

Vision-based robot manipulation of transparent liquid containers in a laboratory setting

Daniel Schober¹, Ronja Güldenring¹, James Love² and Lazaros Nalpantidis¹

Abstract—Laboratory processes involving small volumes of solutions and active ingredients are often performed manually due to challenges in automation, such as high initial costs, semi-structured environments and protocol variability. In this work, we develop a flexible and cost-effective approach to address this gap by introducing a vision-based system for liquid volume estimation and a simulation-driven pouring method particularly designed for containers with small openings. We evaluate both components individually, followed by an applied real-world integration of cell culture automation using a UR5 robotic arm. Our work is fully reproducible: we share our code at <https://github.com/DaniSchober/LabLiquidVision> and the newly introduced dataset *LabLiquidVolume* is available at <https://data.dtu.dk/articles/dataset/LabLiquidVision/25103102>.

I. INTRODUCTION

In life science and chemistry labs, robots are already used in highly standardized processes that deal with large amounts of samples and require high repeatability. However, experiments with small amounts of solutions and active ingredients samples are still mostly executed manually by specialists in the lab. Among the reasons for this automation gap are the high initial investment costs, semi-structured nature of R&D labs, high variability in research protocols, and the current lack of lower-cost interim labor-saving automation options, as attested in [1], [2].

In this work, we introduce a flexible, modular and cost-effective robotic system to allow the flexible automation of small-scale laboratory experiments involving manipulation of liquid containers. Our robot is guided by computer vision to estimate liquid volumes and it adopts a simulation-driven pouring approach particularly designed for containers with small openings.

Our proposed vision-based liquid volume estimation method, illustrated in Fig. 1a, uses a novel two-step Convolutional Neural Network (CNN) architecture. In the first step, a single RGB input image is processed by the first CNN to predict the segmentation and depth of the transparent container and the contained liquid. For the training of the first step, we benefit from existing datasets targeting transparent containers such as *TransProteus* [3] and *Vector-LabPics* [4]. These intermediate predictions are further processed by a second CNN to give an estimate of the liquid volume in the container. For the training of the second network, we create

The authors would like to thank Novo Nordisk for providing all the hardware equipment for the developed robotic setup and the team of Automation and Process Optimization for their invaluable support.

¹Technical University of Denmark, Kongens Lyngby, Denmark
{s212599, ronjag, lanalpa}@dtu.dk

²Automation and Process Optimization, Novo Nordisk, Malmö, Denmark

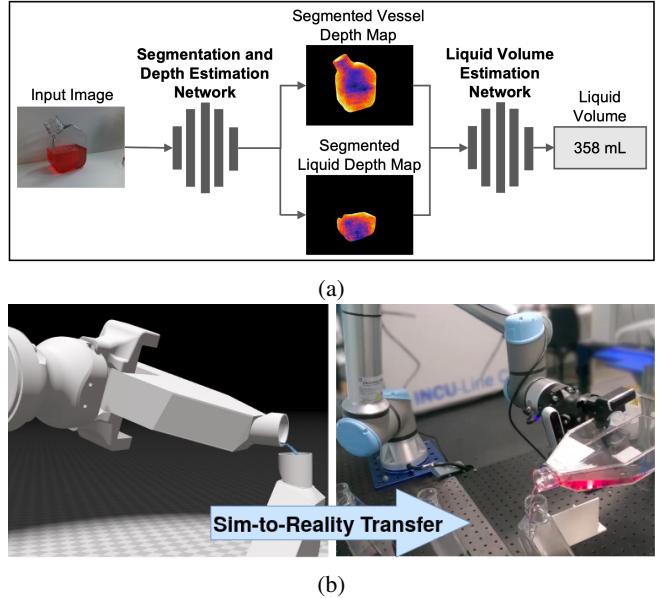


Fig. 1: We present a robotic setup for cell culture automation using exclusively existing laboratory equipment. (a) First, we propose a two-step model architecture to estimate the liquid volume in transparent laboratory containers using a single RGB image. (b) Secondly, from a large pool of simulated pouring trajectories, the simulation that approximates the real-world state best is performed on the UR5. Hereby, the pouring strategy is constrained to rotate around the liquid exit point which is crucial for receiving containers with small openings such as cell culture flasks.

a new dataset *LabLiquidVolume* with 5,451 RGB images of transparent laboratory containers manually labeled with the liquid volume inside the container. Our two-step architecture allows us to keep the size of the newly introduced dataset for training the second step relatively small, as we are able to build upon the knowledge distilled from the publicly available datasets during the first step.

Additionally, we have developed a strategy for the robot to pour from containers with small openings, by confining the robot arm to rotate around a fixed liquid exit point. From a large pool of simulated pouring trajectories applicable for small openings, the simulation that approximates the real-world state best is selected to be executed in the real-world automation setup; this is illustrated in Fig. 1b. Pouring a specific amount of liquid into a container with a small opening is an important and frequently reoccurring manipulation

task for robots in laboratory environments. Hereby we set a focus on the flexible usage, i.e. the pouring procedure can be performed in different experimental setups with visual and auditory variations.

We showcase our developed system on automated cell culture processes, which include various pouring processes such as media changing and passaging cells. We integrate a UR5 robotic arm to fully automate the cell culture process using exclusively existing laboratory equipment. Videos of the automated cell culture process including the different autonomous pouring steps can be seen here: <https://github.com/DaniSchober/LabLiquidVision>.

Our contributions in this work can be summarized as follows.

- We propose a novel two-step CNN architecture to perform liquid volume estimation for transparent laboratory containers. The two-step nature of our architecture has the advantage that we can benefit from a large pool of relevant existing datasets in the first stage and predict intermediate properties such as the depth and segmentation of liquid and vessel.
- We publish a newly created dataset *LabLiquidVolume* with 5,451 images to enable the training of actual liquid estimation in the second step of our proposed two-step CNN architecture. To the best of our knowledge, it is the first real-world dataset including labels for the liquid volume in the transparent container. The data can be downloaded at <https://data.dtu.dk/articles/dataset/LabLiquidVision/25103102>.
- We develop a new pouring strategy for containers with small openings by confining the robot arm to rotate around a fixed liquid exit point, inspired by human execution, and showcase the operation of the entire system on a physical robotic setup.

II. RELATED WORK

Laboratory processes at small scale are often performed manually because the automation of those processes causes high initial costs and includes protocol variability which in most cases does not pay off. Robots constitute a potential solution for the automation of small-scale and variable laboratory processes because they can be deployed for one-to-one replacement of manual labor without the need for any additional specialized equipment as is the case for the automation of large-scale processes. Hereby, different kinds of robots can be relevant such as mobile robots [5], [6] (usually including a moveable arm), stationary robots [7], [8], and fluidic robots [9]. In our work, we aim to modularize the pouring process using standard laboratory containers with small openings, which is relevant for various cell culture procedures and life science laboratory processes. It includes the automated volume estimation of liquids in transparent containers as well as a pouring strategy complying with the small openings of the containers.

A. Vision-based Liquid Volume Estimation

The following publications have been identified to perform vision-based liquid volume estimation in transparent containers. Mottaghi et al. [10] formulate the task as a classification problem and define discrete classes for ranges of the container sizes (50, 100, 200, 300, 500, 750, 1000, 2000, 3000 mL, and above) as well as liquid fillings in percentage (0%, 33%, 50%, 66%, 100% of the container volume). The classification CNN is trained on single RGB images. Zhu et al. [11] enrich the input RGB image with the measurement of a customized tactile sensor and train a multi-modal CNN which predicts the liquid volume with an error of 2 mL for volumes of 40-80 mL. Cobo et al. [12] determine the wine volume on a single-view RGB image with a mean average error lower than 10 mL for volumes of 50-300 mL.

The above-presented approaches have been applied to domains outside laboratory settings. In laboratory settings, only the detection of materials (including liquids) in transparent containers has been heavily researched by Eppel et al.. Initially, they investigated traditional image processing methods such as edge detection [13] and graph-cuts [14]. Recently, they have focused on liquid detection using CNNs trained on a relatively small real-world dataset *Vector-LabPics* [4] as well as a synthetically generated larger dataset *TransProteus* [3].

In our work and to the best of our knowledge, we present the first vision-based liquid volume estimation approach developed for laboratory settings. We take advantage of the publicly available laboratory datasets *Vector-LabPics* [4] and *TransProteus* [3]. We apply a two-step deep learning architecture with two consecutive models. The first model generates a segmentation and depth map, followed by a regression model to estimate the actual liquid volume.

B. Autonomous Pouring

Existing pouring strategies tackle pouring into wide openings, like e.g. in kitchen assistant tasks, leading to moving liquid exit points while pouring. The majority of pouring strategies [15], [16], [17], [18], [19], [20] perform a rotation around the wrist of the robot, while others use more advanced approaches for the trajectory planning [21], [22].

For our laboratory automation, we confine the pouring strategy to rotate around a fixed liquid exit point, which is especially relevant for pouring in small openings, such as cell culture flasks. Finally, the constrained pouring strategy is learned using approximate simulation as in [19], [15]. The final robot movement is performed according to the simulated movement that approximates the real-world scenario closest.

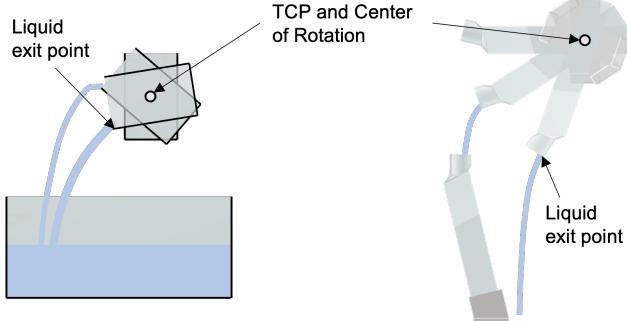


Fig. 2: Visualization of the issue with pouring movements only considering a wrist rotation. Left: Successful pouring only using rotation because of a receiving container with a large opening surface and a pouring container with a small height. Right: Liquid is spilled when only rotating a cell culture flask around the TCP.

III. METHODS

A. Vision Model for Liquid Volume Estimation

The objective is to predict the volume inside transparent laboratory vessels based on a single RGB image. Inspired by Graikos et al. [23] and illustrated in Fig. 1a, we propose a two-step deep learning architecture consisting of two sub-models: (1) the *Segmentation and Depth Estimation (SDE) Network* and (2) the *Liquid Volume Estimation (LVE) Network*.

The SDE Network is based on the work presented by Eppel et al. [3]. In contrast to Eppel et al. [3], we take advantage of a constant camera setup with known camera parameters, enabling the prediction of a 2D depth map instead of a 3D XYZ map. For the architecture, a modified DeepLab-v3 is used, which takes as input an RGB image and outputs both a segmentation map and a depth map. Note that the segmentation map covers three classes: vessel, vessel opening, and liquid.

The LVE Network processes the output of the SDE Network to predict the vessel's liquid volume in mL. The architecture is a simple down-sampling Convolutional Neural Network, consisting of five convolutional layers, followed by four fully connected layers.

B. Pouring in Small Openings

Robotic pouring with rotation of the wrist is not applicable for containers with small openings, such as cell culture flasks, because the exit point of the liquid is constantly changing. As illustrated in Fig. 2, this would result in a large amount of spillage. In research laboratories, particularly where liquids can pose a safety hazard, minimizing the volume of spills is of critical importance.

We propose a method for pouring liquids into containers with a narrow opening inspired by human movements. It mimics the human approach of initially aligning the pouring

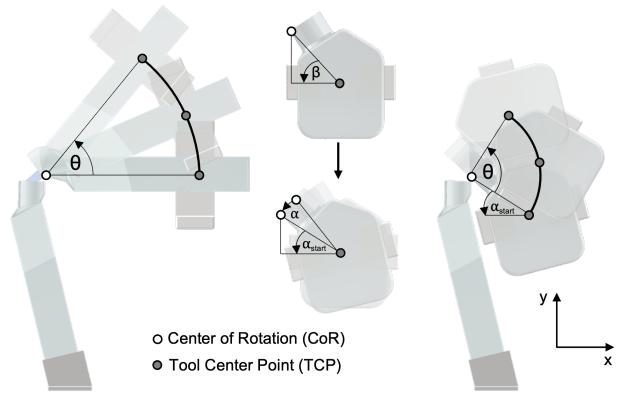


Fig. 3: Visualization of the pouring movement for two exemplary laboratory containers. θ represents the pouring angle, β the angle between the TCP and the CoR of the bottle in the horizontal position, α the rotation of the bottle to the same angle as the receiving flask ($\alpha = 14.5$), and α_{start} the start angle between the TCP and the CoR. Left: Pouring with the cell culture flask. Center: Explanation of the start position of the bottle. Right: Pouring with the media/washing solution bottle.

container's opening with the receiving container's opening, followed by rotating the pouring container around the fixed liquid exit point. In this method, the tool center point TCP undergoes both positional and orientational changes during the movement. The rotation angle (θ) around the liquid exit point, referred to as the center of rotation (CoR), determines the coordinates of the TCP in the XY plane. Depending on the pouring container, the pouring path is calculated based on the distance between the center of rotation and tool exit point $l = |TCP_{start} - CoR|$, the angle β between them, and the rotation at the start α . TCP_x and TCP_y for a specific θ can then be calculated as follows:

For $\theta \leq \alpha_{start}$:

$$TCP_x(\theta) = TCP_{x,start} - l \cos(\alpha_{start}) + l \cos(\alpha_{start} - \theta)$$

$$TCP_y(\theta) = TCP_{y,start} + l \sin(\alpha_{start}) - l \sin(\alpha_{start} - \theta)$$

For $\theta > \alpha_{start}$:

$$TCP_x(\theta) = TCP_{x,start} - l \cos(\alpha_{start}) + l \cos(\theta - \alpha_{start})$$

$$TCP_y(\theta) = TCP_{y,start} + l \sin(\alpha_{start}) + l \sin(\theta - \alpha_{start})$$

The proposed pouring movement is illustrated in Fig. 3.

C. Simulation-to-Reality Transfer of Liquid Pouring

Inspired by Kennedy et al. [15] and Lopez-Guevara et al. [19], a simulation is used to pre-evaluate various pouring movements and identify the optimal robot movement for specific input parameters in real-world scenarios. The main advantage of using pre-simulated pouring results is the reduction of numerous time-consuming real pouring attempts, allowing for more efficient experimentation and adaptation. Moreover, a detailed target volume precision achieved in

simulations is possible because of the possibility of varying various pouring parameters.

We use $\mathbf{p} \in \mathbf{P}$ to denote a set of parameters, $\mathbf{p}^{(i)}$ to denote the i^{th} dimension of the set of parameters, and \mathbf{p}^* to represent an optimal set of parameters. Hence, the combination of the remaining 2D *action parameter* space and the 2D *simulation parameter* space results in a reduced 4D parameter space \mathbf{p} . This consists of the pouring container $\mathbf{p}^{(0)} = c$, the start volume in the pouring container $\mathbf{p}^{(1)} = V_{start}$, the stop angle $\mathbf{p}^{(2)} = \theta_{stop}$, and the stop time $\mathbf{p}^{(3)} = t_{stop}$. The angular velocity and start position were kept constant.

The start volume C_{start} , the received volume $C_{received}$, and the spilled volume C_{spill} of each scene were stored. Based on the given real-world parameters, which are the pouring container, the start volume, and the desired volume to be poured V_{goal} , the best-fitting simulated scene with an optimal set of parameters \mathbf{p}^* can then be chosen by iterating through the results of the simulation and selecting the one with the lowest cost:

$$C(\mathbf{p}) = C_{start} + C_{received} + C_{spill} \quad (1)$$

C_{start} is the magnitude of the difference between the start volume in the simulation scene and the start volume in the real-world scenario. $C_{received}$ is the magnitude of the difference between the received volume in the receiving container in the simulation and the desired real-world volume to be received. C_{spill} is the volume of spilled liquid in the simulation.

IV. EXPERIMENTAL SETUP

A. Cell Culture Robotics Setup

1) *Real-World*: In the planning and design phase of the system, key considerations were modularity, simplicity, and the integration of existing laboratory equipment. It includes a UR5e, a CytoSMART Lux3 BR microscope, an Intel RealSense D415, a VWR INCU-Line ILCO 180 Premium CO₂ incubator, and a thermoelectric chiller. Only consumables utilized in the manual cell culture process carried out by laboratory labor are used to enable seamless system integration. The setup is visualized in Fig. 4. All the devices are either connected to the UR control box or directly to the PC. Similar to the manual process, the autonomous process is split into three high-level workflows.

- Workflow A: Analyzing cell growth.
- Workflow B: Changing media.
- Workflow C: Passaging.

The system should be able to execute the three workflows independently, depending on the user input or the scheduled plan. To enhance reusability and simplicity, the steps to achieve completion of the workflows are structured into workflow modules. Pouring is required for the removal of spent media, adding and discarding of the washing solution, adding of fresh media, and splitting of the cell culture into three empty flasks. Since only a very low amount of dissociation reagent such as trypsin (approximately 0.5 mL

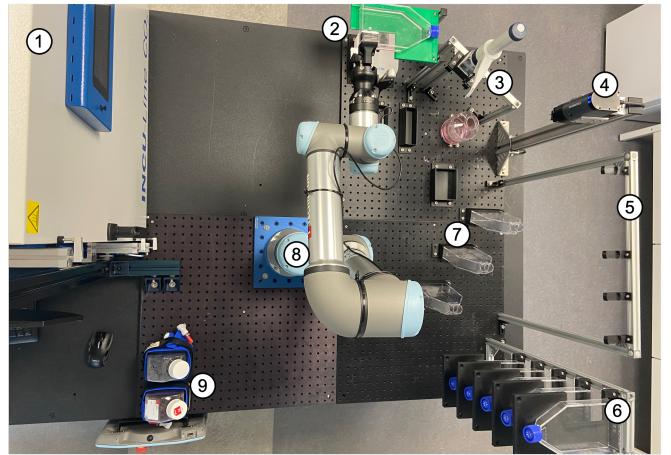


Fig. 4: Overview of the system prototype for cell culture automation. (1) Automated incubator. (2) Microscope. (3) Trypsin unit. (4) Capper/Decapper. (5) Lid holders. (6) Flask storage. (7) Flask holder for pouring. (8) UR5e with 3D-printed gripper fingers. (9) Heating and cooling of liquids unit.

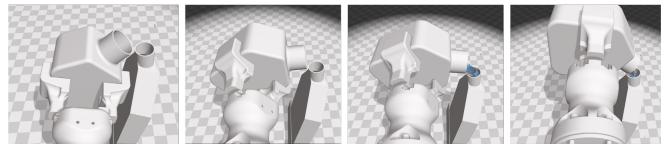


Fig. 5: An exemplary simulation scene of the media bottle executing the new pouring movement.

per 10 cm²) needs to be added to the cell culture flask, pouring is unsuitable. Instead, a bottle dispenser (1-10 mL) was attached to a trypsin bottle to allow vertical insertion of the reagent into the flask. All real-world pouring experiments were executed on that setup.

2) *Simulation*: Just like Kennedy et al. [15] and Lopez-Guevara et al. [19], we used the particle-based simulation library NVIDIA Flex for the simulation of pouring movements. Two scenes are created to simulate the two main liquid pouring movements required for cell culture automation. One for pouring from a cell culture flask, and one for pouring from a media bottle, both into another standing flask. The pouring movement is executed as explained in III-B. An exemplary simulation scene is shown in Fig. 5. In total, 6,805 pouring movements were simulated.

B. Volume Estimation of Liquids

1) *Vision Datasets*: In this work, we consider two recently published datasets containing transparent objects with liquid content, namely *TransProteus* [3] and *Vector-LabPics* [4]. Furthermore, a new dataset that will be called *LabLiquid-Volume* was created for this work.

The *TransProteus* [3] dataset consists of more than 50,000 synthetic images with annotated 3D models of transparent vessels and their content, including segmentation masks and material properties such as color, transparency, reflectance, and roughness. For this work, only transparent vessels with

liquid material are considered, which constitutes $\approx 14,500$ samples.

The *Vector-LabPics* [4] dataset comprises 7,900 real-world images with pixel-wise annotation of vessels and the containing material. The images capture materials in diverse stages and procedures observed in transparent vessels within chemistry labs, medical labs, hospitals, and other relevant settings.

The *LabLiquidVolume* dataset includes 5,451 images of liquids in laboratory containers. The images were taken using an Intel RealSense D415 camera in different environments in the automation laboratory and various research laboratories at the Novo Nordisk R&eD site in Måløv, Denmark. The ground truth of the liquid volume was measured using a Mettler Toledo XSR2002S balance with an accuracy of ± 0.5 mL. Twelve of the most common research laboratory containers, including the consumables used in cell culture processes, were selected. Various liquid volumes (3 - 600 mL), distances to the containers (50 - 600 mm), backgrounds, and camera angles were used. However, all images were taken from above so that the surface of the liquid is visible. The trained SDE model was used to generate the segmentation masks and depth maps automatically. Additionally, for each container, 10 images were manually annotated with a pixel-wise mask for vessel and liquid. Fig. 6 presents example data points of the dataset.

RGB Image	Depth Image (Camera Result)	Liquid Segmentation Map	Vessel Segmentation Map	Segmented Liquid Depth Map	Segmented Vessel Depth Map	Information
						Liquid Volume: 358 mL Vessel Volume: 755 mL Vessel Name: Gibco_500 Liquid Color: Red
						Liquid Volume: 111 mL Vessel Volume: 115 mL Vessel Name: Pyrex_100 Liquid Color: Red
						Liquid Volume: 26 mL Vessel Volume: 125 mL Vessel Name: Duran_100 Liquid Color: Red

Fig. 6: Visualization of the content of three samples of the generated *LabLiquidVolume* dataset.

2) *Model Training Settings*: The SDE network is trained for 75 epochs on the relevant subset of the *TransProteus* dataset and the *Vector-LabPics* dataset. Hereby, images from the *Vector-LabPics* are sampled with 33%, and the depth loss is set to zero, because no depth ground truth is available for those images. Image augmentation for training data includes resizing, centered cropping, Gaussian blurring, decoloring, and darkening. An Adam optimizer was used with an initial learning rate of 1×10^{-5} and a weight decay of 4×10^{-5} . The learning rate is adjusted every five epochs based on the average loss. If the average loss does not decrease for ten consecutive epochs, the learning rate is reduced by 10%. The batch size was set to 6. The loss for each segmentation mask is calculated using the standard per-pixel cross-entropy function. The loss for each depth map is calculated only for the region of the object as given by the specific ground truth mask. The scale-invariant depth loss, as introduced by Eigen et al. [24], is applied. The final model is evaluated on the

subset of the official *TransProteus* test set, which contains 1,500 samples for objects containing liquid.

The LVE network is trained on 80% of the *LabLiquidVolume* dataset. Hereby, the impact of different input information is investigated, as illustrated in Fig. 7.

- (1) Vessel depth map & liquid depth map
- (2) Vessel depth map & liquid depth map & vessel volume
- (3) Vessel segmentation map & liquid segmentation map

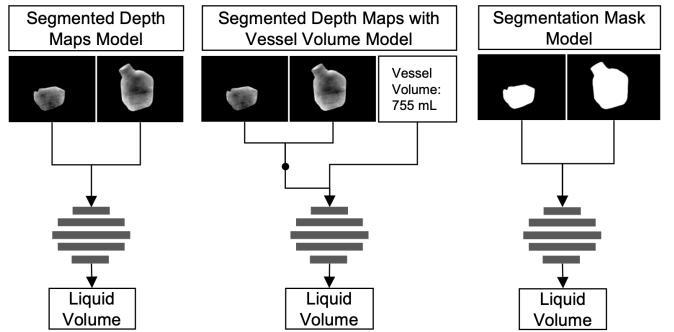


Fig. 7: Overview of the three input variants for liquid volume estimation network, trained on the *LabLiquidVolume* dataset.

For the second variant, the vessel volume is processed by a fully connected layer before being concatenated with the segmented depth maps. All variants are trained for 200 epochs with an Adam optimizer, a learning rate of 0.001, a batch size of 8, and a dropout rate of 0.2. For the loss function, we apply the Mean Squared Error (MSE).

3) *Evaluation Metrics*: For the evaluation of the SDE model only, we use the widely used Intersection over Union (*IoU*) to measure the segmentation performance and the *scale – invariant log RMSE* suggested by Eigen et al. [24] to measure the performance of the depth map prediction. For the evaluation of the full model, three different regression metrics are used: The root-mean-squared error *RMSE*, the mean absolute percentage error *MAPE*, and the *R*² score (coefficient of determination).

V. EXPERIMENTAL RESULTS

A. Vision Model Performance

For the SDE model, results are on par with the ones from Eppel et al. [3]. On the test subset, we achieve a mean *IoU* of 0.92 (Vessel: 0.96, Vessel Opening: 0.94, Liquid: 0.85) for the segmentation performance and a *scale – invariant log RMSE* of 0.031 for the depth estimation. The segmentation performance was additionally evaluated on the 120 manually annotated images of the *LabLiquidVolume* dataset. The performance on the real-world *LabLiquidVolume* images drops with a mean *IoU* of 0.90 (Liquid: 0.87, Vessel: 0.92). Furthermore, we can report a gap between glass and plastic objects for liquid segmentation. While the liquid region inside glass vessels is predicted with an *IoU* of 91.3%, it is significantly lower for liquids inside plastic objects with an *IoU* of 84.3%.

For the full model, we use the trained SDE model presented above and combine it with the three variants for the

LVE model as described in section IV-B.2. The results for the three variants are presented in Table I and indicate that the input combination of segmented depth maps and vessel volume model outperforms their competitors. Omitting depth information entirely and solely relying on liquid and vessel segmentation masks as input significantly degrades the model’s performance. However, in the further course of the

TABLE I: Performance of the three variants for the LVE model combined with the same SDE model.

Model Input	RMSE Training (mL)	RMSE Testing (mL)	MAPE Testing (%)	R^2 Testing
Segmented Depth Maps	13.00	37.90	16.96	0.94
Segmented Depth Maps & Vessel Volume Segmentation Masks	4.63	17.83	9.39	0.99
	32.94	53.92	24.21	0.88

experimental section, we will consider the LVE model that takes segmented depth maps only as input, because it allows the manipulation of vessels of unknown vessel volume and is, therefore, more flexible. Its distribution of real and predicted liquid volumes is shown in Fig. 8.

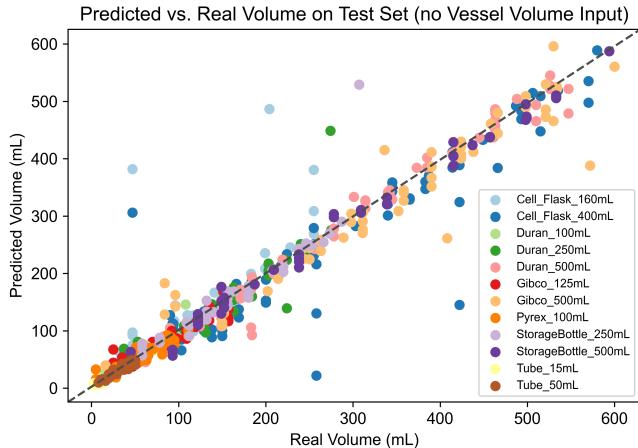


Fig. 8: Real and predicted liquid volumes for images from the *LabLiquidVolume* test set for the model trained only on the segmented depth maps.

We found a moderate negative correlation coefficient of -0.69 between the *IoU* for liquid segmentation and the *MAPE* of liquid volume estimation, indicating that better liquid segmentation performance was linked to more accurate volume predictions.

B. Pouring Performance

1) *Simulation:* We perform a cost analysis by creating sets of every possible target volume for every possible start volume between 0 mL and the maximum of the received volumes with a step size of 1 mL each. For the pouring with the cell culture flask, this results in 9,750 possible combinations of start and target volumes. Fig. 9 shows the minimum cost depending on start and target volumes. The

lowest cost among the combinations is 0.05 mL, and the highest is 16.5 mL. The mean cost for the set is 2.3 mL. For start volumes below 100 mL, the cost is smaller than 7.5 mL for every possible target volume, with a mean cost of 1.8 mL. The cost increases significantly for high start volumes with target volumes above 90 mL. For target volumes below 105 mL, a simulated pour can be found without any spilled liquid.

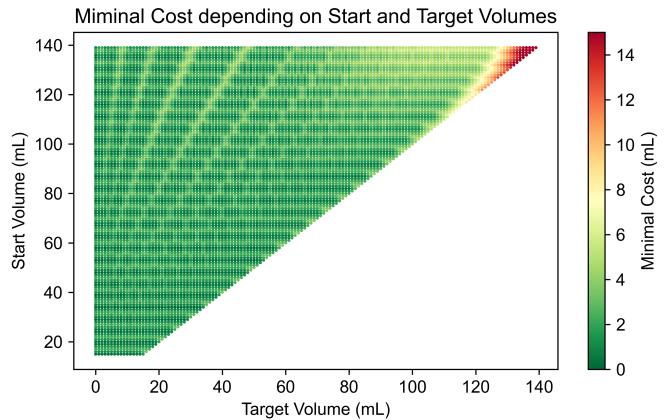


Fig. 9: Minimal costs for pours with a cell culture flask given a defined start and target volume with step sizes of 1 mL.

In Fig. 10, we show the spilled volume in relation to the received volume and stop angle. The spilled volume increases for higher stop angles and received volumes.

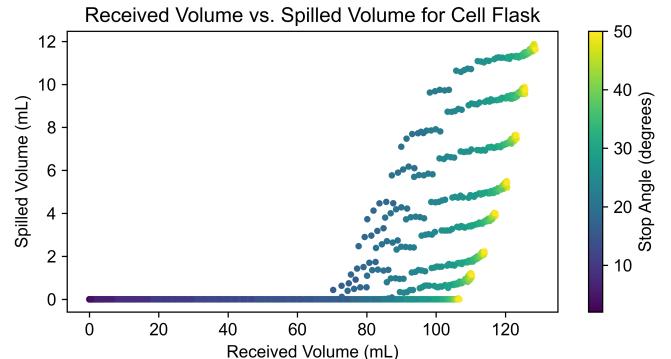


Fig. 10: Visualization of the spilled volume over the received volume for the simulated pours with the cell culture flask. The color encoding shows the stop angle of the simulated pour.

Additionally, it can be observed that the variation in the stop angle is responsible for the coarse differences in the poured volume, while the variation in the stop time results in more subtle differences. Similar results were found for the pouring simulation with the media bottle.

2) *Simulation-to-Reality:* The simulated movements were transferred to be executed on the system prototype using a UR5e as shown in Fig. 11. The results of 50 simulated pours executed in reality with the cell culture flask are visualized in Fig. 12. The *RMSE* is 10.8 mL (*MAPE* = 21%) with a maximum difference of 49.3 mL. The highest

differences occur for ratios of the target volume and start volume between 0.3 and 0.7. For the media bottle, a *RMSE* of 29.2 mL (*MAPE* = 52%) and a maximum difference of 52.6 mL was observed. For both containers, low spilling of on average less than 1 mL occurred during the real movements into a small opening.

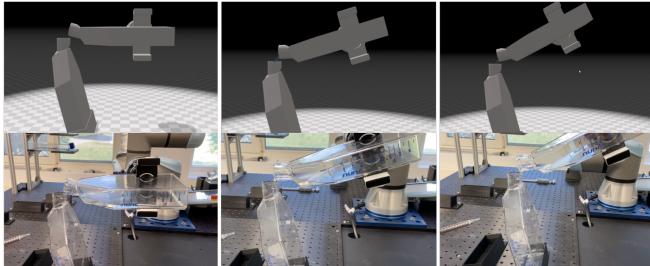


Fig. 11: A simulated pouring sequence with a cell culture flask (up) executed on the real-world robotic system (down).

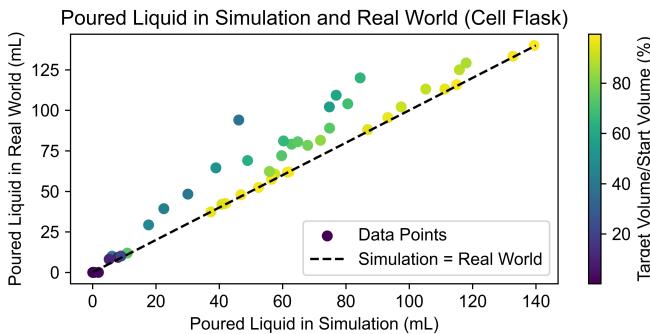


Fig. 12: Simulation-to-reality transfer results of the real experiments executed with a UR5e and a cell culture flask as pouring container.

The results of the spilled volumes in the real executions with the cell culture flask are visualized in Fig. 13. For 35 samples, the real spilled volume is lower than the predicted spilled volume from the simulation, with a maximum difference of 13.2 mL and a mean difference of 2.1 mL. For the remaining 15 pours, the real spilled volume is higher than the simulated one, with a maximum difference of 2.4 mL and a mean difference of 0.5 mL. It can be observed that the real spilled volume is close to the simulated spilled volume for low amounts of spilled liquid and low start volumes.

C. Overall Real-World Performance

The precision of the autonomous pouring approach, which includes the vision-based volume estimation of liquids, the selection of the pour with the minimum cost from the simulation, and the real execution of the pour including picking up and returning the pouring container, was tested. Random start volumes ranging from 100 mL to 500 mL were used. The target volumes were specifically chosen to align with the typical amount of media poured by human scientists, which commonly falls within the range of 30 mL to 50 mL. The results are listed in Table II.

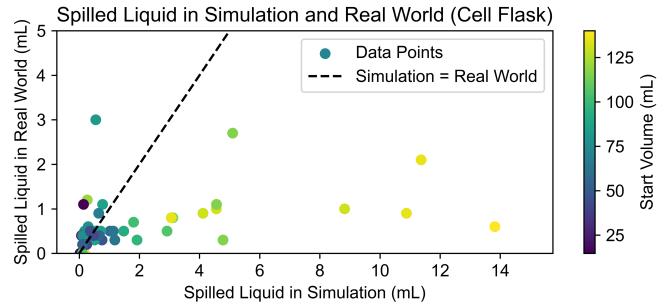


Fig. 13: Spilled volume in the real world compared to the spilled volume in simulation. Again, we show the results for the cell culture flask as a pouring container.

TABLE II: Results of the complete autonomous pouring workflow tested on a UR5e (20 times per target volume).

Target Vol. (mL)	Avg. Exec. Time (s)	RMSE (mL)	MAPE (%)	Max. Error (mL)
30	135.3	21.4	71.3	39.6
50	124.8	26.2	52.4	44.8

The three main cell culture workflows (analyzing cell growth, changing media, passaging) were tested ten times each. The results are shown in Table III. Start conditions were one filled cell culture flask in the flask storage inside the incubator, five empty flasks in the flask storage on the table, a media and a washing solution bottle with sufficient liquid, and a filled trypsin container. Exemplary executions can be seen here: https://github.com/DaniSchober/LabLiquidVision/tree/main/cell_culture_automation#--autonomous-workflows.

TABLE III: Results of the autonomous cell culture workflows executed on the proposed system (10 times each).

Workflow	Avg. Exec. Time (s)	Completion Rate (%)
Analyzing cell growth	78.3	100
Changing media	461.4	100
Passaging	841.2	90

VI. CONCLUSION

In this work, we used a UR5 robot to automate the task of cell culturing. Our approach is independent of additional automation-specific hardware as well as the location of the pouring and receiving containers, which is especially useful for a flexible robotic laboratory assistant. In addition, using a pouring approach instead of pipetting reduces the amount of disposable plastics and avoids integrating non-automation-friendly pipettes.

We introduced a new pouring approach, which is especially relevant for receiving containers with small openings, leading to very low spilling of less than 1 mL on average.

For the vision-based liquid volume estimation, we show that incorporating depth maps alongside segmentation masks

significantly enhances the liquid volume estimation model's performance, underscoring the benefits of utilizing depth information. We make the underlying dataset *LabLiquidVolume* publicly available for the research community.

Our experimental results reveal that our setup is not suitable for precise laboratory experiments since the volume estimation and pouring simulation errors add up. However, improving the accuracy of both elements in the future could enable flexible automation of many research laboratory tasks.

REFERENCES

- [1] A. Wolf and K. Széll, "A review on robotics in life science automation," in *14th International Symposium on Applied Informatics and Related Areas*, 11 2019.
- [2] P. Groth and J. Cox, "Indicators for the use of robotic labs in basic biomedical research: A literature analysis," *PeerJ*, vol. 2017, no. 11, 2017.
- [3] S. Eppel, H. Xu, Y. R. Wang, and A. Aspuru-Guzik, "Predicting 3D shapes, masks, and properties of materials inside transparent containers, using the TransProteus CGI dataset," *Digital Discovery*, vol. 1, no. 1, pp. 45–60, 2022.
- [4] S. Eppel, H. Xu, M. Bismuth, and A. Aspuru-Guzik, "Computer vision for recognition of materials and vessels in chemistry lab settings and the vector-labpics data set," *ACS Central Science*, vol. 6, no. 10, pp. 1743–1752, 2020, pMID: 33145411.
- [5] B. Burger, P. M. Maffettone, V. V. Gusev, C. M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R. S. Sprick, and A. I. Cooper, "A mobile robotic chemist," *Nature*, vol. 583, no. 7815, pp. 237–241, 7 2020.
- [6] S. Kleine-Wechelmann, K. Bastiaanse, M. Freundel, and C. Becker-Asano, "Designing the mobile robot Kevin for a life science laboratory," in *31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 4 2023, pp. 870–875.
- [7] N. Yoshikawa, A. Z. Li, K. Darvish, Y. Zhao, H. Xu, A. Kuramshin, A. Aspuru-Guzik, A. Garg, and F. Shkurti, "Chemistry Lab Automation via Constrained Task and Motion Planning," Tech. Rep., 12 2022. [Online]. Available: <http://arxiv.org/abs/2212.09672>
- [8] Labman Automation Ltd, "MultiDose: Automated XPR Solid Dosing System," 2023. [Online]. Available: <https://www.labmanautomation.com/multidose/>
- [9] S. Steiner, J. Wolf, S. Glatzel, A. Andreou, J. M. Granda, G. Keenan, T. Hinkley, G. Aragon-Camarasa, P. J. Kitson, D. Angelone, and L. Cronin, "Organic synthesis in a modular robotic system driven by a chemical programming language," *Science*, vol. 363, no. 6423, 1 2019.
- [10] R. Mottaghi, C. Schenck, D. Fox, and A. Farhadi, "See the Glass Half Full: Reasoning about Liquid Containers, their Volume and Content," *2017 IEEE International Conference on Computer Vision (ICCV)*, 1 2017.
- [11] F. Zhu, R. Jia, L. Yang, Y. Yan, Z. Wang, J. Pan, and W. Wang, "Visual-Tactile Sensing for Real-time Liquid Volume Estimation in Grasping," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2 2022, pp. 12 542–12 549. [Online]. Available: <http://arxiv.org/abs/2202.11503>
- [12] M. Cobo, I. Heredia, F. Aguilar, L. Lloret Iglesias, D. García, B. Bartolomé, M. V. Moreno-Arribas, S. Yuste, P. Pérez-Matute, and M. J. Motilva, "Artificial intelligence to estimate wine volume from single-view images," *Heliyon*, vol. 8, no. 9, 9 2022.
- [13] S. Eppel and T. Kachman, "Computer vision-based recognition of liquid surfaces and phase boundaries in transparent vessels, with emphasis on chemistry applications," *CoRR*, vol. abs/1404.7174, 2014. [Online]. Available: <http://arxiv.org/abs/1404.7174>
- [14] S. Eppel, "Tracing liquid level and material boundaries in transparent vessels using the graph cut computer vision approach," *CoRR*, vol. abs/1602.00177, 2016. [Online]. Available: <http://arxiv.org/abs/1602.00177>
- [15] M. Kennedy, K. Schmeckpeper, D. Thakur, C. Jiang, V. Kumar, and K. Daniilidis, "Autonomous Precision Pouring from Unknown Containers," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2317–2324, 7 2019.
- [16] C. Do and W. Burgard, "Accurate Pouring with an Autonomous Robot Using an RGB-D Camera," *International Conference on Intelligent Autonomous Systems*, vol. 15, pp. 210–221, 10 2018.
- [17] C. Schenck and D. Fox, "Visual closed-loop control for pouring liquids," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 2629–2636.
- [18] Y. Huang, J. Wilches, and Y. Sun, "Robot gaining accurate pouring skills through self-supervised learning and generalization," *Robotics and Autonomous Systems*, vol. 136, 2 2021.
- [19] T. Lopez-Guevara, N. K. Taylor, M. U. Gutmann, S. Ramamoorthy, and K. Suhr, "Adaptable pouring: Teaching robots not to spill using fast but approximate fluid simulation," in *Proceedings of the 1st Annual Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, S. Levine, V. Vanhoucke, and K. Goldberg, Eds., vol. 78. PMLR, 13–15 Nov 2017, pp. 77–86.
- [20] T. Lopez-Guevara, R. Pucci, N. K. Taylor, M. U. Gutmann, S. Ramamoorthy, and K. Suhr, "Stir to pour: Efficient calibration of liquid properties for pouring actions," in *IEEE International Conference on Intelligent Robots and Systems*. Institute of Electrical and Electronics Engineers Inc., 10 2020, pp. 5351–5357.
- [21] E. Babaian, T. Sharma, M. Karimi, S. Sharifzadeh, and E. Steinbach, "PourNet: Robust Robotic Pouring Through Curriculum and Curiosity-based Reinforcement Learning," in *IEEE International Conference on Intelligent Robots and Systems*, vol. 2022-October. Institute of Electrical and Electronics Engineers Inc., 2022, pp. 9332–9339.
- [22] Z. Pan, C. Park, and D. Manocha, "Robot Motion Planning for Pouring Liquids," in *Proceedings of the Twenty-Sixth International Conference on Automated Planning and Scheduling*, 2016, pp. 518–526.
- [23] A. Graikos, V. Charisis, D. Iakovakis, S. Hadjidimitriou, and L. Hadjileontiadis, "Single image-based food volume estimation using monocular depth-prediction networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12189 LNCS. Springer, 2020, pp. 532–543.
- [24] D. Eigen, C. Puhrsch, and R. Fergus, "Depth Map Prediction from a Single Image using a Multi-Scale Deep Network," in *NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems*, 6 2014, pp. 2366–2374.