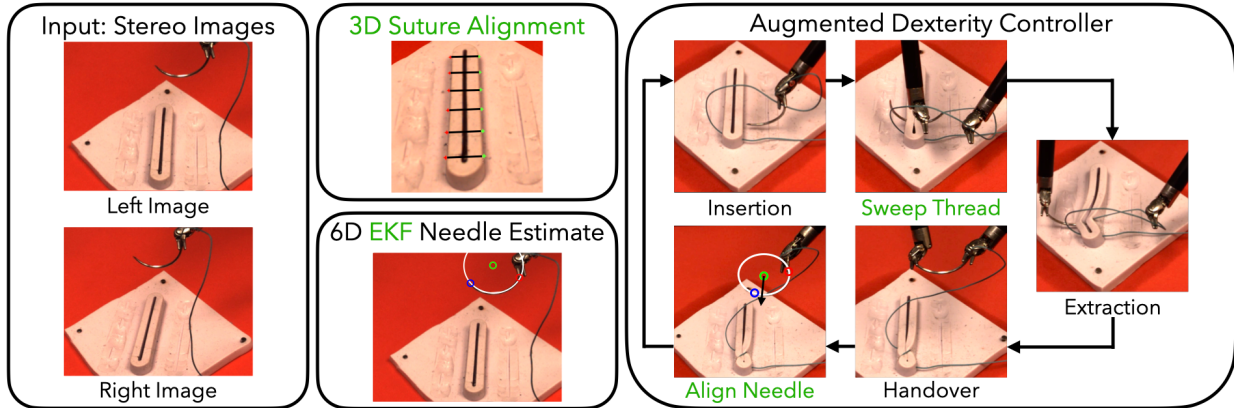


Fig. 2. Overview of the STITCH 2.0 pipeline with extensions from STITCH 1.0 [1] in green. STITCH 2.0 integrates a 3D Suture Alignment, 6D EKF Needle Pose Estimator, and Augmented Dexterity controller. The 3D suture alignment determines optimal insertion points, the 6D EKF Pose Estimator provides geometric data for precise needle handling, and the Augmented Dexterity controller executes suturing.

- 6) Flood-filled stereo images based on U-Net masks to reduce needle pointcloud noise from light specularities.
- 7) Skeletonized U-Net mask applied to depth image to eliminate needle boundary pointcloud noise.

These improvements allow STITCH 2.0 to perform 66% more sutures in 38% less time compared to STITCH 1.0.

II. RELATED WORK

A. Needle Pose Estimation

Accurate needle pose estimation is crucial for autonomous suturing to ensure precise suture placement and prevent tissue damage. Previous approaches include segmentation-based methods [1] [6], which can struggle with noisy 3D reconstruction, along with end-to-end learning methods [13], which require detailed CAD models and separate training for different needle types. Alternatively, keypoint estimation methods, both in simulation [14] and in reality [15] [16], face challenges during needle manipulation. Although recent work [17] achieves submillimeter accuracy using gripper motion priors, tracking at high accuracy without such priors remains essential. Li's concurrent work [18] on monocular keypoint estimation struggles with common suturing occlusions. Needle tracking in STITCH 2.0 improves on [1] by using flood-filled stereo imaging to reduce the 3D reconstruction error common to segmentation-based methods.

B. Suturing Subtask Automation

The multistep nature of suturing has led researchers to focus on automating specific subtasks. These include needle handover [6], suture placement [8] [9], thread manipulation [7], and needle pickup [15], all with success rates exceeding 90%. However, chaining these subtasks together into a full autonomous pipeline will accumulate error and lower the success rate for full wound suturing.

C. Suturing Automation Systems

While some studies have explored suture trajectory planning [19], they primarily focus on simulation or avoid continuous

suturing. Previous work on continuous suturing uses threadless needles or specialized equipment such as painted needles and gripper attachments [11] [20]. The Smart Tissue Autonomous Robot (STAR) system [10] demonstrates autonomous suturing in biological tissue but uses industrial robots and 3D tracking markers not available in surgical settings.

D. End-to-End Transformer Model for Surgical Tasks

In recent work using a large transformer model, the Surgical Robot Transformer (SRT) [21] employs imitation learning from demonstrations to perform isolated subtasks like needle handover and knot tying. SRT performs a single subtask while STITCH 2.0 performs 6 continuous suturing cycles for wound closure. Additionally, SRT uses four cameras (one stereo pair plus wrist-mounted cameras on each gripper), while STITCH 2.0 only uses one stereo pair.

III. PROBLEM STATEMENT

We consider a laparoscopic surgery setting with a raised wound phantom and bimanual classic dVRK Intuitive Surgical robot (Fig. 1). Given a stereo camera video stream, the objective is to evenly suture the wound closed by planning suitable needle entry and exit points and manipulating a surgical needle with thread.

We make the following assumptions:

- 1) The needle is a semi-circle with known radius, common for clinical suturing.
- 2) There is sufficient thread attached to the needle to fully suture the wound.
- 3) The thread starts untangled at the beginning of each experiment.
- 4) We use a raised linear wound phantom, common for resident training [24]. The wound is placed orthogonal to the stereo camera pair for easier suture alignment. See Limitations and Future Work section for more challenging wound configurations.
- 5) The stereo camera pair is calibrated and its intrinsic and extrinsic parameters are known.
- 6) The dVRK is calibrated with an RNN based on [25] to reduce kinematic errors.

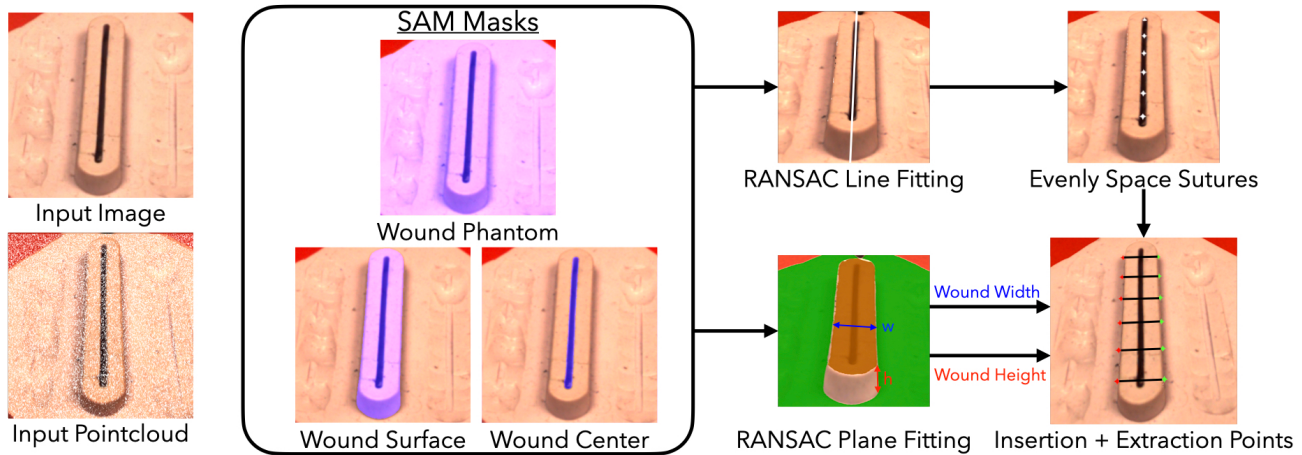


Fig. 3. The **3D Suture Alignment** starts with generating a scene pointcloud using RAFT-Stereo [22], followed by segmentation of the wound center, surface, and phantom with Segment Anything (SAM) [23]. The wound height is calculated by measuring the distance between the wound and phantom surfaces through RANSAC fitting. A 3D line representing the wound center is obtained by projecting points onto the top wound surface plane and fitting with RANSAC. Suture positions are evenly distributed along the centerline, and the wound width and height are used to determine the insertion points for needle placement.

We evaluate STITCH 2.0 based on 3 evaluation metrics: completion time in seconds, number of successful sutures, and percentage of wound gap closed. A successful suture throw is defined by passing the needle through the wound at the desired insertion/extraction points and performing sufficient suture cinching without thread over-tightening, thread tangling, or cross-stitching. Percentage of wound gap closed is defined as the number of stitches that closed the wound divided by the total number of planned stitches.

IV. METHODS

STITCH 2.0 iterates over 3 phases as specified in Fig. 2.

- A. **3D Suture Alignment** autonomously determines the optimal insertion/extraction points for each suture based on stereo imaging, improving on previous work [1] where the insertion/extraction points were manually provided.
- B. **6D Needle Pose Estimator with EKF**.
- C. **Augmented Dexterity Controller** manages gripper movements to complete suture throws, needle extraction, needle handover, and thread slack management based on perception inputs from Phases A and B.

A. 3D Suture Alignment

Uniform suture alignment can facilitate patient healing and reduce tissue scarring. It is important to properly space sutures because insufficient spacing can restrict tissue oxygenation, while excessive spacing can fail to provide adequate closure force [26]. While STITCH 1.0 required user-specified insertion and extraction points, STITCH 2.0 performs autonomous suture alignment (Fig. 3).

The system starts by building a 3D wound model. It creates a scene pointcloud by passing stereo images into RAFT-Stereo [22], an off-the-shelf stereo network. It then employs Segment Anything (SAM) [23] to segment the wound center, surface, and phantom, creating separate pointclouds for each component. Because we assume the wound is flat and raised, the system calculates the height by measuring the distance

between the top wound surface plane and the bottom phantom surface plane. Both planes are determined through Random Sample Consensus (RANSAC) fitting of their respective pointclouds.

To determine the suture positions, STITCH 2.0 projects the wound center pointcloud to the top wound surface plane and fits a center line using RANSAC. Based on the force closure model from [9], the robot attempts to evenly distribute 6 suture positions along this centerline.

To find the 3D needle insertion-extraction points, the system first finds the wound width vector. The direction is computed as the cross product of the wound surface normal and centerline. The magnitude is calculated by projecting the wound surface pointcloud onto this direction unit vector and finding the maximum distance between projected points. With the wound width and wound height, the system calculates the 3D insertion-extraction points by offsetting the centered suture positions by half the wound width and half the wound height in their respective directions, so the needle insertion-extraction points are at mid-phantom height on the wound boundary.

B. 6D Needle Pose Estimator with EKF

The STITCH 2.0 6D pose estimator with EKF determines the best-fit 3D circle for the surgical needle, providing the circle center coordinates, needle endpoint positions, and the normal vector of the needle plane as outlined in Fig. 4. Previous realtime needle pose estimation methods [16] [17] [18] use keypoints and struggle in common suturing scenarios where many keypoints on the needle can be occluded by the grippers or the wound tissue.

We begin by training a U-Net [27] for needle segmentation. The U-Net is trained on 1200 stereo image pairs showing the needle in random poses and lighting conditions. The number of training images was iteratively increased until the U-Net achieved robust performance across image conditions typical in suturing, including scenarios with both needle and its shadow present. Training data is collected by painting the needle with ultraviolet (UV) paint and performing color

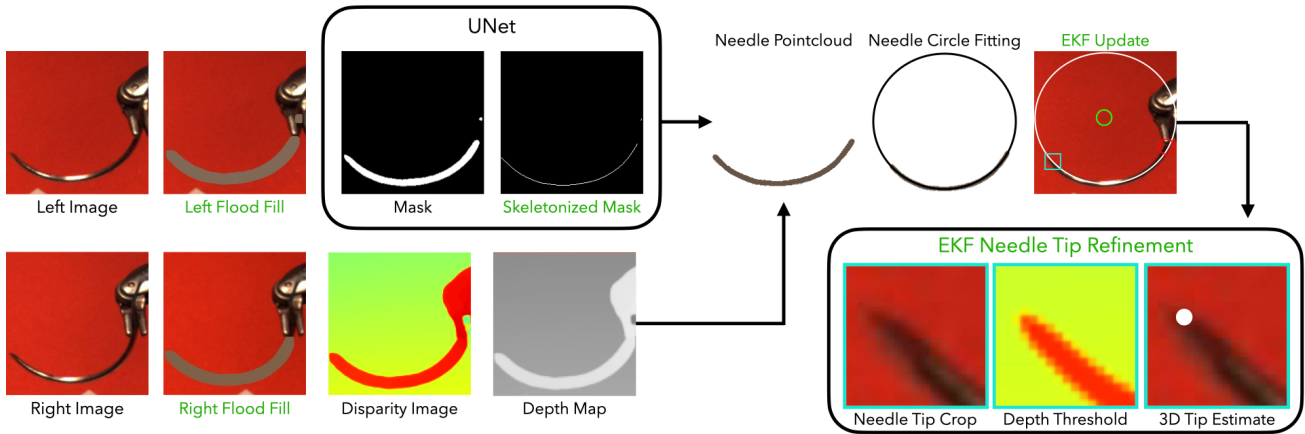


Fig. 4. The **6D Needle Pose Estimator** begins by training a U-Net for needle segmentation. A flood-fill approach is applied to reduce noise and eliminate specularities before processing through RAFT-Stereo [22] to generate a disparity image. The disparity image is converted into a depth map, and a skeletonized mask is used to improve pose estimates and accelerate circle fitting. The pose is further stabilized with an Extended Kalman Filter (EKF). Needle tip estimation is refined by adaptive thresholding and computing the intersection of the pixel ray with the EKF 3D circle estimate, enabling precise needle manipulation for insertion and suture alignment. **New contributions with STITCH 2.0 from STITCH 1.0 [1] are in green.**

segmentation on UV images to generate ground truth needle masks. [28]. However, UV paint degradation can result in incomplete masks, so we use SAM [23] to generate complete ground truth masks for the affected images.

During suturing, the system uses the U-Net to segment the needle in both the left and right images and flood-fill the needle regions with the average color. This is significant because in STITCH 1.0, the stereo pointcloud was noisier in areas where the needle had white spots caused by the needle’s reflective surface. Replacing the entire needle region with a uniform average color in the left and right images eliminates these visual specularities. Therefore, the system passes the flood-filled left and right images into RAFT-Stereo [22] to generate a disparity image and depth map.

From there, the system applies a skeletonized U-Net mask to the depth map and creates the needle pointcloud. In STITCH 1.0 [1], we applied the unmodified U-Net mask to the depth map, and observed the needle pointcloud was noisier at the boundaries. Here, the system uses Zhang’s method [29] to iteratively remove border pixels and create a single pixel-width skeletonized mask. Because we use the skeletonized mask to obtain the needle pointcloud, STITCH 2.0 yields robust needle pose estimates while simultaneously accelerating the subsequent circle fitting by 2-3x by processing just a fraction of the original pointcloud.

After these filtering steps, the pointcloud is stable enough to perform the 3D fitting process from STITCH 1.0 [1]. This involves RANSAC plane fitting to remove lingering pointcloud outliers and estimate the needle normal vector. The algorithm iteratively samples random subsets of the needle pointcloud, fits a plane to each subset, and selects the subset with the most inliers for the final plane fit. The RANSAC inliers are then projected onto the best-fit plane to optimize circle parameters, with the two farthest points selected as needle endpoints.

EKF Needle Estimate: While circle fitting provides reliable estimates for the needle normal vector and center point, the endpoints have up to 3-5 mm error due to noisy pointcloud data and U-Net segmentation errors at the endpoints. This

inaccuracy results in poor needle insertions and sutures.

To address this in STITCH 2.0, the system generates a new Extended Kalman Filter (EKF) before each needle interaction step (needle extraction, handover, alignment, etc.) when the needle is stationary. We do this as opposed to having a single EKF track the needle throughout the procedure because the needle estimates are noisiest during needle motion and STITCH 2.0 only requires needle pose estimation directly before needle interaction steps, as described in the Augmented Dexterity Controller section (Sec. C). We use EKFs as opposed to standard Kalman Filters because EKFs can handle non-Gaussian noise by performing linearization around the current needle state mean and covariance.

The EKF needle state is a 13-dimensional vector comprising the needle center point (c_x, c_y, c_z) , left endpoint (l_x, l_y, l_z) , right endpoint (r_x, r_y, r_z) , normal vector (n_x, n_y, n_z) , and radius r :

$$\mathbf{x} = [c_x, c_y, c_z, l_x, l_y, l_z, r_x, r_y, r_z, n_x, n_y, n_z, r]^T$$

Since the measurement vector has identical structure to the state vector, the measurement model uses the identity matrix: $\mathbf{H} = \mathbf{I}_{13}$. Because the needle remains stationary between EKF steps, the state transition model is also the identity matrix: $\mathbf{F} = \mathbf{I}_{13}$. These state and measurement models can now be directly applied to the EKF.

For each EKF, the initial state and covariance is determined by averaging the first seven measurements. Then, the system applies the EKF update step to the next three measurements that are within 3 standard deviations of the initial state to get the final needle estimate. The number of initial and update measurements are empirically determined by iterating from zero until the estimates consistently converge.

With this stable needle pose estimate from the EKF, the robot can now accurately determine the needle tip point. We first crop a 200x200 patch centered around the EKF-estimated needle tip in the disparity image (Fig. 4). Then, we perform contour detection [29] with adaptive depth thresholding to create a skeletonized needle mask in the image crop. Since this mask maintains a consistent 1-pixel width, the end corner

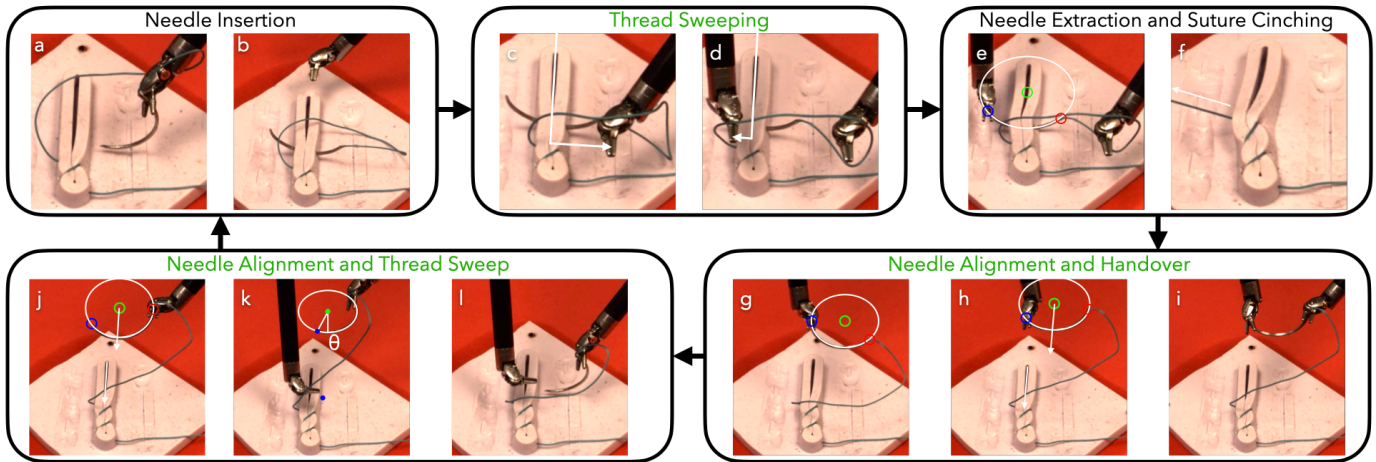


Fig. 5. The **Augmented Dexterity Controller** enables suturing automation through five components: (a), (b) needle insertion, (c), (d) thread sweeping, (e), (f) needle extraction with thread cinching, (g)-(i) needle alignment and handover, and (j)-(l) pre-insertion alignment. Needle insertion involves precise suture positioning. Thread sweeping prevents tangling, while extraction involves cinching for wound closure. Needle alignment and handover correct needle orientation, and pre-insertion alignment facilitates a high-quality insertion. **New STITCH 2.0 contributions from STITCH 1.0 [1] are in green.**

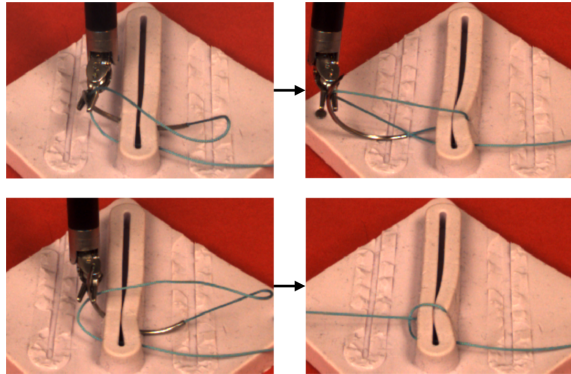


Fig. 6. **Previous Thread Tangling Failures.** (top) Gripper grasps thread and needle together leading to thread tangling. (bottom) Gripper grasps needle while thread is behind the needle leading to cross-stitch.

of the mask corresponds to the needle tip in pixel space. While the disparity image signal is strong enough for an adaptive threshold, directly projecting that pixel to a 3D point is too noisy. Since the EKF merges prior measurements resulting in a stable needle circle estimate, the system determines the depth by deprojecting the needle tip pixel to a 3D ray and computing the intersection between the ray and the EKF 3D circle estimate. This provides a reliable 3D estimate of the needle tip, enabling the augmented dexterity controller to maintain proper needle alignment with the wound during suturing, as described in the next section.

C. Augmented Dexterity Controller

The augmented dexterity controller implements suturing automation through the two dVRK grippers (g_1 and g_2) with five key subtasks (Fig. 5): needle insertion, thread sweeping, needle extraction with thread cinching, needle alignment and handover, and pre-insertion needle alignment. Because these subtasks are negatively affected by the inaccurate cable-driven kinematics, we perform an RNN calibration [25] for each dVRK gripper (Fig. 1).

1) *Surgical Needle Insertion:* The insertion process begins with positioning the needle tip at the planned insertion point (Fig. 5(a)). Gripper g_1 executes a 3-step twist motion that

reduces strain on the wound phantom as the needle is inserted. This motion was empirically determined so that g_2 can grasp the needle for extraction. Initially, the needle is translated into the wound by one wound width (9 mm), ensuring the needle tip completely penetrates through the wound. The needle is then rotated 40 degrees around the wound center line, leveraging the circular needle geometry. Finally, the needle undergoes an additional 2 mm translation to prevent the needle from slipping out the wound. Afterward, g_1 releases the needle when insertion is completed (Fig. 5(b)).

2) *Thread Slack Management:* Poor thread management can cause thread tangling (Fig. 6 top), or “cross-stitching” that prevents proper suture tightening and wound closure (Fig. 6 bottom). Both of these issues occur when the thread rests behind the needle during extraction as seen in Fig. 6. In STITCH 1.0 [1], thread management involves just moving g_2 down the center of the wound to push the thread forward. However, the thread often springs back behind the needle leading to thread tangling errors. For STITCH 2.0, we empirically determine a trajectory where g_1 and g_2 coordinate to ensure the thread remains in front of the needle: g_1 moves down the wound center, lifts the thread, and moves to the right (Fig. 5(c)). This provides ample clearance for g_2 to move down the wound center and lift the thread on the other side to ensure the thread stays in front of the needle (Fig. 5(d)). Then, g_2 returns to the home position, but g_1 maintains its position to hold the thread during needle extraction. This move prevents thread tangling across various configurations (Fig. 6). When the thread is already positioned in front of the needle, the move does not disrupt this configuration, allowing the thread to remain untangled throughout the extraction process.

3) *Surgical Needle Extraction with Thread Cinching:* Extraction begins with g_2 grasping the needle 2 mm from its tip (based on the 6D pose estimate) to avoid dulling the needle tip (Fig. 5(e)). Once grasped, the needle is twisted to extract it out mirroring the earlier insertion twist motion.

Subsequently, the surgical thread is pulled to close the wound and then the needle is moved to the handover position (Fig. 5(f)). The following equation is used to determine the

TABLE I
STITCH 2.0 WITH ABLATIONS FOR 15 TRIALS OF 6 SUTURES FOR EACH METHOD (MAXIMUM 450 SUTURE TRIALS).

	Avg. Number of Sutures	Single-Suture Success Rate (%)	Wound Gap Closure Rate (%)	Mean Time per Suture (s)	Error types					Needle Estimate Success Rate (%)
					A	T	I	M	Total	
STITCH 1.0 [1]	2.93 ± 0.70	69.4	22.3	159.3	0	2	8	5	15	42
STITCH 1.0 [1] + 2 Human Corr.	4.47 ± 0.99	83.3	35.3	141.9	0	0	8	5	13	42
STITCH 2.0 w/o Thread Mgmt.	2.87 ± 0.83	74.1	39.4	81.02	2	7	2	4	15	-
STITCH 2.0 w/o EKF	3.20 ± 0.68	76.2	44.5	61.60	2	4	3	6	15	85.98
STITCH 2.0	4.87 ± 0.83	86.9	74.4	98.74	3	3	5	0	11	97.56
STITCH 2.0 + 1 Human Corr.	5.53 ± 0.52	92.2	87.8	103.40	2	2	1	0	5	97.56
STITCH 2.0 + 2 Human Corr.	6.00 ± 0.00	100.0	100.0	101.60	0	0	0	0	0	97.56

STITCH 2.0 Comparisons (N = 15): We evaluate each method on 15 trials of 6 sutures each. STITCH 2.0 executes more sutures than both ablations, STITCH 1.0, and STITCH 1.0 with human corrections. Furthermore, allowing for up to 2 human corrections led to 100% wound closure rate with 0 errors as seen in the bottom row of the table. See the Suturing Experiments section for more details about the different error types.

Number of Sutures Achieved

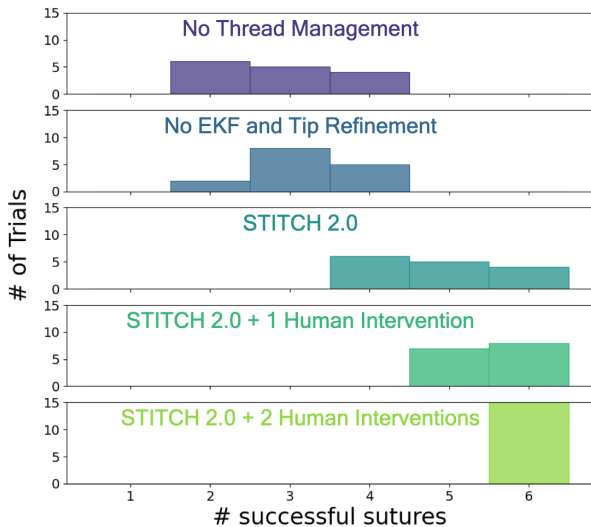


Fig. 7. **Histogram of the Number of Sutures Achieved by Method.** The results in the histogram shown above is for 15 trials of 6 sutures per each of the 5 methods in the ablation up to 450 suture trials.

appropriate goal point to pull the thread for cinching:

$$D_i = [n * d - ((i - 1) * d)] \quad (1)$$

Where D_i is the thread cinching translation, n is the number of planned sutures, i is the current suture number, and d is the per-suture thread length determined by the 3D suture alignment.

4) *Needle Alignment and Handover:* A common failure case for needle handover in STITCH 1.0 [1] is that the needle snapped to an orientation where it was misaligned with the wound when g_2 released the needle. To mitigate this, STITCH 2.0 rotates the needle such that the needle normal vector is parallel to the wound allowing for orthogonal tissue entry (Fig. 5(g)). Then, STITCH 2.0 rotates the needle around the normal until both endpoints are level (Fig. 5 (h)). Leveling the needle endpoints is crucial for preventing needle deflection and misalignment after gripper release by ensuring g_1 has the same approach angle as g_2 . With the needle properly oriented, gripper g_1 grasps 2 mm below the estimated needle endpoint connected to the thread (Fig. 5(i)). Once g_1 and g_2 have grasped the needle, g_1 releases the needle and it remains at the same orientation.

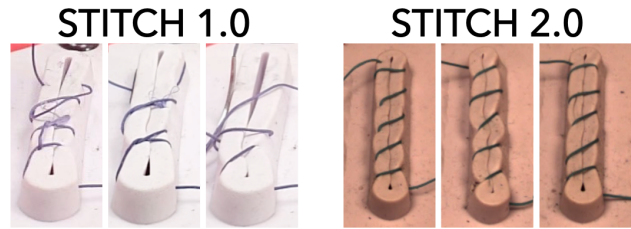


Fig. 8. **Qualitative Comparison.** While STITCH 1.0 (left) produces messy sutures that are insufficient for wound closure, STITCH 2.0 (right) delivers uniform sutures that can achieve successful closure.

5) *Pre-Insertion Needle Alignment:* To ensure the needle enters the tissue orthogonally for a high quality stitch and minimal tissue damage, the system first rotates the needle so its normal is aligned to the wound similar to the previous step. (Fig. 5(j)). The needle is then rotated around its normal vector to establish a 20-degree angle between the vertical axis and the needle tip-to-center line (Fig. 5(k)). This empirically-determined angle improves insertion quality by leveraging the flat, linear tip design, which translates smoothly into tissue when positioned at this orientation. With the needle in this configuration, g_2 then pushes the thread and holds it forward while g_1 moves to the next insertion point (Fig. 5(l)). Overall, this process iterates until all planned sutures are completed.

V. PHYSICAL EXPERIMENTS

Experiments evaluate (1) the ability of STITCH 2.0 to suture a wound closed compared to STITCH 1.0, (2) the impact of the EKF and thread management on STITCH 2.0, (3) the impact of augmented dexterity on STITCH 2.0 by allowing for minimal human corrections, and (4) the needle pose estimation accuracy across all suturing trials.

For robot execution, we use a bimanual classic dVRK [12]. The wound phantom is a soft tissue suture pad [24] used for medical education and training. We use a pair of fixed RGB stereo cameras (Prosilica GC 1290, Allied Vision). The stereo cameras record 1280x960 images at 33 frames per second. We use variable-length (30-75 cm) green braided Ethibond Excel™ suture thread. Attached to the thread is a V-37 40mm half-circular tapercut surgical needle [30].

A. Suturing Experiments

We evaluated STITCH 2.0 across 15 experiments of six consecutive suture trials each. A suture is successful if the

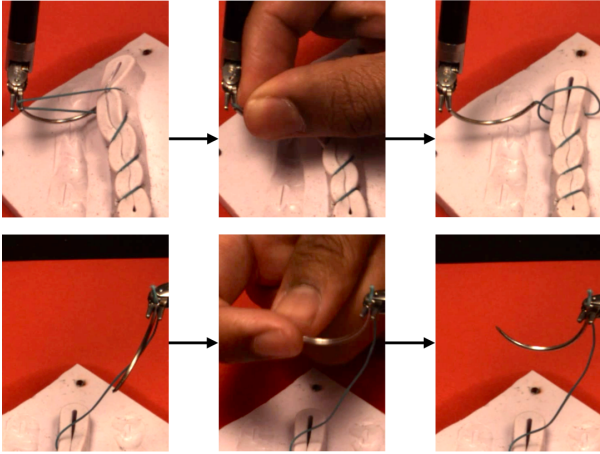


Fig. 9. **Sample Human Interventions for STITCH 2.0.** (top) Human removes thread from gripper to prevent tangling during extraction. (bottom) Human corrects needle orientation after handover.

needle is inserted and extracted from the phantom without damaging the wound and the suture thread is correctly tightened across the wound for closure. We measure mean sutures achieved, single-suture throw success rate, wound gap closure rate, mean time per suture, and needle estimate success rate. During experiments, we observed 4 different types of failures: (A)lignment error where the needle alignment step failed, leaving the needle unaligned to the wound causing a failed insertion, extraction, or handover, (T)hread management error where the thread gets tangled with the robot or is improperly cinched over the wound leading to a low quality suture, (I)nsertion error where the needle misses the phantom or is inserted in the phantom such that it cannot be extracted, and (M)issed grasp where an inaccurate needle pose estimate leads to the gripper missing a grasp during either extraction or handover.

Table I and Fig. 7 display experimental results. STITCH 2.0 achieves an 86.9% single-suture success rate and 74.4% wound gap closure rate. Additionally, STITCH 2.0 achieves an average of 4.87 sutures at 98.74 seconds per suture—a 66% increase in higher-quality sutures and 38% decrease in time compared to STITCH 1.0 (Fig. 8). Furthermore, STITCH 2.0 averages 6 sutures with 2 human interventions (See Augmented Dexterity in Experiments for more details).

B. Ablations

We compare STITCH 2.0 against two ablation configurations: one without thread management and one without EKF and needle tip refinement (Table I).

STITCH 2.0 completes 2.00 additional sutures on average compared to the ablation without thread management resulting in a 35% increase in wound gap closure. This improvement comes from the thread management method that reduces thread tangling by 57% despite taking 17.72 more seconds.

Similarly, STITCH 2.0 averages 1.67 additional sutures compared to the ablation without the EKF and needle tip refinement. While this additional EKF and needle tip refinement step averages an extra 37.14 seconds, it reduces alignment and missed grasp failures by 50%.

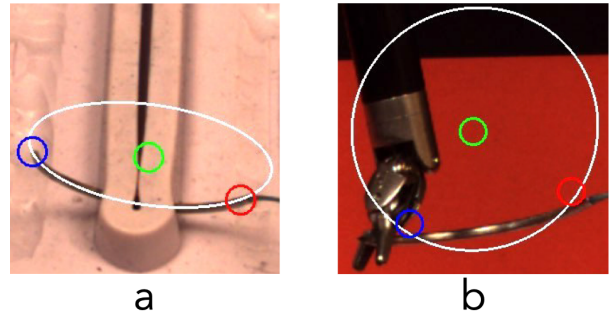


Fig. 10. **Needle Pose Estimate Success and Failure.** a. Accurate needle pose estimate: the projected blue and red circles contain the needle endpoints and the projected white circle follows the needle curvature. b. Inaccurate needle pose estimate.

Because STITCH 2.0 has a higher average suture count compared to its ablations, it has increased error counts in certain individual categories. However, STITCH 2.0 still achieves fewer total errors across all categories combined, as demonstrated in the Total Error column of Table I.

C. Augmented Dexterity in Experiments

In the context of Augmented Dexterity, we allow a human supervisor to intervene when a trial appears nonrecoverable and cannot proceed to the next stage [3]. When this happens, we note the failed suture and allow up to 2 human corrections as seen in the last 2 rows of Fig. 7. Example human interventions include needle realignment, thread untangling, thread removal from gripper, or needle replacement after it is dropped.

With respect to human interventions, we found that STITCH 2.0 with 1 intervention averaged 5.55 sutures with a 87.8% wound gap closure (Table I). STITCH 2.0 with 2 interventions consistently performed all 6 sutures with a 100% wound gap closure (Table I).

D. Needle Pose Estimation Ablation

To evaluate needle pose estimation accuracy, we overlaid the needle pose estimate on the image and compared it to the needle position in the image, with measurements taken prior to each needle grasping or manipulation action during suturing trials. Success and failure examples are shown in Fig 10. The complete STITCH 2.0 pipeline achieved 97.56% accuracy, outperforming the 85.98% accuracy of the system without EKF and the 42% accuracy of STITCH 1.0 (Table I).

The STITCH 2.0 needle pose estimation pipeline performs worse when the needle is oriented away from the camera, where it no longer appears circular in the image, resulting in needle endpoint errors up to 5 mm as shown in Fig 10b.

VI. LIMITATIONS AND FUTURE WORK

STITCH 2.0 has three primary limitations. First, the most common failure occurs during needle insertion. Despite accurate needle pose estimation and deep calibration of the robot [25], sub-millimeter gripper inaccuracies sometimes result in the needle tip being at the wrong height for stable insertion. In future work, we will improve upon the RNN gripper calibration [25] to reduce the dVRK kinematic inaccuracies.

Second, while STITCH 2.0 suture completion time is faster than previous work, it still takes 98.74 seconds per suture when a surgeon could perform a suture in 5-10 seconds. This is primarily due to needle estimation operating at 1 Hz, limited by the computational demands of stereo reconstruction and pointcloud processing to get stable estimates. Because needle pose estimation relies on 10-15 estimates depending on the state confidence from the Extended Kalman Filter (EKF), needle pose estimation accounts for 63.76% of total suture time. Future work will explore faster alternatives such as cropping stereo images on the needle to reduce computational load for faster stereo reconstruction and integrating the STITCH 2.0 needle pose estimator with concurrent, realtime monocular keypoint methods [14] [16] [18].

Third, while STITCH 2.0 addresses challenges in needle pose estimation, thread management, and suture placement relevant for any environment, several important steps are required before STITCH 2.0 can generalize to ex-vivo and clinical settings. Currently, STITCH 2.0 uses an external stereo camera pair rather than an endoscope commonly used in laparoscopic settings. Contemporary stereo endoscopes have a narrow baseline, resulting in noisier pointclouds. Future work will extend STITCH 2.0's noise reduction methods to better handle pointcloud noise for stereo endoscopes. Additionally, we assume the wound is raised and orthogonal to the camera, which simplifies needle insertion since the needle does not require full rotation through tissue. Additionally, the orthogonal positioning provides clear visibility of insertion and extraction points for suture alignment. Future work will extend the STITCH 2.0 suture alignment and needle placement components to handle more complex wound geometries, including nonlinear and 3D wounds.

ACKNOWLEDGMENT

We thank Madison Veliky, Chung Min Kim, and Justin Kerr for their feedback on paper drafts and insightful conversations about the project. We thank Intuitive Surgical for the dVRK.

REFERENCES

- [1] K. Hari, H. Kim, W. Panitch, K. Srinivas, V. Schorp, K. Dharmarajan, S. Ganti, T. Sadjadpour, and K. Goldberg, "STITCH [1.0]: Augmented dexterity for suture throws including thread coordination and handoffs," in *2024 International Symposium on Medical Robotics (ISMR)*. IEEE, 2024.
- [2] "Intuitive announces preliminary fourth quarter and full year 2024 results," *Intuitive Surgical*, 2025.
- [3] K. Goldberg and G. Guthart, "Augmented dexterity: How robots can enhance human surgical skills," *Science Robotics*, vol. 9, no. 95, 2024.
- [4] E. Lawson, "The global primary care crisis," *British Journal of General Practice*, vol. 73, no. 726, pp. 3–3, 2023.
- [5] M. Byrne and A. Aly, "The surgical suture," *Aesthetic surgery journal*, vol. 39, no. Supplement_2, pp. S67–S72, 2019.
- [6] A. Wilcox, J. Kerr, B. Thananjeyan, J. Ichnowski, M. Hwang, S. Paradis, D. Fer, and K. Goldberg, "Learning to localize, grasp, and hand over unmodified surgical needles," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 9637–9643.
- [7] N. Joglekar, F. Liu, F. Richter, and M. C. Yip, "Autonomous image-to-grasp robotic suturing using reliability-driven suture thread reconstruction," *IEEE Robotics and Automation Letters*, 2025.
- [8] R. Bendikas, V. Modugno, D. Kanoulas, F. Vasconcelos, and D. Stoyanov, "Learning needle pick-and-place without expert demonstrations," *IEEE Robotics and Automation Letters*, vol. 8, no. 6, 2023.
- [9] V. Ramakrishnan, J. Isaac, H. Jalan, H. Polavaram, A. Adler, H. Kim, D. M. Fer, and K. Goldberg, "Automating 3d surgical suture planning," in *2024 IEEE 20th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2024, pp. 2111–2116.
- [10] H. Saeidi, J. D. Opfermann, M. Kam, S. Wei, S. Léonard, M. H. Hsieh, J. U. Kang, and A. Krieger, "Autonomous robotic laparoscopic surgery for intestinal anastomosis," *Science robotics*, vol. 7, no. 62, 2022.
- [11] S. Sen, A. Garg, D. V. Gealy, S. McKinley, Y. Jen, and K. Goldberg, "Automating multi-throw multilateral surgical suturing with a mechanical needle guide and sequential convex optimization," in *2016 IEEE international conference on robotics and automation (ICRA)*, 2016.
- [12] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. DiMaio, "An open-source research kit for the da vinci® surgical system," in *2014 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 6434–6439.
- [13] Y. Xu, K.-Y. Lin, G. Zhang, X. Wang, and H. Li, "Rnnpose: 6-dof object pose estimation via recurrent correspondence field estimation and pose optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
- [14] Y. Jiang, H. Zhou, and G. S. Fischer, "Markerless suture needle tracking from a robotic endoscope based on deep learning," in *2023 International Symposium on Medical Robotics (ISMR)*. IEEE, 2023, pp. 1–7.
- [15] C. D'Ettorre, G. Dwyer, X. Du, F. Chadebecq, F. Vasconcelos, E. De Momi, and D. Stoyanov, "Automated pick-up of suturing needles for robotic surgical assistance," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1370–1377.
- [16] Z.-Y. Chiu, A. Z. Liao, F. Richter, B. Johnson, and M. C. Yip, "Markerless suture needle 6d pose tracking with robust uncertainty estimation for autonomous minimally invasive robotic surgery," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 5286–5292.
- [17] Z.-Y. Chiu, F. Richter, and M. C. Yip, "Real-time constrained 6d object-pose tracking of an in-hand suture needle for minimally invasive robotic surgery," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 4761–4767.
- [18] B. Li, B. Lu, H. Lin, Y. Wang, F. Zhong, Q. Dou, and Y.-H. Liu, "On the monocular 3d pose estimation for arbitrary shaped needle in dynamic scenes: An efficient visual learning and geometry modeling approach," *IEEE Transactions on Medical Robotics and Bionics*, 2024.
- [19] P. Marra, S. Hussain, M. Caianiello, and F. Ficuciello, "Mpc for suturing stitch automation," *IEEE Transactions on Medical Robotics and Bionics*, 2024.
- [20] K. L. Schwaner, I. Iturrate, J. K. Andersen, P. T. Jensen, and T. R. Savarimuthu, "Autonomous bi-manual surgical suturing based on skills learned from demonstration," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021.
- [21] J. W. Kim, T. Z. Zhao, S. Schmidgall, A. Deguet, M. Kobilarov, C. Finn, and A. Krieger, "Surgical robot transformer (SRT): Imitation learning for surgical tasks," in *8th Annual Conference on Robot Learning*, 2024.
- [22] L. Lipson, Z. Teed, and J. Deng, "Raft-stereo: Multilevel recurrent field transforms for stereo matching," in *2021 International Conference on 3D Vision (3DV)*. IEEE, 2021, pp. 218–227.
- [23] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, et al., "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4015–4026.
- [24] "Directional suture pad, 3d med."
- [25] M. Hwang, B. Thananjeyan, S. Paradis, D. Seita, J. Ichnowski, D. Fer, T. Low, and K. Goldberg, "Efficiently calibrating cable-driven surgical robots with rgbd fiducial sensing and recurrent neural networks," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5937–5944, 2020.
- [26] T. G. Barnes and A. A. Sheikh, "Suture material, knot tying and wound closure in surgery," *Journal of Operating Department Practitioners*, vol. 1, no. 1, pp. 15–20, 2013.
- [27] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.
- [28] B. Thananjeyan, J. Kerr, H. Huang, J. E. Gonzalez, and K. Goldberg, "All you need is luv: Unsupervised collection of labeled images using uv-fluorescent markings," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 3241–3248.
- [29] T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Communications of the ACM*, vol. 27, no. 3, 1984.
- [30] "Ethicon suture, ethibond excel, tapercut, v-37, 4-30", size 2."