SCIENCE ROBOTICS | RESEARCH ARTICLE

MEDICAL ROBOTS

Deep learning–based autonomous retinal vein cannulation in ex vivo porcine eyes

Peiyao Zhang[1,2]*, Peter Gehlbach[3], Russell H. Taylor[1,2,4], Iulian Iordachita[1,2], Marin Kobilarov[1,2]

Retinal vein cannulation (RVC) is an emerging method for treating retinal vein occlusion (RVO). The success of this procedure depends on surgeon expertise and, recently, robotic assistance. This paper proposes an autonomous RVC workflow leveraging deep learning and computer vision. Two Steady-Hand Eye Robots (SHERs) controlled a 100-micrometer metal needle and a medical spatula to execute precise tasks. Three convolutional neural networks were trained to predict needle movement direction and identify contact and puncture events. A surgical microscope with an intraoperative optical coherence tomography (iOCT) system captured the surgical field through a microscope and cross-sectional images (B-scans). The goal was to enable the robot to autonomously carry out the critical steps of the RVC procedure, especially those that are challenging and require expert knowledge. The less technically demanding tasks were assigned to the user, who also supervised the robot during these steps. Our method was tested on 20 ex vivo porcine eyes, achieving a success rate of 90%. In addition, we simulated eye movements caused by breathing on six other ex vivo porcine eyes. With the eyes moving in a sinusoidal pattern, we achieved a success rate of 83%, demonstrating the robustness and stability of the proposed workflow. Our results demonstrate that the autonomous RVC workflow, incorporating deep learning and robotic assistance, achieves high success rates in both static and dynamic conditions, indicating its potential to enhance the precision and reliability of RVO treatment.

INTRODUCTION

Retinal vein occlusion (RVO) is a condition where a vein in the retina becomes occluded by a thrombus. Left untreated, RVO can lead to vision loss due to complications such as macular edema, macular ischemia, ocular neovascularization, vitreous hemorrhage, and tractional retinal detachment. RVO is classified into three types: branch retinal vein occlusion (BRVO), hemiretinal vein occlusion (HRVO), and central retinal vein occlusion (CRVO), with HRVO considered an intermediate condition between BRVO and CRVO (*1*). RVO is the second leading cause of retinal vascular blindness after diabetic retinopathy. Rogers *et al.* (*2*) estimated that 16.4 million adults were affected by RVO in the United States, Europe, Asia, and Australia on the basis of an analysis of pooled data from 68,751 individuals. Among these, 85% were affected by BRVO and occasional HRVO, whereas the remaining 15% had CRVO.

Given the absence of safe, repeatable, and efficacious methods for routine removal of thrombus from retinal veins, the current treatment for RVO focuses on managing the complications resulting from the condition. Intravitreal anti–vascular endothelial growth factor (anti-VEGF) therapy is the most common approach, involving injections of anti-VEGF medications such as bevacizumab (*3*), ranibizumab (*4*), or aflibercept (*5*). These treatments help reduce macular edema by minimizing fluid leakage. They also inhibit abnormal blood vessel growth that leads to hemorrhage and tractional retinal detachment. However, this therapy requires frequent visits for ongoing intravitreal injections. For patients who do not respond well to anti-VEGF therapy, intravitreal steroids like triamcinolone (*6*) and dexamethasone implants (Ozurdex) (*7*) are often used to reduce inflammation and retinal swelling. Focal or grid laser photocoagulation has been historically used as a second-line treatment for patients with BRVO (*8*), providing an additional option for managing the condition. Although these treatments improve clinical outcomes such as visual acuity shortly after treatment, they still present challenges, including frequent visits, high costs, procedural risks, and incomplete recovery. They do not address the underlying cause of RVO.

Retinal vein cannulation (RVC) is an emerging treatment aimed at removing the causative thrombus in the occluded vein(s) to restore blood flow. Today, it involves a surgeon inserting a needle into the eye under a microscope, navigating it freehand to puncture the retinal vein, and injecting anticoagulant drugs such as recombinant tissue plasminogen activator (*9*). Several challenges limit the consistent application of RVC. One major issue is human physiological limitations such as hand tremors. The root mean amplitude of hand tremor at the needle tip during retinal microsurgery has been reported to be 182 μm (*10*). In comparison, the average size of a retinal vein in healthy individuals is reported to be 151 μm (*11*). However, in the case of RVO, the occluded vessels are dilated and therefore wider than those in healthy individuals. To mitigate hand tremors and improve precision, three classes of robotic systems have been developed: "smart" surgical tools, telemanipulation, and comanipulation (*12*, *13*). The handheld instrument Micron, developed at Carnegie Mellon University, demonstrated a 90% reduction in hand tremor in a pointing task and achieved less than 25 μm error in handheld circle tracing tasks (*14*). Telemanipulation systems, such as the Preceyes robotic system (*15*), typically rely on motion scaling to reduce hand tremor. In a comanipulation system like SHER (*16*), the surgeon and the robot share control of a task, typically in a cooperative manner. A comparative study by Esfandiari *et al.* (*17*) demonstrated that cooperative control resulted in shorter task completion time in achieving a vessel-following task than the teleoperation control, whereas the teleoperation mode provided advantages such as motion scaling of hand movements.

---

[1]Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD 21211, USA. [2]Laboratory for Computational Sensing and Robotics, Johns Hopkins University, Baltimore, MD 21211, USA. [3]Wilmer Eye Institute, Johns Hopkins University, Baltimore, MD 21211, USA. [4]Department of Computer Science, Johns Hopkins University, Baltimore, MD 21211, USA.
*Corresponding author. Email: pzhang24@jhu.edu

Depth perception is also challenging when relying on a binocular microscope to view the surgical scene. Several studies have explored methods to estimate depth using only two-dimensional (2D) images (*18*) or additional imaging systems, such as optical coherence tomography (OCT) (*19*). For example, Yang *et al.* (*20*) developed a technique to reach the target in the eye on the basis of instrument shadow convergence observed in 2D microscope images, achieving a mean depth error of 0.347 mm for intraretinal targets. Our previous work trained a convolutional neural network (CNN) to predict the distance vector from the tool-tip position to a user-defined target on the retina, using both a silicone eye phantom (*21*) and ex vivo porcine eyes (*22*). OCT is a noninvasive imaging modality that uses reflected light to generate high-resolution cross-sectional images of the region of interest (ROI). Kang and Cheon (*23*) introduced a handheld common-path swept source OCT for precise subretinal injections (SRIs) with micrometer-level depth control. The intraoperative OCT (iOCT) system integrates OCT imaging with a surgical microscope, providing real-time images alongside B-scans. Sommersperger *et al.* (*24*) proposed a CNN-based segmentation network to segment the tool surface, internal limiting membrane (ILM), and retinal pigment epithelium (RPE) boundaries from iOCT B-scans, estimating depth on the basis of the 3D surface point clouds generated from these segments. In our prior work (*25*) with stationary ex vivo porcine eye experiments, we achieved autonomous subretinal injections with a mean depth error of less than 25 μm by registering the tool-tip position between the robot and iOCT coordinates.

Multiple studies have investigated the feasibility of RVC using various models, including eye phantoms (*26*), open-sky porcine eyes (*27*), and in vivo porcine eyes (*28*). The team from KU Leuven achieved the world's first-in-human, robot-assisted RVC in four patients as part of a phase 1 clinical trial (*29*). The robot system, originally developed at KU Leuven, is now being further advanced by Mynutia. However, there is no standardized workflow for achieving safe, repeatable, and successful RVC. One key factor is the choice of cannulation needle. Most studies used glass micropipettes (*9*, *28*, *30*) that can be made thinner and sharper at the needle tip than metal needles. However, the glass micropipettes are fragile and difficult to visualize under the microscope. Needle tip coating has been proposed as one solution, but tip breakage remains a problem (*27*, *28*, *30*, *31*). Researchers from Yokohama City University (*27*) have successfully injected saline solution into the retinal veins of porcine eyes using a 50-μm micropipette and a 50-μm microneedle near the optic nerve head. This is one of the few groups to achieve successful RVC using metal needles, despite the cost and technical challenges associated with manufacturing metal needles with a tip size of 60 μm or smaller. Because in part of needle availability, only a few robotic systems have, to date, demonstrated RVC using a metal needle, such as SHER from Johns Hopkins University (*32*) and a robotic manipulator developed at the University of Leuven (*33*). Another key factor lies in the method of controlling the needle during RVC. Some studies have used manual operation (*34*, *35*), whereas others have required robotic assistance (*35*, *36*). In both cases, it is challenging for the human operator to stop the robot immediately after detecting venous puncture during RVC. To the best of the authors' knowledge, only one group from the University of California, Los Angeles attempted automated RVC on silicone phantoms (*37*). They defined three steps (vein approach, targeting, and cannulation) for successful vein cannulation using an 80-μm glass micropipette. An OCT C-scan was used to target the vein, with a scanning period of 35.6 s. Cannulation was confirmed by detecting a change in color inside the glass micropipette from the pressure difference between the blood analog in the phantom and the dyed water in the micropipette. The final key factor is the implementation of the remote center of motion (RCM) constraint at the scleral incision point (as seen in Fig. 1A). Any forces applied at this point should be avoided to reduce the potential damage to the sclera and distortion of the ocular surface leading to degradation of the view into the eye. Different groups have implemented the RCM constraint in various ways, such as through mechanical design (*37*), adaptive control to maintain the scleral force at a safe level (*38*), and the establishment of a virtual RCM point using robot kinematics and algorithms (*39*).

To address these challenges in achieving successful RVC, automatic control of SHER is used to eliminate the effects of hand tremor, a surgical microscope with an iOCT system is used to provide precise depth information, a 100-μm metal needle is manipulated for cannulation, and the RCM constraint is enforced algorithmically throughout the autonomous needle navigation process. Unlike prior work (*39*), where the robot is manually controlled by a user pressing keyboard keys to guide the needle's movement, the main contributions of this work are as follows. We developed a workflow to achieve autonomous RVC in ex vivo porcine eyes using a 100-μm beveled bent metal needle. Three CNNs were trained using microscope images and B-scans as inputs to provide critical information for key steps in the workflow. These networks were trained to ensure that the knowledge necessary for guaranteeing the success of RVC was well preserved. This allowed users without expertise in RVC to achieve results comparable to those of an experienced surgeon. The SHER system fully controlled needle navigation, guiding it to a user-defined target using 3D detection and tracking across multiple imaging modalities while strictly enforcing the RCM constraint (see Supplementary Materials and Methods). The user was only involved in the initial preparation, target selection via mouse click, and the infusion step, which was activated by pressing a foot pedal. In addition, an *XYZ* microlinear stage reproduced a sinusoidal motion along the vertical axis, simulating the breathing motion during surgery. The motion amplitude was large enough to move the needle tip out of the vein during the infusion step in the original workflow. To follow the motion pattern of the linear stage, an extra breathing compensation control was implemented on top of the proposed workflow. Last, the proposed workflow was validated using 20 ex vivo porcine eyes fixed on the *XYZ* microlinear stage, achieving a 90% success rate. Six additional ex vivo porcine eyes were tested with sinusoidal motion to demonstrate the workflow's robustness and consistency.

RESULTS

System overview and workflow

Figure 1 provides an overview of the proposed autonomous RVC system. The system contains two SHERs, a microscope with an iOCT system, a vitrectomy machine, and a microlinear stage system. Each robot has five degrees of freedom, consisting of three translational joints and two rotational joints. In this experiment, the human operator was an expert in using the SHER robots, the medical needle and spatula, the vitrectomy machine, and the iOCT system but lacked a clinical background.

The workflow was based in part on our prior work (*39*) involving robot-assisted manual control via a keyboard. However, in this study, we eliminated all manual controls, enabling the robot to operate

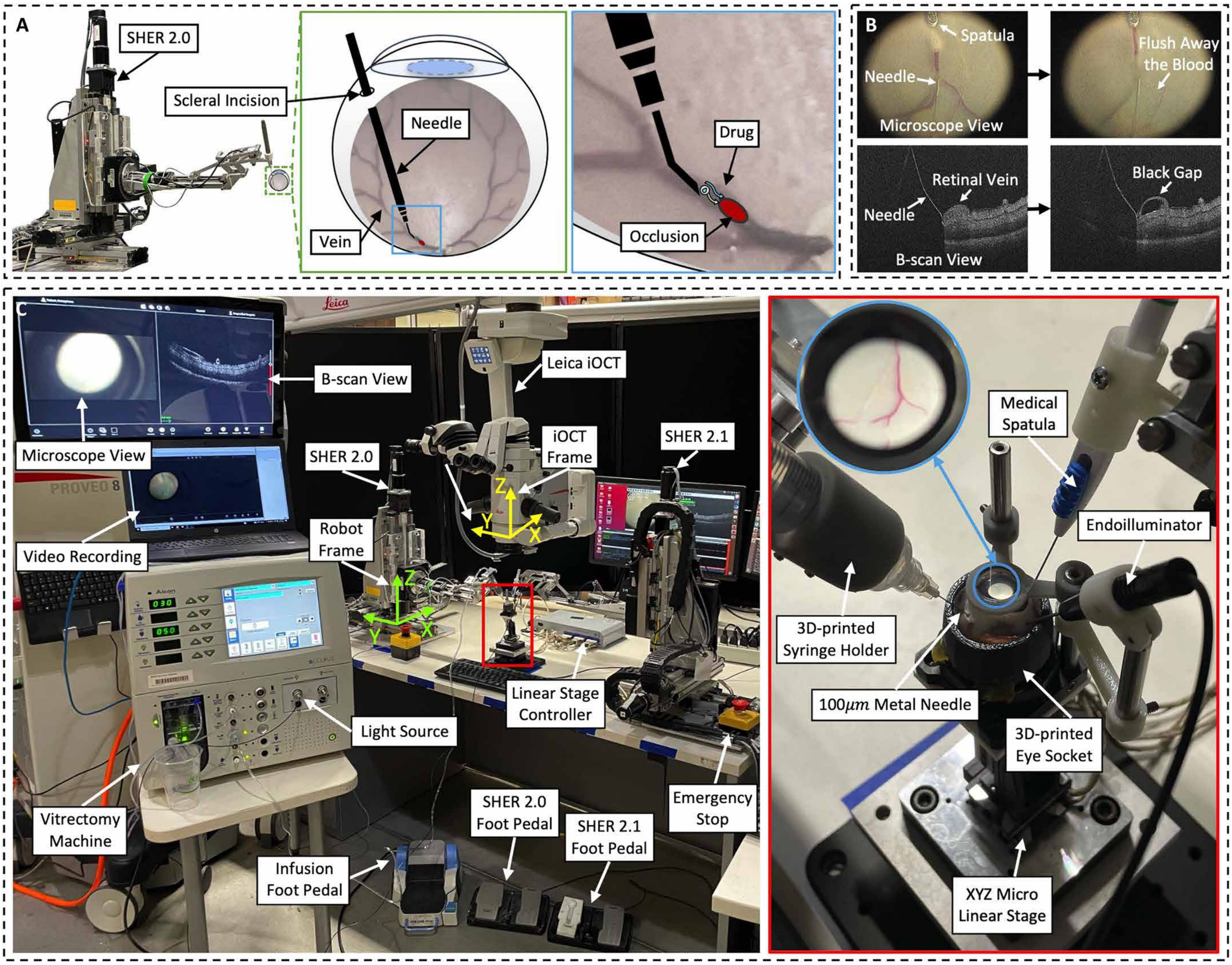

Fig. 1. System overview of autonomous RVC. (**A**) During RVC, a surgical needle is inserted into the eye through a scleral incision to penetrate the occluded retinal vein and deliver anticoagulant drugs to dissolve the blood clot. (**B**) A successful RVC is defined as the blood being flushed away in the microscope view and a black gap being created inside the vein in the B-scan view, both caused by the injected fluid. (**C**) The system features two SHERs (SHER 2.0 and SHER 2.1). SHER 2.0 functions as the primary robot, holding the microneedle, whereas SHER 2.1 serves as the auxiliary robot, holding the medical spatula. A laptop is connected to the Leica iOCT system to record the screen for documentation purposes.

autonomously on the basis of the visual data from the microscope and B-scans. The complete workflow is illustrated in Fig. 2, including initial preparation, horizontal and vertical navigation, puncture and retraction, infusion, and needle removal. After the initial setup, the user was only required to select a target and perform the final infusion step. The target selection was accomplished by clicking on a pixel in the top-down microscope image using a mouse. At the same time, the infusion step required pressing the foot pedal of the vitrectomy machine. All robot movements were driven on the basis of visual feedback from the microscope images and iOCT B-scans. Three deep learning models were trained to provide essential information for robot control.

The initial preparation involved setting up all necessary equipment, preparing the porcine eyes, and preprocessing for cannulation. The equipment setup included configuring the two robots, attaching the needle to SHER 2.0 and the spatula to SHER 2.1, and setting up the vitrectomy machine to provide infusion pressure and the light source. The needle was connected to a syringe, part of the vitrectomy machine's infusion system (viscous fluid control mode). The syringe and spatula were mounted to the end effectors of the two SHERs using 3D-printed tool holders. Filtered water was used in place of viscous fluid for infusion. To fit the porcine eye into our experimental setup, the surrounding tissues were removed, and three scleral incision holes were cut to insert the needle, the endoilluminator, and the medical spatula manually. Vitrectomy cannulas were not used in this study because of constraints associated with the use of a bent needle, given that the 30° bend prevented it from passing smoothly through the cannula, altering both the needle tip and the trocar valve system. The endoilluminator was inserted into the eye to a depth that ensured illumination of the entire ROI

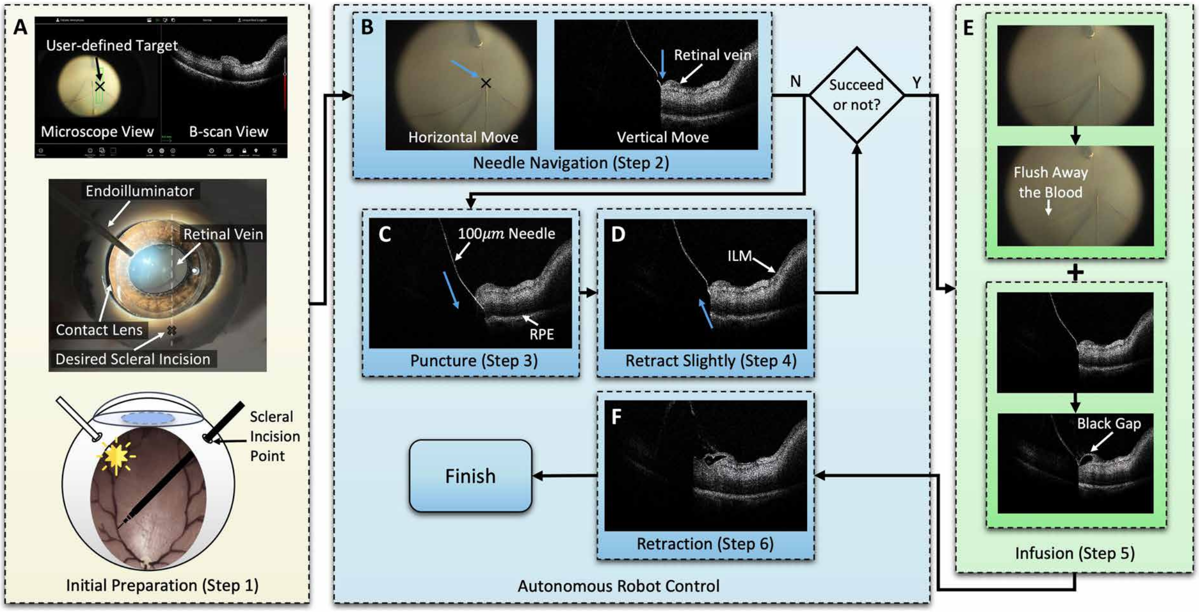

Fig. 2. Illustration of proposed workflow. The workflow consists of six steps: (**A**) initial preparation, (**B**) needle navigation, (**C**) puncture, (**D**) slight needle retraction, (**E**) infusion, and (**F**) final retraction. The initial preparation and infusion steps involve human interaction, whereas the robot fully manages the autonomous robot control block. The blue arrow indicates the direction of the needle movement.

without obstructing the microscope's view or the needle's path. Next, the spatula was inserted to gently press on the upstream section of the targeted vein (thereby moving residual blood to the target site for venous distension), followed by needle insertion to the initial position where it became visible under the microscope, ~4 to 5 mm away from the target. The incision hole for the needle should ensure that the inserted needle aligned roughly with the orientation of the target vein. This was achieved by placing a contact lens on top of the cornea and then using the endoilluminator as a light source to identify the orientation of the vein. The incision on the sclera should be positioned at the intersection of the extended line of the vein and the sclera, ensuring that the inserted needle is adequately aligned with the vein's trajectory, as shown in the initial preparation block of Fig. 2. The eye was then mounted onto the eye socket. The contact lens was placed on top. The gap between the contact lens and the cornea was filled with filtered water as the coupling agent.

To navigate the needle toward a target in the *XY* plane of the microscope, we used visual feedback from the top-down microscope view to control the robot. A CNN was trained to predict the direction of the needle's movement toward a user-defined target, which was specified by a simple mouse click. The SHER controlled the needle with a constant velocity. The target was converted into a 512 pixel–by–512 pixel binary image with a black background and a 40 pixel–by–40 pixel white square, where the center represented the target pixel. We used ResNet18 (*40*) as our backbone, modifying the input size to 512 pixels by 512 pixels by 4 pixels and the number of output classes to nine. The input image consisted of four channels, which included a three-channel RGB microscope image and a one-channel binary target image (Fig. 3A). Previous work (*41*) showed that adding this binary image improved prediction accuracy. The actual image size of the microscope was 1920 pixels by 1080 pixels. To fit the input size, we first performed a center crop to obtain a 900 pixel–by–900 pixel image and then resized it to 512 pixels by 512 pixels.

The first eight output classes corresponded to eight possible movement directions, each representing a 45° range (see Fig. 3A). These eight classes covered all 360° of movement direction. The number of movement direction classes leveraged a balance between the efficiency of needle control and the accuracy of network predictions. More output classes allowed for smoother needle control but required a larger dataset to maintain the same accuracy as models with fewer output classes. The last class was the stop class, which indicated that the needle was very close to the target. The stop class was triggered when the distance between the needle tip and the target pixel was less than 2.5 pixels. This allowed for a maximum error of 1 pixel on one axis and 2 pixels on the other (1 pixel ≈ 11.33 μm).

A contact network was used to determine whether the needle tip contacted the upper outer wall of the blood vessel. The network started predicting contact results from the initial position until a contact was confirmed. The contact network was a binary classification network based on YOLOv8 (*42*). The input to the network was a 512 pixel–by–512 pixel–by–3 pixel B-scan. The output was a binary indicator, where 0 represented no contact and 1 represented contact (see Fig. 3B). Because of the multiple factors such as calibration errors and distortion caused by the contact lens and internal lens inside the eye, the OCT scanning position on the screen might deviate from the actual scanning position on the target. This made it challenging to align the scanning position with the needle and capture the entire needle tip in a single B-scan. In this work, we proposed using a C-scan to capture small-volume information around the target position. This C-scan contained five consecutive B-scans spanning a width of 100 μm. Each B-scan was 1000 pixels by 1024 pixels, representing an area of 6 mm by 3.38 mm. Each B-scan was resized to match the input size required by the network. Because the diameter of the needle tip used was 100 μm, the small-volume C-scan ensured that at least one B-scan captured the tool shaft near its center when the center of the C-scan captured part of the needle. The

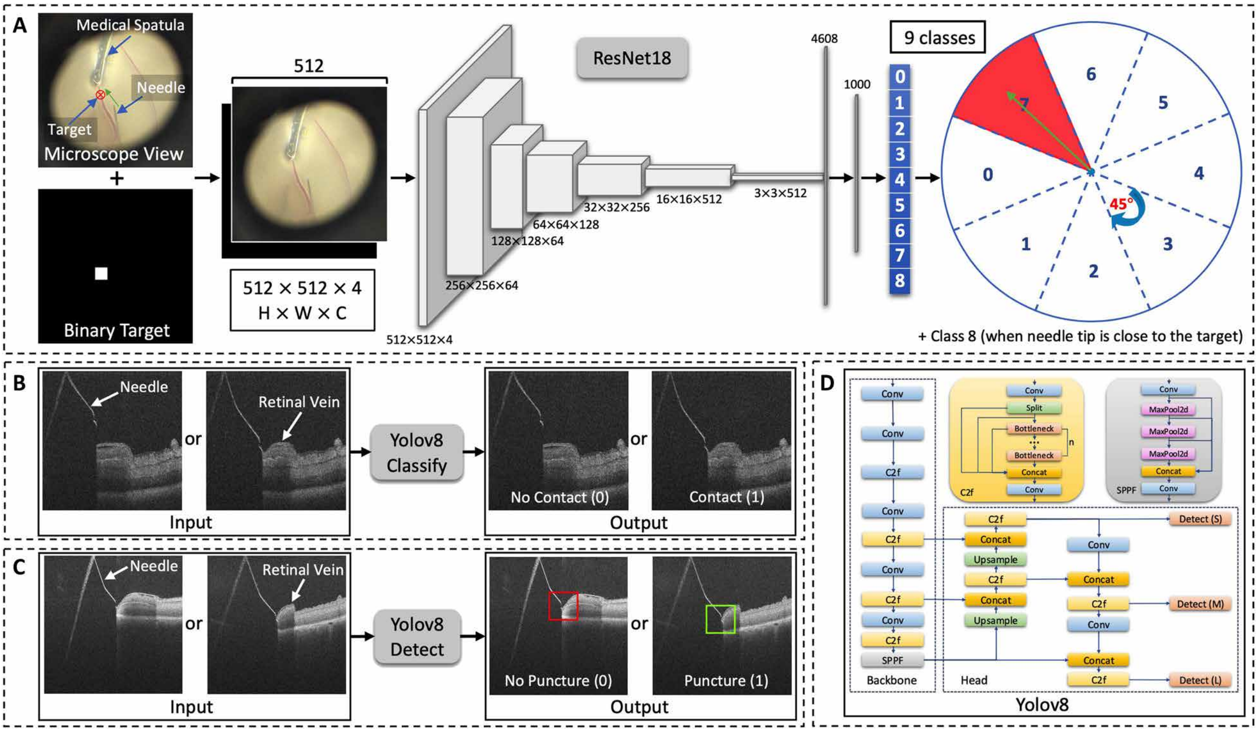

Fig. 3. Network architecture. (**A**) The architecture of the orientation network is based on ResNet18. The output of the orientation network consists of nine classes. The first eight classes (classes 0 to 7) represent eight possible movement directions, each spanning a 45° range, with the actual movement direction located at the center of each class. For example, the predicted class for the left microscope image is class 7 (red sector), whereas the green arrow indicates the actual movement direction controlled by the robot. The final class (class 8) represents the stop class, which signifies that the distance between the needle tip and the target is less than 2.5 pixels. (**B**) The architecture of the contact network is based on the YOLOv8 classification model. (**C**) The architecture of the puncture network is based on the YOLOv8 detection model. (**D**) The detailed architecture of YOLOv8.

C-scan scanning frequency was around 6.67 Hz. The contact network checked each B-scan in the C-scan until at least one predicted binary indicator was 1. In this work, contact was confirmed when the classification probability of the contact class exceeded 90%.

After the needle tip contacted the upper outer wall of the blood vessel, we proceeded to the next step, the puncture. In this work, the puncture step involved a slight push of the needle. Otherwise, if the needle was advanced gradually, the needle tip might not penetrate the vein wall without damaging the retinal pigment epithelium because of the large needle size and the elasticity of the vein wall. Therefore, we controlled and increased the robot's velocity from 0.2 to 5.4 mm/s for 0.1 s. The movement direction was aligned with the tool shaft orientation. During this process, the RCM constraint was automatically maintained, and no additional rotational movement of the needle was required. During the puncture, we measured the distance traveled by the needle tip before and after the puncture. After the puncture, we retracted the needle along the tool shaft by four-fifths of the recorded distance. This retraction step was necessary because the needle tip always appeared to be inside the vein after the puncture in the B-scan (see step 3 in Fig. 2). However, in some cases, retracting the needle revealed that the needle tip was outside the vein. The puncture and retraction steps were repeated until the needle tip was confirmed to be inside the vein.

This confirmation was done through a puncture network based on YOLOv8. Initially, we used a classification model to predict whether an entire B-scan represented a puncture or no puncture, but this approach achieved only ~60% accuracy. This result was expected because we needed to know only whether the needle tip was inside the vein, which was determined by a small area in the B-scan. The rest of the B-scan was irrelevant and added noise to the model. Therefore, we switched to an object detect model for puncture detection. We treated a small rectangular area around the needle tip as an object and manually labeled this object as either puncture or no puncture.

The input to the puncture network was the same 512 pixel–by–512 pixel–by–3 pixel B-scan used for the contact network. The output included the position of the bounding box, the prediction class (0 for no puncture and 1 for puncture), and a confidence score (see Fig. 3C). The YOLOv8 model had two types of confidence: box confidence and class confidence. Box confidence represents how confident the model is that a bounding box contains an object, whereas class confidence represents how certain the model is that the object belongs to a specific class. The final confidence score output by YOLOv8 combined these two confidences, allowing a balance between them. In this work, we considered a puncture confirmed when the confidence score exceeded 60% in ex vivo experiments.

After confirming the puncture, the user pressed the foot pedal (see Fig. 1) to initiate the infusion process. The maximum infusion pressure was set to 20 mmHg when the pedal was fully pressed. The user had to gradually press the pedal to increase the infusion pressure from 0 to 20 mmHg until there was an indication of either a successful or failed infusion. Gradually pressing the pedal helped reduce the risk of damaging the retinal layers caused by a sudden increase in the fluid flow speed. A successful injection was determined when the blood inside the vein was flushed away from the puncture point in the microscope image and a black gap between the upper and lower vessel walls (see step 5 in Fig. 2) was visible in the B-scan because of the infusion of sufficient filtered water. A failed infusion occurred if the filtered water was injected only into the vitreous cavity or resulted in subretinal injection instead of vein cannulation. Because the purpose was to demonstrate the feasibility of the proposed workflow, the volume of liquid infused into the vein was not strictly regulated. In most cases, an overdose of water was injected to signal a successful infusion. After the infusion step, the needle was automatically retracted along its shaft to a safe position, away from the retinal layers, allowing the user to easily remove the needle from the eye without risking damage to the retinal tissue.

Network training results

All three networks were trained on a local NVIDIA TITAN RTX GPU. The inference for all three networks was used with an NVIDIA RTX 3090 GPU at ~15 Hz. Details of data collection and training are provided in Supplementary Materials and Methods.

Orientation network

For the best model of the orientation network, the mean validation error for all classes was 0.08. We also tested the model on 300 unseen microscope images from six intact eyes. Ten random pixels in the ROI were selected as targets for each image. The total number of predictions was 3000, resulting in a mean error of 0.09. When setting the target directly at the tool tip, the mean error increased to 0.6, with 45 false predictions out of 300 images. These results were expected because of the prediction and labeling inaccuracies caused by the blurriness in the microscope image. The error was calculated using the following criteria

$$\text{abs\_error} = |\,\text{prediction} - \text{label}\,|$$

$$\text{error} = \begin{cases} \text{abs\_error}, \ \text{if abs\_error} < 5, \\ 8 - \text{abs\_error}, \ \text{if } 4 < \text{abs\_error} < 8, \\ 4, \ \text{if label} = = 8 \text{ and prediction} \neq \text{label} \end{cases}$$

Given that the eight orientation classes covered 360° of motion direction, class 0 and class 7 were adjacent. Therefore, if the predicted class was 0 and the actual label was 7, the absolute error was considered to be 1 instead of 7. The maximum error between the predicted class and the true label was 4, corresponding to opposite directions. For instances when the actual label was the stop class but the predicted class was not, we assigned the maximum error of 4 to that prediction.

Contact network

For the best contact network model, the validation top-one accuracy was 0.991. We tested the model on 950 unseen B-scans, each labeled as contact or no contact, collected from 15 intact eyes. Of these, 248 B-scans were manually labeled as contact, whereas the remaining 702 were labeled as no contact. The top-one test accuracy was 0.987.

Puncture network

The validation and test results are shown in Table 1 for the best model of the puncture network. A set of 720 unseen B-scans collected from nine intact eyes was used as the test dataset. The evaluation metric mAP50 represents the average precision calculated at an intersection over union (IoU) threshold of 0.50. The mAP50-95 represents the average precision calculated at varying IoU thresholds, ranging from 0.50 to 0.95. We ended up with 2808 validation images instead of 2880 because some duplicate images were automatically removed upon importing the validation dataset into the Roboflow platform (online platform for labeling) (*43*).

Fixed porcine eye experiments

The workflow described in Fig. 2 was tested on 20 ex vivo porcine eyes. The assumption was that the eye would remain static throughout the entire process, given that it was fixed on the *XYZ* microlinear stage with no additional motion introduced. We began with the initial preparation for each trial as detailed in the system overview section. After this preparation, the proposed workflow was executed on the basis of feedback from both microscope images and B-scans. The visualization of the surgical scene and outputs from both robots and networks are shown in Fig. 4.

The user selected the target on the vein by clicking the mouse. The needle then navigated toward the target horizontally, guided by the predicted moving direction from the orientation network. If the needle tip reached the target or a stop class was detected, it moved along the negative *Z* axis until it contacted the upper outer wall of the retinal vein in the B-scan. A contact was confirmed when the probability output of the contact network exceeded 90%. Simultaneously, if the needle tip deviated from the target point while moving downward, the robot adjusted as described in algorithm S1. Upon contact, the robot initiated a vein puncture along the tool's axis,

Table 1. Validation and test results of puncture network. The best model of the puncture network was validated using 2808 unseen images and tested using 720 unseen images. "Val no punc" and "Test no punc" indicate validation and test images without a puncture event, whereas "Val punc" and "Test punc" indicate validation and test images with a puncture. "Val all" and "Test all" include all images in the respective datasets. Image no. denotes the number of images in each subset. Precision, recall, mAP50, and mAP50-95 are standard object detection metrics.

Phase	Image no.	Precision	Recall	mAP50	mAP50-95
Val no punc	1762	0.965	0.957	0.989	0.710
Val punc	1046	0.902	0.970	0.960	0.676
Val all	2808	0.933	0.963	0.974	0.693
Test no punc	418	0.972	0.904	0.983	0.686
Test punc	302	0.872	0.969	0.969	0.663
Test all	720	0.922	0.937	0.976	0.674

retracting slightly afterward. These two steps were repeated until the puncture network confirmed a successful puncture (the confidence score exceeded 60%). The user then pressed the pedal of the vitrectomy machine to initiate the infusion, continuing until the user determined either a successful or failed infusion. After infusion, the robot retracted the needle to a safe position, and the user removed the needle and spatula from the eye, marking the completion of one trial.

The duration of each step, excluding the initial preparation, was recorded. The microscope images and B-scans were captured at a frequency of ~15 Hz to analyze the needle-tissue interaction, and a video was recorded for demonstration purposes. Of the 20 trials, the proposed workflow achieved an 18 of 20 success rate. The true success rate range (69.9 to 97.2%) was calculated using the Wilson score interval with a 95% confidence interval. The duration for each step is provided in Table 2, and a comparison between autonomous RVC and robot-assisted (manual) RVC from previous work (*39*) is included and shown in fig. S1. Figure 5 presents the microscope images and B-scans from six trials, both before and after the infusion step. The two failed cases involved double punctures. In one case, both subretinal injection and RVC were achieved simultaneously. In the other, subretinal injection occurred first, followed by an RVC, as illustrated in fig. S2.

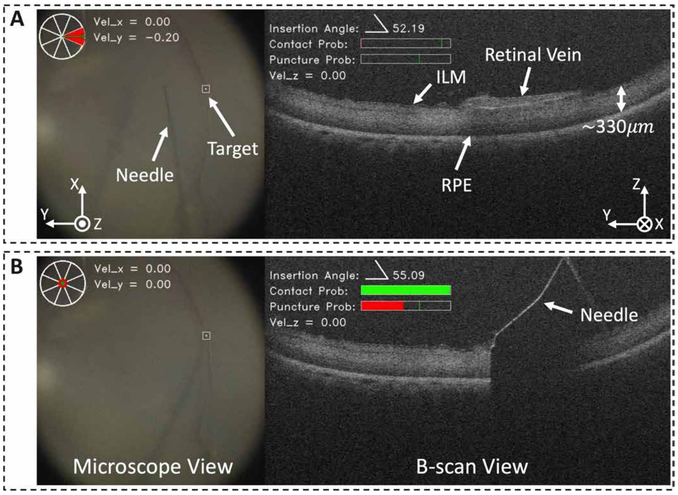

Fig. 4. The visualization of the microscope image, B-scan, needle tip velocities, insertion angle of the needle shaft, and outputs from the three networks. The output of the orientation network is represented as a red sector spanning 45°, with the center of the sector indicating the moving direction, shown by a green arrow in the top left of (**A**). If a stop class is detected, the result is visualized in the top left of (**B**). Vel_x, Vel_y, and Vel_z represent the velocities of the needle tip along the *X* axis, *Y* axis, and *Z* axis, respectively, in the robot frame. The insertion angle is calculated as the angle between the needle shaft and the *XY* plane. The output from the contact network represents the probability of contact, with a green threshold line set at 90% in the probability bar. The output from the puncture network represents the probability of a puncture, corresponding to the confidence score produced by YOLOv8. A threshold of 60% is indicated by the green line in the bar.

Porcine eye experiments with breathing simulation

To demonstrate the robustness of the proposed workflow, we simulated the effect of breathing on eye motion and incorporated an additional control algorithm to follow the breathing pattern. Eye motion caused by physiological factors such as breathing and heartbeat was unavoidable. Several studies have analyzed the frequencies and the amplitudes of these motions using various methods and models. Using OCT distance estimation, Ourak *et al.* (*44*) observed the frequency and amplitude of respiration in conjunction with heart rate. On the basis of their RVC experiments on in vivo porcine eyes, the frequency was ~11 breaths per minute for breathing and 120 heartbeats per minute, with peak-to-peak amplitudes of around 300 and 100 μm, respectively. Lan *et al.* (*45*) used optical coherence elastography to measure the frequencies and amplitudes of respiratory motion on healthy human corneas, resulting in a mean value of 0.13 ± 0.01 Hz and a peak-to-peak amplitude of 68.2 ± 6.6 μm. Kinkelder *et al.* (*46*) measured heartbeat-induced axial motion of the retina using OCT on healthy individuals, reporting ~1 Hz and an average peak-to-peak amplitude of 81 ± 3.5 μm. Using a head-mounted robot, Posselli *et al.* (*47*) estimated the vertical displacement of the eye (126 ± 92 μm) through the optical tracker and rigid body transformation analysis. On the basis of these studies, we controlled the *XYZ* microlinear stage to move vertically following a sinusoidal function

$$z = A\sin\left(2\pi f t\right) \tag{1}$$

where A was set to be 100 μm, f was chosen as 0.23 Hz, and t represents the time variable. The amplitude was selected to be large enough to affect the current workflow if no compensation was applied, for instance, causing the needle tip to move outside the vein because of this breathing motion. The frequency was chosen to be ~14 breaths per minute.

To compensate for this sinusoidal motion, we controlled the robot to follow the sinusoidal pattern on top of the original control logic. An optical flow method was implemented to calculate the vertical difference between two B-scans in real time, as shown in fig. S3.

Table 2. Mean duration for ex vivo porcine eye experiments. The mean duration was calculated separately for steps 2, 3 and 4, 5, and 6 in the proposed workflow. Ten successful trials were recorded for the fixed-eye robot-assisted (manual) experiments (*39*), 18 for the fixed-eye autonomous experiments, and 5 for the breathing simulation experiments.

Step (see Fig. 2)	Mean time (s) ± SD fixed-manual (10 trials)	Mean time (s) ± SD fixed-auto (18 trials)	Mean time (s) ± SD breathing (5 trials)
Step 2 needle navigation	57.45 ± 14.56	30.56 ± 12.60	23.52 ± 10.12
Step 3 puncture and step 4 retract slightly	43.55 ± 26.77	9.08 ± 6.34	10.33 ± 2.47
Step 5 infusion	67.58 ± 50.08	28.93 ± 8.78	27.92 ± 7.08
Step 6 retraction	19.99 ± 3.09	1.40 ± 0.02	1.83 ± 0.10

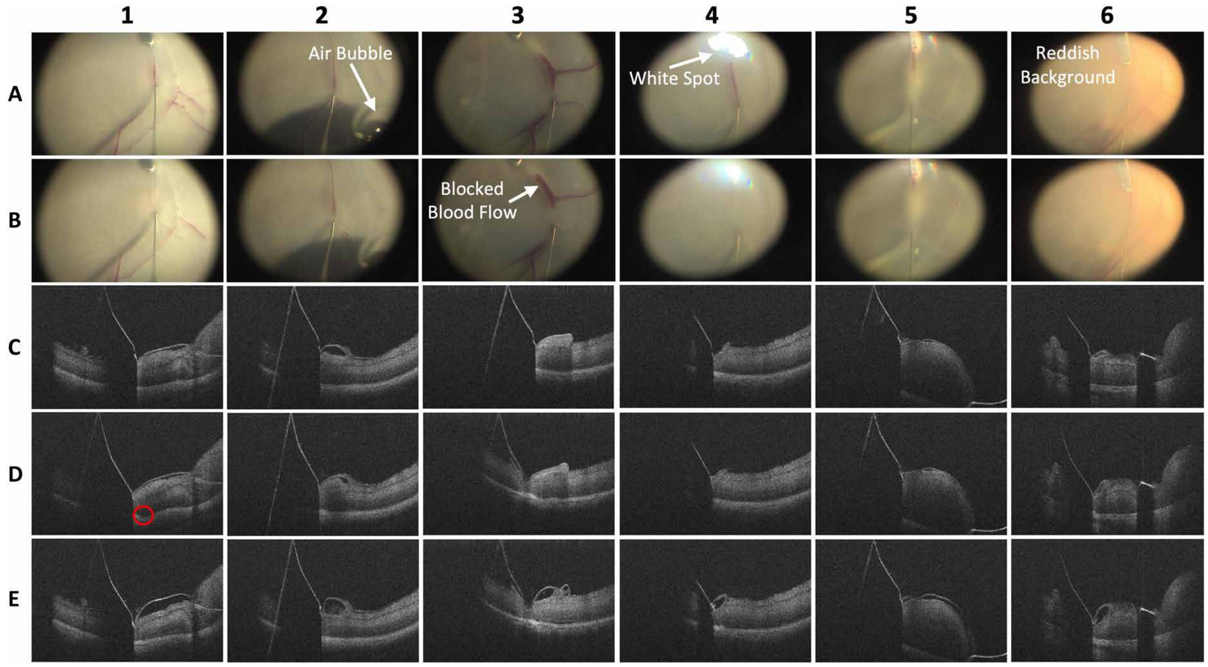

Fig. 5. Demonstration of six successful autonomous RVC trials. Each column represents a trial. Trial 2 contains an air bubble in the view, trial 4 shows a white spot, trial 5 exhibits a detached retina, and trial 6 has a reddish background. (**A**) Top-down microscope view before infusion. (**B**) Top-down microscope view after infusion. In most trials, the infused water flushes the blood out of the vein. In B3, the spatula blocks the blood in the corresponding branch, causing some blood to remain in the vein. (**C**) B-scan after contact and before puncture. (**D**) B-scan after puncture and before infusion. The red circle in D1 highlights a dent caused by the puncture, which disappears after the infusion. (**E**) B-scan after infusion, showing a black gap formed between the upper and lower vein walls.

For each B-scan, the Shi-Tomasi Corner Detector (*48*) was used to detect key features, followed by the Lucas-Kanade method (*49*) to calculate motion between two images. Although outliers caused by medical spatula and surgical needle features might exist, they represented a small fraction of the detected features. In this work, we first calculated all vertical differences between key features in two B-scans and then selected the median value of all differences as the desired distance to compensate. Last, we controlled the robot to move along the Z axis using a fine-tuned velocity gain to account for that desired distance.

The breathing simulation was applied to six ex vivo porcine eye experiments in addition to 20 trials with fixed eye motion. After inserting the needle into the eye, a sinusoidal motion was induced using the XYZ microlinear stage. Simultaneously, an additional control algorithm was used to ensure that the needle followed this sinusoidal pattern based on the distance differences between B-scans calculated using the optical flow method (see Fig. 6). The control gain was fine-tuned so that the peak-to-peak compensation amplitude matched the amplitude specified in Eq. 1. In Fig. 6 (C to F), two trajectories of the needle tip from step 2 to step 6 are shown. The needle tip's position along the Z axis during the infusion step was compared with the desired sinusoidal motion. The same workflow described in Fig. 2 was followed to achieve RVC. A success rate of five of six trials was achieved. The true success rate range (43.6 to 97.0%) was calculated using the Wilson score interval with a 95% confidence interval. Table 2 presents the mean duration of the five successful trials. The failure occurred early when the control gain was too high, causing the needle tip to move out of the vein while following the sinusoidal motion. The infused water did not enter the vein during the infusion but went into the vitreous cavity.

DISCUSSION

This work attempts autonomous RVC in ex vivo porcine eyes using a metal needle, which has not been previously achieved and can be considered an essential step toward safe robot-assisted micromanipulation inside the eye. A success rate of 18 of 20 was achieved in the fixed-eye experiments, and a success rate of 5 of 6 was achieved in the breathing simulation experiments. These results demonstrate the feasibility of autonomous RVC using a 100-μm metal needle. Although the stop class prediction accuracy for the orientation network was only 85% across 300 test images, this does not affect the experimental outcomes. The only observed effect was that the needle occasionally moved around the user-defined target until a stop class was predicted. However, only a tiny fraction of the trials exhibited this motion, and none of these trials ultimately failed to predict the stop class. Data augmentation techniques enhanced the robustness of the orientation network, enabling it to handle various image conditions. For instance, as shown in Fig. 5, an air bubble is present in A2, a white spot caused by the light reflection from the spatula is visible in A4, eye decay causes blurriness in A5, and some porcine eyes exhibit a reddish background for unknown reasons as shown in A6. Despite these conditions being absent from the training dataset, the orientation network can still accurately predict the correct movement direction.

A B-scan is classified as either a contact class or a puncture object on the basis of the contact probability and the box confidence score. If the contact probability or the box confidence score exceeds 50%, the B-scan is classified accordingly. This threshold is raised to 90% for the contact network and 60% for the puncture network to improve prediction accuracy. This adjustment is made because images are labeled only when the distinction between contact or puncture is

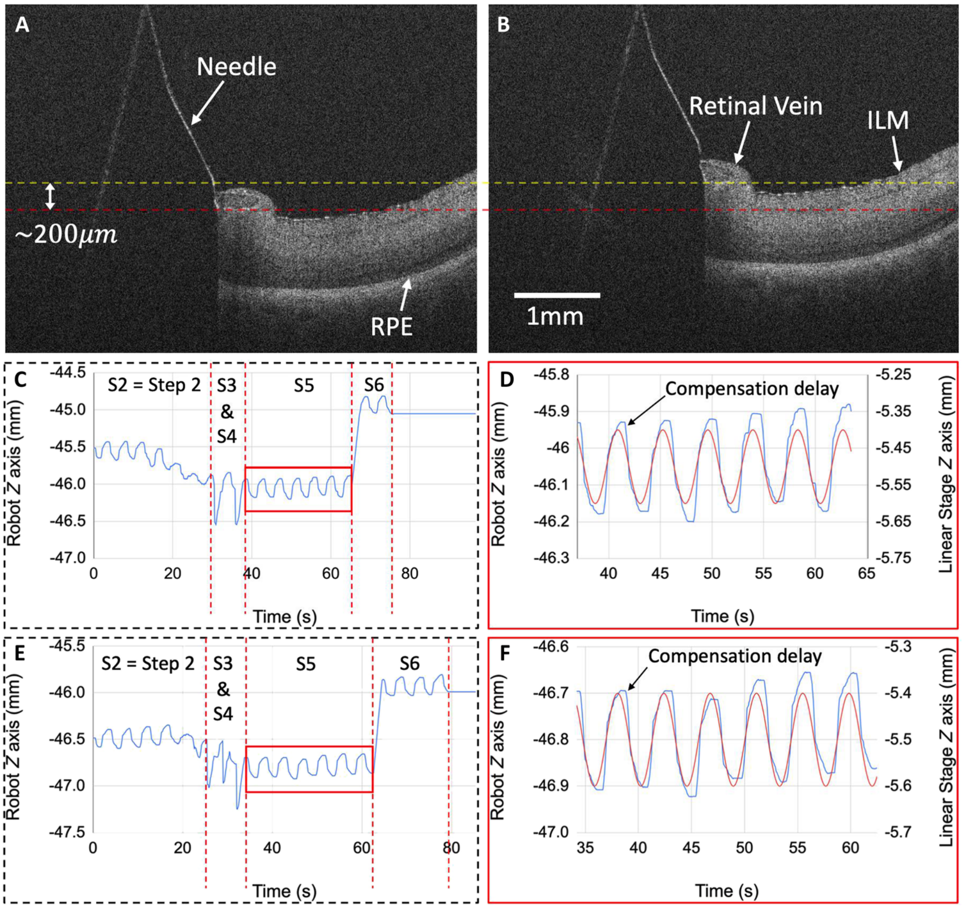

Fig. 6. Motion compensation for breathing simulation. (**A**) The bottom peak position of the retina during the breathing simulation. (**B**) Top peak position of the retina during the breathing simulation. (**C** and **E**) Needle tip positions along the Z axis from step 2 to step 6. In steps 3 and 4, the first trial involves two punctures, whereas the second involves three punctures. (**D** and **F**) Breathing compensation during the infusion step, with the first trial using a higher control gain than the second one.

relatively straightforward during the training of both networks. More challenging images where human classification is uncertain are excluded from the dataset. This approach enhances prediction accuracy, even for unseen images that are difficult for humans to classify.

The networks used in this work are selected to enable real-time inference with limited computational resources while maintaining sufficient accuracy with limited training data. We chose ResNet18 and YOLOv8 for their favorable trade-offs between training effort and performance on the available dataset. For the puncture detection task, we selected the YOLOv8 detection model over the ResNet18 classification model because the ROI is localized around the needle tip rather than the entire image. A supplementary ablation study comparing ResNet18 and YOLOv8 in terms of validation and test accuracy is provided in table S1 and Supplementary Materials and Methods. With the increasing availability of data in the future, we may further enhance network performance by adopting more complex architectures, such as the Vision Transformer (*50*) and ConvNeXt V2 (*51*).

A comparison (Table 2 and fig. S1) between the robot-assisted (manual) and autonomous experiments aimed to demonstrate the feasibility of the proposed autonomous workflow and its advantages over the robot-assisted approach presented in our prior publication (*39*). The mean duration of autonomous experiments was notably shorter than that of robot-assisted experiments. The P values for steps 2, 3 and 4, 5, and 6 between autonomous and robot-assisted experiments were 2.14×10^{-5}, 2.27×10^{-4}, 1.14×10^{-2}, and 1.11×10^{-16}, respectively. This reduction was expected, given that autonomous experiments eliminate procedures that require human judgment, which typically involve time-consuming analysis and decision-making. With the aid of deep learning networks, decision-making is integrated into the training process. Consequently, the user does not need to be a highly experienced expert; instead, they can rely on data labeled by experts. The mean infusion time varied for several reasons. First, the infusion pressure differed between the two methods, being 12 mmHg for the robot-assisted experiments and 20 mmHg for the autonomous experiments. In addition, the user became more proficient with controlling the foot pedal of the infusion system as more experiments were conducted. Last, the equipment played a role: The infusion speed was regulated by the rubber gasket inside the syringe, which was pushed by air from the air compressor of the vitrectomy machine. The edge of the rubber gasket wore out after multiple uses, increasing friction between the gasket and the syringe's inner wall, making the infusion process more difficult. The rubber gasket was replaced with a new one when conducting the autonomous experiments, which likely contributed to the improved performance.

In some trials, the high velocity used during the puncture step may affect the retinal pigment epithelium. However, the retinal pigment epithelium returned to its original shape shortly without sustaining damage, as shown in D1 and E1 in Fig. 5, where D1 shows the state before infusion and E1 shows the state after infusion for the same trial. In some trials, the medical spatula blocked the blood flow, preventing the blood from being flushed away in the microscope image. Despite this, a successful infusion can be confirmed in the B-scan, as illustrated in E3 in Fig. 5, which shows a successful infusion in a trial with blocked blood flow. In one trial, the retina was already detached, yet successful RVC was achieved, demonstrating the robustness of the contact and puncture networks (as seen in D5 and E5 in Fig. 5). Blood reflux was observed in 10 of 18 successful trials, which was expected because of the excess infusion volume and the large needle size. This study did not focus heavily on the reflux effect, given that cadaveric porcine eyes were used, which do not maintain the same blood pressure, and blood vessels no longer react as they do in vivo. Both failed cases involved double punctures, resulting in subretinal injection and RVC. All double punctures occurred after a single puncture, suggesting that the selected velocity for the puncture may be too high for those eyes. Another contributing factor could be the reduced elasticity of the vein as the postmortem eye decays. Alternatively, the lack of blood pressure allowed the anterior and posterior walls of the vein to be appositional at the time of needle puncture. Future work could involve reducing the puncture velocity and increasing the number of punctures, allowing for more gradual needle advancement to enhance safety.

The breathing compensation motion experienced some delays (as shown in Fig. 6) because of how we captured the distance between two B-scans using optical flow. Specifically, we compensated for the motion after detecting it. The control gain influenced the peak-to-peak amplitude of the compensatory motion. This work implemented an open-loop control, meaning that the robot controlled the needle without feedback regarding the needle tip position. The only input was the displacement between two B-scans. This approach was feasible if the infusion period was short for our current setup. As shown in Fig. 6, the needle tip may deviate from the desired sinusoidal pattern over time. This deviation error could accumulate if the infusion time was extended. In the future, we could track the relative position between the needle tip and retinal veins using a real-time segmentation network to adjust the control gain.

This work has some additional limitations. The size of the needle tip prevented it from fully penetrating the vein lumen, which measured ~100 to 250 μm in ex vivo porcine eyes in our experiments. Instead, the aim was to insert a small portion of the needle tip into the vein to enable the infusion process. Given that the remaining part of the beveled needle tip stayed in the vitreous cavity, some infused water inevitably entered the vitreous cavity. Using a thinner needle with a smaller diameter would address this issue but at a higher cost. This study determined and fixed the scanning position during the initial preparation step. Although a small-volume scan was used to increase the scanning width from a single slice (one B-scan) to 100 μm (five B-scans), ensuring that the needle remains visible in the B-scan after reaching the target required substantial effort during the initial preparation. This issue could be mitigated by increasing the scanning volume width, albeit at the cost of reducing scanning frequency. When roughly aligning the needle with the orientation of the vein, this work did not account for situations where such alignment was not possible, including extreme cases in which the needle was nearly perpendicular to the vein. In these uncommon scenarios, several solutions are possible. These include, but are not limited to, using an angled needle cannula introduced via a sheath or a sclerotomy without a trocar; using a second instrument to receive a handoff and reorient the needle approach; using a snake-like robotic end effector to adjust the needle angle; changing the orientation of the eye to better present the target; and using a second tool to modify the orientation of the target vein. If such a scenario occurred, the network would need to be retrained and fine-tuned on the basis of the specific tools used, and the puncture orientation would need to be adjusted to ensure proper alignment with the vein. In addition, external factors such as people moving around the workspace, resting arms on the table, or touching the screen of the Leica iOCT system may cause vibrations in the iOCT's scanning arm, leading to image instability. One potential solution is to mount the iOCT system on the ceiling of the operating room. The breathing simulation used in this work only models vertical motion caused by physiological activity, which is relatively small compared with the horizontal displacement, such as head drift. Although horizontal movement was not discussed in the current work, our previous study (*22*) demonstrated that head drift can be tracked during needle navigation by combining optical flow and deep learning techniques as well. As the next step, we aim to analyze both vertical and horizontal motions caused by breathing and head drift with multiple frequencies and amplitudes.

This work demonstrates the feasibility of combining deep learning methods with robotic assistance to achieve autonomous RVC. In the long term, it represents a foundational step toward precision microsurgical systems capable of performing complex intraocular procedures and delicate retinal treatments with minimal human intervention. As a next phase, we plan to transition the workflow to in vivo animal experiments, providing a critical bridge toward human clinical translation. Future work will also focus on aligning the system with clinical and preclinical data synthesis methodologies (*52*) and ensuring compliance with safety, risk management, and regulatory requirements to support its eventual deployment in real-world surgical environments.

MATERIALS AND METHODS

Study design

This study evaluated the feasibility of autonomous robot-assisted RVC, with comparisons with previously reported robot-assisted manual techniques (*39*). The autonomous method was hypothesized to offer advantages in procedural efficiency, integration of domain knowledge, and dynamic compensation for physiologic motion, such as respiratory-induced eye displacement. Cannulation success was defined by a visible blood flush at the cannulation site on the microscope image and a corresponding gap on B-scan OCT.

Twenty autonomous trials were conducted on stationary eyes, with 18 successful cannulations (90%). An additional six trials simulated respiratory motion, yielding five successes (83.3%). Twelve robot-assisted manual trials achieved 10 successes (83.3%), as reported in the prior work (*39*).

The sample size was based on prior pilot data and the availability of ex vivo specimens. No formal power analysis was performed because of the exploratory nature of the work. Inclusion criteria required complete, synchronized recordings of microscope, B-scan, and iOCT data with time-stamped procedural steps. Trials with corrupted or incomplete data were excluded; no statistical outliers were removed. Success criteria were prespecified and uniformly applied. Each eye underwent a single cannulation attempt, for a total of 26 autonomous trials (stationary and breathing conditions).

Experimental setup for autonomous RVC

Our experimental setup (see Fig. 1) included two SHER robots: SHER 2.0 (*16*) and SHER 2.1 (*53*). SHER 2.0 was the primary robot, with a syringe attached to its end effector using a 3D-printed syringe holder. The syringe was part of a viscous fluid control system (ALCON Constellation Vision System) connected to a vitrectomy machine (Alcon Accurus 800CS) that provided infusion pressure. The user could control the infusion pressure via a foot pedal, and the maximum pressure could be set on the vitrectomy machine. The vitrectomy machine also served as a light source for a 20-gauge endoilluminator (Alcon Accurus Sapphire Wide Angle Endoilluminator). The endoilluminator was mounted on a Q-522 Q-motion *XYZ* microlinear stage (Physik Instrumente) using 3D-printed components. The stage was controlled by an E-873.3QTU linear stage controller (Physik Instrumente). A 100-μm beveled-tip bent metal needle

(MedOne MicroTip Beveled Cannula, 25 g/40 g), with additional polishing at the tip (*39*), was mounted on the syringe for cannulation. SHER 2.1 functioned as an auxiliary robot, holding a medical spatula (DORC extendible curved spatula, 23 g/0.6 mm) to press the vein inside the eye, simulating blood pressure. A surgical microscope with an iOCT system (Leica Proveo 8) provided en face and cross-sectional images of the retina in the ROI. The eye was placed inside a 3D-printed eye socket, attached to the *XYZ* microlinear stage. A contact lens (Oculus Surgical Super View Hassan-Tornambe Contact Lens) with a 36° field of view was mounted on the cornea of each eye to ensure that the microscope and OCT focus in the same range, allowing for clear visualization of both the en face images and B-scans simultaneously.

Needle navigation

Needle navigation was the key step in the complete workflow (initial preparation, needle navigation, puncture and retraction, infusion, and needle removal) that ensured that the needle tip reached the user-defined target on the vein. SHER 2.0 was encoded so that it could be controlled using body velocities, which consisted of three translational velocities along the *XYZ* axes and two rotational velocities along the *X* and *Y* axes. The relationship between SHER's body velocities and joint angle velocities is derived in Supplementary Materials and Methods. Needle navigation involved two parts: horizontal navigation and vertical navigation. The complete navigation logic is presented in algorithm S1. The output of the contact network determined the stop condition for navigation. Navigation should stop if the output was 1 (True), so we constantly checked the contact network output. If no contact was detected, then we controlled the robot toward the user-defined target until contact was confirmed. Starting from the initial position, we first obtained the predicted movement direction from the orientation network. The next step was to control the robot to move horizontally in the robot space toward that direction until it reached the user-defined target on that specific plane. The target on this plane was not the final target on the vein. It was the vertical projection of the target on the vein onto the current *XY* plane at the height of the needle tip. To reach this desired target on the vein, we moved downward along the negative *Z* axis in the robot space. Theoretically, if the robot *Z* axis was perfectly aligned with the iOCT *Z* axis, the needle should remain close to the defined target while moving downward. However, this was not the case in actual experiments, given that we could not guarantee that the robot and iOCT were placed on the same vertical ground. This meant that when moving the needle along the negative *Z* axis, the needle tip might have deviated from its current position in the *XY* plane as seen in the top-down microscope view. For this reason, we continuously predicted the orientation class during horizontal and vertical navigation. If the output class was not a stop class (class 8) during vertical navigation, then we fine-tuned the robot to move toward the target in the current *XY* plane using smaller linear and angular control gains.

Statistical analysis

The true success rate range was calculated using the Wilson score interval. The function proportion_confint from the Python library statsmodels.stats.proportion was used, with α set to 0.05 and method set to wilson.

The statistical analysis comparing the mean duration of each step of the proposed workflow between the autonomous and robot-assisted (manual) experiments was conducted in Python using the SciPy package (version 1.15). Unpaired one-sided *t* tests were used to demonstrate that the mean duration of each step of the autonomous method was shorter than that of the robot-assisted method, with a *P* value of less than 0.05 considered statistically significant.

Supplementary Materials

The PDF file includes:
Materials and Methods
Figs. S1 to S5
Table S1
Legends for movies S1 and S2
Algorithm S1
References (*54*–*59*)

Other Supplementary Material for this manuscript includes the following:
Movies S1 and S2
MDAR Reproducibility Checklist

REFERENCES AND NOTES

1. M. Ip, A. Hendrick, Retinal vein occlusion review. *Asia Pac. J. Ophthalmol.* **7**, 40–45 (2018).
2. S. Rogers, R. L. McIntosh, N. Cheung, L. Lim, J. J. Wang, P. Mitchell, J. W. Kowalski, H. Nguyen, T. Y. Wong, International Eye Disease Consortium, The prevalence of retinal vein occlusion: Pooled data from population studies from the United States, Europe, Asia, and Australia. *Ophthalmology* **117**, 313–319.e1 (2010).
3. D. L. Epstein, P. V. Algvere, G. von Wendt, S. Seregard, A. Kvanta, Bevacizumab for macular edema in central retinal vein occlusion: A prospective, randomized, double-masked clinical study. *Ophthalmology* **119**, 1184–1189 (2012).
4. P. A. Campochiaro, J. S. Heier, L. Feiner, S. Gray, N. Saroj, A. C. Rundle, W. Y. Murahashi, R. G. Rubio, BRAVO Investigators, Ranibizumab for macular edema following branch retinal vein occlusion: Six-month primary end point results of a phase III study. *Ophthalmology* **117**, 1102–1112.e1 (2010).
5. F. G. Holz, J. Roider, Y. Ogura, J.-F. Korobelnik, C. Simader, G. Groetzbach, R. Vitti, A. J. Berliner, F. Hiemeyer, K. Beckmann, O. Zeitz, R. Sandbrink, VEGF Trap-Eye for macular oedema secondary to central retinal vein occlusion: 6-month results of the phase III GALILEO study. *Br. J. Ophthalmol.* **97**, 278–284 (2013).
6. M. S. Ip, I. U. Scott, P. C. VanVeldhuisen, N. L. Oden, B. A. Blodi, M. Fisher, L. J. Singerman, M. Tolentino, C. K. Chan, V. H. Gonzalez, SCORE Study Research Group, A randomized trial comparing the efficacy and safety of intravitreal triamcinolone with observation to treat vision loss associated with macular edema secondary to central retinal vein occlusion: The Standard Care vs Corticosteroid for Retinal Vein Occlusion (SCORE) study report 5. *Arch. Ophthalmol.* **127**, 1101–1114 (2009).
7. J. A. Haller, F. Bandello, R. Belfort Jr., M. S. Blumenkranz, M. Gillies, J. Heier, A. Loewenstein, Y.-H. Yoon, M.-L. Jacques, J. Jiao, X.-Y. Li, S. M. Whitcup, OZURDEX GENEVA Study Group, Randomized, sham-controlled trial of dexamethasone intravitreal implant in patients with macular edema due to retinal vein occlusion. *Ophthalmology* **117**, 1134–1146.e3 (2010).
8. A. R. Berger, A. F. Cruess, F. Altomare, V. Chaudhary, K. Colleaux, M. Greve, A. Kherani, E. D. Mandelcorn, H. Parsons, M.-A. Rhéaume, E. Tourville, Optimal treatment of retinal vein occlusion: Canadian expert consensus. *Ophthalmologica* **234**, 6–25 (2015).
9. K. A. van Overdam, T. Missotten, L. H. Spielberg, Updated cannulation technique for tissue plasminogen activator injection into peripapillary retinal vein for central retinal vein occlusion. *Acta Ophthalmol.* **93**, 739–744 (2015).
10. C. N. Riviere, P. S. Jensen, "A study of instrument motion in retinal microsurgery," in *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE, 2000), vol. 1, pp. 59–60.
11. D. Goldenberg, J. Shahar, A. Loewenstein, M. Goldstein, Diameters of retinal blood vessels in a healthy cohort as measured by spectral domain optical coherence tomography. *Retina* **33**, 1888–1894 (2013).
12. I. I. Iordachita, M. D. De Smet, G. Naus, M. Mitsuishi, C. N. Riviere, Robotic assistance for intraocular microsurgery: Challenges and perspectives. *Proc. IEEE* **110**, 893–908 (2022).
13. E. Vander Poorten, C. N. Riviere, J. J. Abbott, C. Bergeles, M. A. Nasseri, J. U. Kang, R. Sznitman, K. Faridpooya, I. Iordachita, "Robotic retinal surgery," in *Handbook of Robotic and Image-Guided Surgery* (Elsevier, 2020), pp. 627–672.
14. S. Yang, R. A. MacLachlan, C. N. Riviere, Manipulator design and operation of a six-degree-of-freedom handheld tremor-canceling microsurgical instrument. *IEEE ASME Trans. Mechatron.* **20**, 761–772 (2014).

15. J. Cehajic-Kapetanovic, K. Xue, T. L. Edwards, T. C. Meenink, M. J. Beelen, G. J. Naus, M. D. de Smet, R. E. MacLaren, First-in-human robot-assisted subretinal drug delivery under local anesthesia. *Am. J. Ophthalmol.* **237**, 104–113 (2022).
16. A. Üneri, M. A. Balicki, J. Handa, P. Gehlbach, R. H. Taylor, I. Iordachita, "New steady-hand eye robot with micro-force sensing for vitreoretinal surgery," in *2010 3rd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics* (IEEE, 2010), pp. 814–819.
17. M. Esfandiari, J. W. Kim, B. Zhao, G. Amirkhani, M. Hadi, P. Gehlbach, R. H. Taylor, I. Iordachita, "Cooperative vs. teleoperation control of the steady hand eye robot with adaptive sclera force control: A comparative study," in *2024 IEEE International Conference on Robotics and Automation* (*ICRA*) (IEEE, 2024), pp. 8209–8215.
18. J. Yang, Z. Zhao, M. Maier, K. Huang, N. Navab, M. A. Nasseri, "Shadow-based 3D pose estimation of intraocular instrument using only 2D images," in *2024 IEEE International Conference on Robotics and Automation* (*ICRA*) (IEEE, 2024), pp. 1323–1329.
19. J. W. Kim, S. Wei, P. Zhang, P. Gehlbach, J. U. Kang, I. Iordachita, M. Kobilarov, Towards autonomous retinal microsurgery using RGB-D images. *IEEE Robot. Autom. Lett.* **9**, 3807–3814 (2024).
20. J. Yang, Z. Zhao, S. Shen, D. Zapp, M. Maier, K. Huang, N. Navab, M. A. Nasseri, EyeLS: Shadow-guided instrument landing system for target approaching in robotic eye surgery. *IEEE Robot. Autom. Lett.* **9**, 3664–3671 (2024).
21. P. Zhang, J. W. Kim, M. Kobilarov, "Towards safer retinal surgery through chance constraint optimization and real-time geometry estimation," in *2021 60th IEEE Conference on Decision and Control* (*CDC*) (IEEE, 2021), pp. 5175–5180.
22. P. Zhang, J. W. Kim, P. Gehlbach, I. Iordachita, M. Kobilarov, Autonomous needle navigation in retinal microsurgery: Evaluation in ex vivo porcine eyes. arXiv:2301.11839 [cs.RO] (2023).
23. J. U. Kang, G. W. Cheon, Demonstration of subretinal injection using common-path swept source OCT guided microinjector. *Appl. Sci.* **8**, 1287 (2018).
24. M. Sommersperger, J. Weiss, M. A. Nasseri, P. Gehlbach, I. Iordachita, N. Navab, Real-time tool to layer distance estimation for robotic subretinal injection using intraoperative 4D OCT. *Biomed. Opt. Express* **12**, 1085–1104 (2021).
25. P. Zhang, J. W. Kim, P. Gehlbach, I. Iordachita, M. Kobilarov, Autonomous needle navigation in subretinal injections via iOCT. *IEEE Robot. Autom. Lett.* **9**, 4154–4161 (2024).
26. B. Gonenc, N. Tran, P. Gehlbach, R. H. Taylor, I. Iordachita, "Robot-assisted retinal vein cannulation with force-based puncture detection: Micron vs. the steady-hand eye robot," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (*EMBC*) (IEEE, 2016), pp. 5107–5111.
27. K. Kadonosono, A. Arakawa, S. Yamane, E. Uchio, Y. Yanagi, An experimental study of retinal endovascular surgery with a microfabricated needle. *Invest. Ophthalmol. Vis. Sci.* **52**, 5790–5793 (2011).
28. K. Willekens, A. Gijbels, L. Schoevaerdts, L. Esteveny, T. Janssens, B. Jonckx, J. H. Feyen, C. Meers, D. Reynaerts, E. Vander Poorten, P. Stalmans, Robot-assisted retinal vein cannulation in an in vivo porcine retinal vein occlusion model. *Acta Ophthalmol.* **95**, 270–275 (2017).
29. A. Gijbels, J. Smits, L. Schoevaerdts, K. Willekens, E. B. Vander Poorten, P. Stalmans, D. Reynaerts, In-human robot-assisted retinal vein cannulation, a world first. *Ann. Biomed. Eng.* **46**, 1676–1685 (2018).
30. M. D. de Smet, J. M. Stassen, T. C. Meenink, T. Janssens, V. Vanheukelom, G. J. Naus, M. J. Beelen, B. Jonckx, Release of experimental retinal vein occlusions by direct intraluminal injection of ocriplasmin. *Br. J. Ophthalmol.* **100**, 1742–1746 (2016).
31. K. Willekens, A. Gijbels, J. Smits, L. Schoevaerdts, J. Blanckaert, J. H. Feyen, D. Reynaerts, P. Stalmans, Phase I trial on robot assisted retinal vein cannulation with ocriplasmin infusion for central retinal vein occlusion. *Acta Ophthalmol.* **99**, 90–96 (2021).
32. A. Alamdar, N. Patel, M. Urias, A. Ebrahimi, P. Gehlbach, I. Iordachita, Force and velocity based puncture detection in robot assisted retinal vein cannulation: In-vivo study. *IEEE Trans. Biomed. Eng.* **69**, 1123–1132 (2021).
33. A. Gijbels, K. Willekens, L. Esteveny, P. Stalmans, D. Reynaerts, E. B. Vander Poorten, "Towards a clinically applicable robotic assistance system for retinal vein cannulation," in *2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics* (*BioRob*) (IEEE, 2016), pp. 284–291.
34. L.-O. Hattenbach, J. Puchta, I. Hilgenberg, Experimental endoscopic endovascular cannulation: A novel approach to thrombolysis in retinal vessel occlusion. *Invest. Ophthalmol. Vis. Sci.* **53**, 42–46 (2012).
35. T. Ueta, T. Nakano, Y. Ida, N. Sugita, M. Mitsuishi, Y. Tamaki, Comparison of robot-assisted and manual retinal vessel microcannulation in an animal model. *Br. J. Ophthalmol.* **95**, 731–734 (2011).
36. M. D. de Smet, T. C. Meenink, T. Janssens, V. Vanheukelom, G. J. Naus, M. J. Beelen, C. Meers, B. Jonckx, J.-M. Stassen, Robotic assisted cannulation of occluded retinal veins. *PLOS ONE* **11**, e0162037 (2016).
37. M. J. Gerber, J.-P. Hubschman, T.-C. Tsao, Automated retinal vein cannulation on silicone phantoms using optical-coherence-tomography-guided robotic manipulations. *IEEE ASME Trans. Mechatron.* **26**, 2758–2769 (2020).
38. A. Ebrahimi, M. G. Urias, N. Patel, R. H. Taylor, P. Gehlbach, I. Iordachita, Adaptive control improves sclera force safety in robot-assisted eye surgery: A clinical study. *IEEE Trans. Biomed. Eng.* **68**, 3356–3365 (2021).
39. P. Zhang, P. Gehlbach, M. Kobilarov, I. Iordachita, "A feasible workflow for retinal vein cannulation in ex vivo porcine eyes with robotic assistance," in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (*EMBC*) (IEEE, 2024), pp. 1–5.
40. K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), pp. 770–778.
41. J. W. Kim, P. Zhang, P. Gehlbach, I. Iordachita, M. Kobilarov, "Towards autonomous eye surgery by combining deep imitation learning with optimal control," in *Conference on Robot Learning* (PMLR, 2021), pp. 2347–2358.
42. G. Jocher, A. Chaurasia, J. Qiu, YOLO by Ultralytics (2023); https://github.com/ultralytics/ultralytics.
43. B. Dwyer, J. Nelson, T. Hansen, Roboflow (Version 1.0) [Software] (2024); https://roboflow.com.
44. M. Ourak, J. Smits, L. Esteveny, G. Borghesan, A. Gijbels, L. Schoevaerdts, Y. Douven, J. Scholtes, E. Lankenau, T. Eixmann, H. Schulz-Hildebrandt, G. Hüttmann, M. Kozlovszky, G. Kronreif, K. Willekens, P. Stalmans, K. Faridpooya, M. Cereda, A. Giani, G. Staurenghi, D. Reynaerts, E. B. Vander Poorten, Combined OCT distance and FBG force sensing cannulation needle for retinal vein cannulation: In vivo animal validation. *Int. J. Comput. Assist. Radiol. Surg.* **14**, 301–309 (2019).
45. G. Lan, B. Gu, K. V. Larin, M. D. Twa, Clinical corneal optical coherence elastography measurement precision: Effect of heartbeat and respiration. *Transl. Vis. Sci. Technol.* **9**, 3 (2020).
46. R. de Kinkelder, J. Kalkman, D. J. Faber, O. Schraa, P. H. Kok, F. D. Verbraak, T. G. van Leeuwen, Heartbeat-induced axial motion artifacts in optical coherence tomography measurements of the retina. *Invest. Ophthalmol. Vis. Sci.* **52**, 3908–3913 (2011).
47. N. R. Posselli, P. S. Wellborn, P. S. Bernstein, R. J. Webster III, J. J. Abbott, Head-mounting surgical robots for passive compensation of patient motion. *J. Med. Robot. Res.* **9**, 2350006 (2024).
48. J. Shi, "Good features to track," in *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 1994), pp. 593–600.
49. B. D. Lucas, T. Kanade, "An iterative image registration technique with an application to stereo vision," in *IJCAI'81: 7th International Joint Conference on Artificial Intelligence* (ACM, 1981), vol. 2, pp. 674–679.
50. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010.11929 [cs.CV] (2020).
51. S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I. S. Kweon, S. Xie, "Convnext v2: Co-designing and scaling convnets with masked autoencoders," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2023), pp. 16133–16142.
52. N. R. Posselli, E. S. Hwang, Z. J. Olson, A. Nagiel, P. S. Bernstein, J. J. Abbott, Head-mounted surgical robots are an enabling technology for subretinal injections. *Sci. Robot.* **10**, eadp7700 (2025).
53. X. He, D. Roppenecker, D. Gierlach, M. Balicki, K. Olds, P. Gehlbach, J. Handa, R. Taylor, I. Iordachita, "Toward clinically applicable steady-hand eye robot for vitreoretinal surgery," in *ASME International Mechanical Engineering Congress and Exposition* (American Society of Mechanical Engineers, 2012), vol. 45189, pp. 145–153.
54. Supervisely, "All computer vision in one platform"; https://supervisely.com.
55. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization. arXiv:1412.6980 [cs.LG] (2014).
56. R. M. Murray, Z. Li, S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation* (CRC Press, 2017).
57. P. Chiacchio, S. Chiaverini, L. Sciavicco, B. Siciliano, Closed-loop inverse kinematics schemes for constrained redundant manipulators with task space augmentation and task priority strategy. *Int. J. Rob. Res.* **10**, 410–425 (1991).
58. J. N. Weiss, L. A. Bynoe, Injection of tissue plasminogen activator into a branch retinal vein in eyes with central retinal vein occlusion. *Ophthalmology* **108**, 2249–2257 (2001).
59. M. Sohan, T. Sai Ram, R. Reddy, C. Venkata, "A review on yolov8 and its advancements," in *International Conference on Data Intelligence and Cognitive Informatics* (Springer, 2024), pp. 529–545.

Acknowledgments: We thank N. Navab and M. Ali Nasseri for discussion and feedback on the experiments. **Funding:** This work was supported by the US National Institutes of Health under the grants 2R01EB023943-04A1 and 1R01EB025883-01A1 and partially by Johns Hopkins University internal funds. **Author contributions:** I.I., M.K., and P.Z. conceived and designed the project. P.Z. collected the data, labeled the data, trained the networks, implemented algorithms and visualizations, conducted the experiments, analyzed the results, and wrote the manuscript. P.G., R.H.T., I.I., and M.K. supervised the project, provided feedback, and reviewed

the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to support the conclusions of this manuscript are included in the main text or Supplementary Materials. The data and code necessary for reproducing this work are available at https://doi.org/10.5061/dryad.3ffbg79zd.

Submitted 27 January 2025
Accepted 18 November 2025
Published 17 December 2025
10.1126/scirobotics.adw2969

Deep learning–based autonomous retinal vein cannulation in ex vivo porcine eyes

Peiyao Zhang, Peter Gehlbach, Russell H. Taylor, Iulian Iordachita, and Marin Kobilarov

Sci. Robot. **10** (109), eadw2969. DOI: 10.1126/scirobotics.adw2969

Science Robotics (ISSN 2470-9476) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Robotics* is a registered trademark of AAAS.