2025 IEEE/SICE International Symposium on System Integration (SII)
January 21-24, 2025, Munich, Germany

Empowering Lab Education: Integrating a Vision-Based Monitoring System with Small-Scale Self-Driving Experiment Platforms

Junru Ren[1], Nis Fisker-Bødker[2], Ronja Güldenring[1],
Jin Hyun Chang[2], Tejs Vegge[2], Ole Ravn[1], Lazaros Nalpantidis[1]

***Abstract*— The rise of self-driving laboratories has seen significant growth across various research domains, particularly in chemistry, materials science and life science. However, a major challenge persists—the majority of self-driving systems are costly due to the use of highly precise lab equipment, robotic platforms, and case-specific algorithms, rendering these systems less accessible for educational purposes. This paper takes a multidisciplinary approach; we first introduce a small-scale self-driving experiment platform tailored for educational use, focusing on liquid materials mixing tasks commonly seen in chemistry and life sciences. To understand the operational status in real-time while maintaining self-driving capability and efficiency, we propose a novel system concept: employing a mobile robot as the lab supervisor to monitor the experiment process across multiple identical self-driving platforms. Specifically, this paper focuses on implementing a vision-based monitoring system. A deep learning architecture with a new training strategy is presented to jointly address two tasks: (a) vessel and content material segmentation and (b) volume estimation. The two tasks can be trained independently but can be inferred end-to-end by integrating them into the Mask R-CNN framework. Through evaluating the monitoring module on a real dataset, the results showcase promising detection capabilities, good real-time performance, and compatibility with the self-driving platform, indicating the feasibility of our proposed system.**

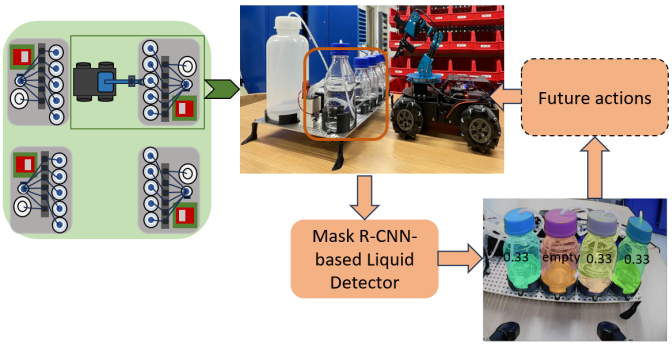


Fig. 1: Illustration of the proposed concept targeting liquid materials discovery experiments for educational purposes. First, several small-scale self-driving materials discovery platforms have been designed and built. To monitor the status of various experiments, a small vision-based mobile robot is deployed to navigate around and check the self-driving platforms, informing the students about the status of the experiments and potentially initiating subsequent manipulation actions.

I. Introduction

Research activities in chemistry, materials science and life science often involve extensive laboratory work, which demands continuous physical presence, considerable time and expertise from researchers. To address this challenge, materials acceleration platforms (MAPs) or self-driving labs (SDLs) have gained prominence, often jointly referred to as 'lab automation' [17, 3, 1, 2]. SDLs require multidisciplinary skill sets, as they integrate robotic platforms, advanced lab equipment, and AI algorithms to execute experimental laboratory tasks automatically and, in some cases, autonomously. However, such platforms are prohibitively expensive to build and require consistent and meticulous maintenance. Consequently, access is typically restricted to professional staff, limiting broader applicability.

To alleviate the aforementioned limitations and make SDLs part of lab education, we present a concept of implementing several stationary self-driving platforms in a lab for university students conducting class-level experiments, with supervision from a vision-enabled mobile robot. To be more specific, in our case, a self-driving platform is built for accelerated materials discovery, where a certain percentage of raw materials stored in separate containers is pumped into a test cell, and the characteristics of the mix are measured and compared against the target goal. Simultaneously, a mobile robot is implemented to mimic a human operator monitoring the experiments. The robot navigates around the platform(s), locates containers, and estimates the level of materials inside. If the estimation is in disagreement with lab expectation, the robot will take action, such as replacing the empty vessel with a new one. The illustration of our concept is shown in Figure 1.

The design of the self-driving platform is modular with all liquid containers aligned on one side, as shown in 2a. This arrangement facilitates easy monitoring and replacement of the equipment. Having an on-site monitoring system can assist in maintaining constant and safe operation. A Mask R-CNN-based architecture is proposed to mainly perform two vision monitoring tasks: vessel-liquid segmentation and

The authors acknowledge the Digital PhD Program at the Technical University of Denmark (DTU), and the Digital Hub for Energy Materials at DTU Energy for support. The authors would like to thank students Louie Lucas Bisgaard Nyeland, Rógvi Ziskason, and Kshitij Gambhir for helping with building the self-driving platform.

[1]Department of Electrical and Photonics Engineering, Automation and Control Group, Technical University of Denmark {junre, ronjag, oravn, lanalpa}@dtu.dk

[2]Department of Energy Conversion and Storage, Technical University of Denmark {nisfi, teve, jchang}@dtu.dk

liquid-level detection. The architecture is designed to be trained separately on different datasets for the two tasks. The training results will be combined through end-to-end inference to solve more complex jobs. The whole architecture is trained on public datasets only, avoiding the time-consuming dataset collection process. We use the Labpics[6] dataset to train the model for vessel-liquid instance segmentation and the COQE[13] dataset for the liquid-level detection task. To show our approach's feasibility for our considered application we tested the trained model on a custom dataset. It is worth noting that real-time performance is highly demanded in the monitoring system, leading us to experiment with various backbones to enable this two-stage model to work in real-time. In the end, the segmentation and level detection results show the ability to support future tasks such as container manipulation or liquid replenishment.

The main contributions of our work can be summarized as follows:

- We present a novel self-driving liquid handling platform tailored for educational purposes, providing a cost-effective and easily reproducible solution.
- We develop a deep learning architecture capable of monitoring experiment status and performing multiple tasks, i.e. segmentation and liquid volume detection. Compared to existing works, our model demonstrates good accuracy for real-time data processing.
- We specifically design our network architecture to showcase how different datasets can be fused to solve more complex jobs, avoiding collecting a new large custom dataset.

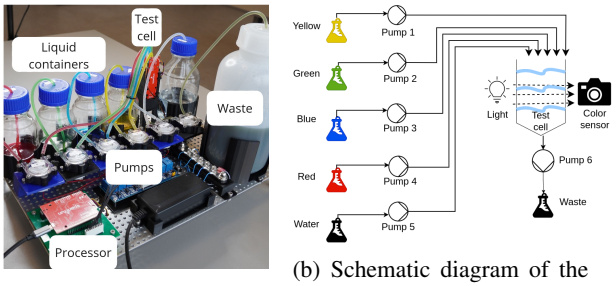
To clarify the objective of this paper, our primary focus is on the perception task, which is to demonstrate the feasibility of integrating a vision-based monitoring module with the self-driving liquid handling platform, enabling the robot to understand the status of each liquid container, provide feedback to students, and suggest future actions. The self-driving platform's software and robot manipulation algorithm are not within this paper's scope.

II. System Description and Application Case

A. Small-scale Self-driving Platform

The platform aims to provide students with the opportunity to experiment with machine learning algorithms and optimize towards a predetermined color mixture given by the instructor. Instead of employing robotic manipulators for liquid handling, the platform utilizes peristaltic pumps to transfer fluids from reservoirs to the test cell. This ensures simplicity and allows students without prior robotic knowledge to use the platform. Control of the pumps is facilitated by an Arduino microcontroller and an 8-channel relay.

The platform features five reservoirs, which are filled with water for flushing, as well as water mixed with red, green, yellow, and blue colorings. The pumps then blend the desired combination of colors within the test cell. This cell consists of acrylic transparent flat surfaces, white LEDs to illuminate the cell and a color sensor that detects the current color of



(a) Self-driving platform

(b) Schematic diagram of the color-mixing experiment

Fig. 2: Pictures and diagram of the self-driving materials discovery platform designed in our project. The bidirectional pumps are used for liquid handling, and a sensor is mounted on the test cell to detect the experiment result. The containers are aligned on one side of the platform which enables easy monitoring by the robot.

the cell. After color sensing, the cell is drained using a pump and flushed 2 times with water before it is ready for the next run. Waste is collected in a bottle. The peristaltic pumps are calibrated with weight versus runtime to ensure reliable and similar dispensing across the pumps.

The RGB color from the color sensor is then fed back to the machine learning algorithm to optimize toward the predetermined color given by the instructor.

A schematic illustration of conducting the color-mixing experiment on the platform is shown in Figure 2b. The demonstration of the working process is shown in the supplement video.

B. Mobile Robot Supervisor

Implementing a vision-equipped mobile manipulator as the lab supervisor, aims to enhance the cost efficiency and operation of the SDL. Consider a scenario where a mobile robot serves 4 stationary platforms in operation. It drives toward each platform to identify containers that need filling or emptying. If the liquid level is too low, the manipulator will execute further action, such as replacing the containers. To execute an entire supervision workflow, a mobile manipulator is anticipated to fulfill various tasks, such as: (i) Navigating to the selected platform, (ii) Adjusting robot arm pose to get ready for taking pictures, (iii) Perceiving working status, (iv) Manipulating containers if needed.

We have selected a small mobile robot manipulator from Hiwonder, which operates on the Robot Operating System (ROS)[15] and provides open-source software. The robot carries an RGB camera on the last link of the robotic arm. The camera has 480p resolution and a 180^o-wide field of view. Students working on the SDL will be given access to set the robot's movements, controlling the robot to navigate to the workstations regularly for status checks and manipulation if needed.

III. Related Work

Lab automation is a popular topic being extensively researched, e.g., in biology, chemistry, materials science and

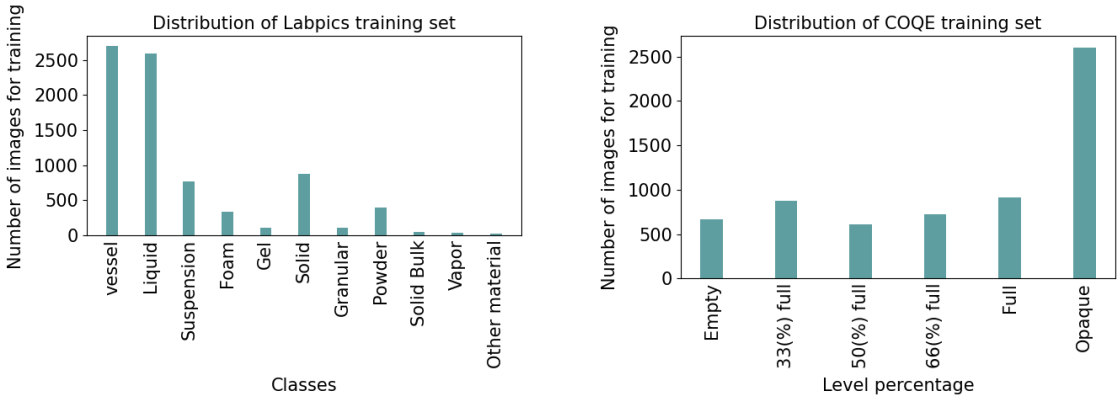


Fig. 3: Distribution of two training datasets. In Labpics, the amount of data for liquid and vessel classes is even. In COQE, the data for each measurable class is balanced.

food science. Most works aim to build an automatic experiment platform for high-level research [2, 17, 3, 1, 18], while [12] provides a review on existing works for low-cost SDLs, which are called 'frugal twins', that can be implemented for research and education purposes. Similar to the works discussed in [12], the self-driving platform proposed by us is for liquid mixing purposes and has been modularized to facilitate easy and real-time maintenance in the laboratory environment. Among recent works, mobile robots are mainly used as couriers without monitoring purposes. To the authors' knowledge, our system is the first to integrate multiple small-scale platforms and a mobile robot supervisor as a cost-effective educational experiment platform.

The vision-based perception system enables the lab operators, both humans and robots, to understand the situation and plan for the following actions[20]. Many experiments in chemistry and materials science fields are conducted using transparent glassware equipment, with the materials in a liquid phase. Therefore, building a computer vision system capable of detecting transparent objects is essential. Many works on transparent object detection have collected comprehensive datasets in both 3D simulators and the physical world, and models for detection and segmentation are developed[21, 6]. Regarding liquid detection and liquid-level estimation, some works used traditional computer vision methods [22], such as background subtraction and calculating the height of the detected bounding box. Those methods have constraints on specific environment setups, for example, high contrast between the background and the foreground, and the bounding box height calculation method relies on box detection accuracy. However, works like [14, 13, 4] used neural network models to predict the liquid volume and level, which inspired our work. In our work, we are interested in detecting the liquid level as a percentage of the vessel because the specific volume value detection is easily distracted by the view distance, as well as the capacity and shape of the container.

We aim to solve two tasks: vessel-liquid instance segmentation and liquid-level detection. We choose Mask R-CNN [7] as the primary framework of the monitoring system. The framework is still a popular tool used in instance segmentation and shows good competitiveness in many works. Transparent vessels and liquids are difficult to detect compared to solid and opaque objects due to reflections and weak edges[9], which means it is essential to extract image features and select the region of interest (ROI). Therefore, Mask R-CNN is selected due to its specialized Region Proposal Network (RPN), ROI align module and strong performance in instance segmentation tasks. Compared to state-of-the-art one-stage detectors, Mask R-CNN is worse on inference time but exhibits high predicting precision[23]. As the system would be applied within a lab scenario, we prioritize lightweight backbones for the Mask R-CNN that are sufficient to handle image features and reduce inference time.

Regarding liquid level detection, [4] demonstrated that the level prediction result would be improved when introducing the bounding box information of each vessel during training. Inspired by this, we would like to involve the bounding box as input in both training and inference, where the Mask R-CNN will provide the bounding box information during the inference process. Thus, the system can be implemented as an ensemble of two tasks: instance segmentation on vessels and liquid and liquid-level prediction.

IV. Datasets

Numerous novel datasets for transparent object and liquid detection have been proposed recently, captured both in the real world and in a synthetic way. An interesting challenge is how to efficiently utilize existing public datasets and tailor them to our specific needs instead of relying on building custom training sets.

Targeting vessel-liquid segmentation and liquid level estimation tasks, we focus on two datasets: Labpics-v1[6] and COQE[13], briefly summarized in Figure 3. Labpics offers classes of vessels, liquids, solids and other chemistry materials annotated with pixel-level masks. It is closely aligned with our use case, providing scenarios in chemistry and medical labs, making it an ideal choice for training and testing the model. Additionally, the various material

types make Labpics a valuable resource for expanding our architecture on non-liquid detection in the future. Even though COQE contains images from daily scenes—not from labs—it uniquely labels the liquid level in discrete percentage levels: empty, 33%, 50%, 66%, and full, which converts liquid level estimation to a classification task. This is similar to our task, making COQE an appealing option for our work.

For model training, we selected the classes 'vessel' and 'liquid' from Labpics to train the segmentation task. The dataset contains balanced number of data for these two classes. Simultaneously, we trained the liquid level detector using the six classes from COQE, to classify liquid volume into categories.

In order to test the model performance, we collected 50 images using the RGB camera mounted on the mobile robot as the custom test dataset. The dataset contains scenarios of multiple vessels with colored and uncolored transparent liquids inside, and images were captured from random angles. Only the bounding box and liquid level were labeled on the custom test dataset to evaluate the performance of the liquid level detection network and the performance of the entire architecture.

V. Network Architecture

There are two tasks for the monitoring system: vessel-liquid segmentation and liquid-level detection. We aim to propose a method that on the one hand allows the two tasks to be trained separately on two public datasets, and on the other hand, allows inference end-to-end by integrating a level detector attaching to Mask R-CNN. The entire architecture is shown in Figure 4. Mask R-CNN first takes the RGB image as input and outputs the categories, bounding boxes and masks of the detected objects. As a part of the system output, the segmentation masks can be used for future tasks such as object localization and color detection. The detected bounding boxes, together with the original RGB image, crop the input as a single-vessel-focused RGB image. Then, the cropped image is sent to the level detector for the level detection task. The integrated network architecture enables the system to predict both the object type (vessel or liquid) and object property (liquid level) from one detection, and it also enables the training of two tasks independently on two datasets. This refrains dataset collection and relabeling when a new project starts and boosts data efficiency.

Mask R-CNN, as a two-stage object detector, is well-known for high precision, but its main disadvantage is the long inference time compared to one-stage detectors. Therefore, the selection of a backbone, which has sufficient capacity for image features and is simple enough for faster detection, is crucial. In our case, the backbone is responsible for extracting features from the input image and works with the feature pyramid network (FPN)[11] to enrich the extracted features. The proposed regions of interest (ROI) aligned with the features are sent to heads for prediction, respectively.

By using the COQE dataset, the level detection is seen as a classification task, where the volume is divided into 5 classes: 'empty', '33% full', '50% full', '66% full' and 'full'. The liquid level is predicted based on the proportion of the liquid inside the vessel instead of a specific value. During training, the image is first cropped according to the ground truth vessel bounding box, while in inference, the image is cropped based on the bounding box output from the Mask R-CNN. The cropped result enables level classification on every single vessel, helping improve detection performance.

VI. Experiments

In this section, the experimental results of the proposed monitoring methods are presented to demonstrate our system's applicability. Based on the architecture introduced in the former section, several models were trained on the public datasets and tested on the custom dataset to select a proper backbone and liquid-level detector.

A. Proposed architecture evaluation

We trained the Mask R-CNN with various backbones using the Labpics dataset for the segmentation task. Only the 'vessel' and 'liquid' classes were used during training, which was sufficient in our scenario. We aimed to find a lightweight backbone that can both keep the precision and guarantee real-time performance. All networks were trained on a single Nvidia V100 GPU and tested on Nvidia RTX 3090.

The loss function was selected from the original Mask R-CNN model with no modification. The Mask R-CNN treats classification, bounding box detection, and mask segmentation as three parallel tasks, from which the loss function is the sum of these tasks as $L = L_{cls} + L_{box} + L_{mask}$. The hyper-parameters used during training were specified as batch size, optimizer, and learning rate. We set the batch size as 8 due to the limited memory of our GPU and used the adaptive moment estimation (Adam) optimizer[10] with a constant learning rate of 0.0001 for simplicity. Considering the properties of transparent objects, specific data augmentation methods were applied during training: blurring, safe crop, and color jitter with $P = 0.5$. The evaluation result of the model with various backbones is shown in Table I. The result illustrates that the mentioned data augmentation pipeline can help improve the model performance in general, and that the ResNets provide slightly better precision but longer inference time. In the end, we use MobileNet-v2[16] as the backbone due to its low inference time and considerable performance.

We trained the liquid level detector on the COQE dataset. In line with the practice of [13], the level detector was trained with a 4-layer input: a 3-layer RGB image and a bounding box binary mask, and then inference by only RGB image. However, in our work, we modified the training and the inference process by using the cropped RGB image as input. The input was cropped by the ground truth during training, while it was cropped by the Mask-RCNN predictions during inference. Similar to the segmentation task, we aimed to select a fast and precise liquid-level detector. Several popular models were trained and compared with both original and

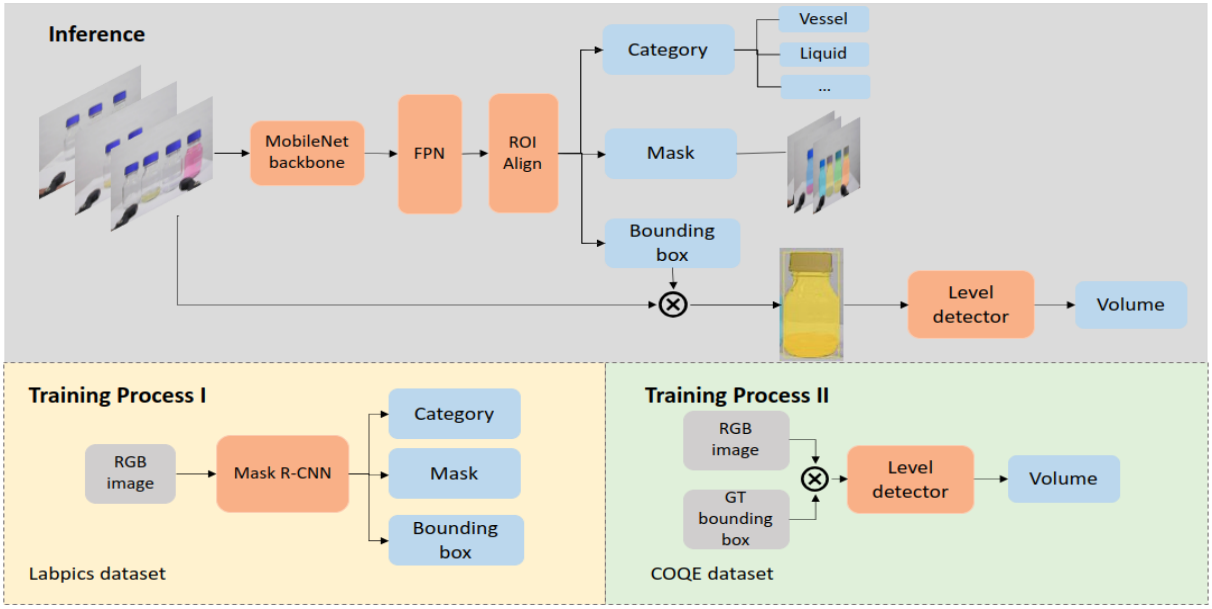


Fig. 4: Proposed network architecture. The figure illustrates the processes of inference and training. The training process I trains the Mask R-CNN to predict the mask and bounding box of the vessel and liquid instances. The training process II teaches the level detector to estimate the liquid volume, where the input is the RGB image cropped by the annotated vessel bounding boxes. During inference, the entire architecture takes the RGB image as the input and sends it to the Mask R-CNN first. The bounding box for each vessel instance, predicted by Mask R-CNN, is used to crop the RGB input and send it to the Level detector. In the end, the predicted categories, masks, and liquid levels are outputed as the result of the monitoring system.

cropped input. The loss function used was cross entropy $L(\hat{y}, y) = -\sum_k^K y^{(k)} \log \hat{y}^{(k)}$, where k represents the k_{th} class. The detectors were trained with the Adam optimizer, constant learning rate at 0.00001, and batch size 32. Data augmentation was also applied. Several popular networks were trained, where ResNet[8], MobileNet[16] and EfficientNet[19] had input dimensions of 64×64, whereas the ViT [5] 224×224 pixels.

TABLE I: Performance of Mask R-CNN with different backbones. All models used data augmentation unless otherwise stated. Models were trained and tested on the Labpics dataset. Result shows that data augmentation improves model performance. ResNet-50 achieves the highest average precision (AP), and MobileNet has the fastest inference time.

Backbones	AP	AP50	AP75	AR	#params	Inf. (msec)
ResNet-50 w/o data aug.	48.7	65.1	50.6	52.9	43M	33.8
ResNet-50	**51.5**	**67.7**	**55.6**	**56.1**	43M	42.8
ResNet-18	48.2	64.6	51.1	52.2	**30M**	35.2
MobileNet-v2	48.1	65.3	50.9	51.9	87M	**21.2**

TABLE II: Performance of different methods for the liquid-level detection task, by testing on COQE dataset. The best performance is marked in bold. Each method was trained and tested on both non-cropped images (METHOD w/o) and cropped images (METHOD w/). Mean Average Precision (mAP), mean Average Recall (mAR), F1 score (F1), and inference time are listed. ViT-b gives the highest Precision and Recall with acceptable inference time.

Methods	mAP	mAR	F1	Inf. [msec]
ResNet-18 w/o	20.3	19.3	0.198	3.8
ResNet-18 w/	36.2	35.6	0.359	**3.7**
MobileNet-v3 w/o	20.6	19.1	0.196	6.1
MobileNet-v3 w/	29.0	28.4	0.287	6.3
EfficientNet w/o	21.4	20.7	0.210	17.4
EfficientNet w/	38.6	37.4	0.380	16.8
ViT-b w/o	27.3	25.5	0.264	7.6
ViT-b w/	**46.7**	**46.6**	**0.466**	7.7

The evaluation results are shown in Table II, which indicates that the cropped input can significantly improve model performance. We select ViT-b as the liquid level detector because of its overwhelming performance among other tested models—the AP of ViT using cropped input is 40% higher than using non-cropped input, and 20% higher than other classifiers. The inference time of ViT-b is 7.7 msec which is still acceptable. We noticed that during training, the ViT-b got over-fitting after 10 epochs, while other networks were trained for 50 epochs. This indicates the good learning ability

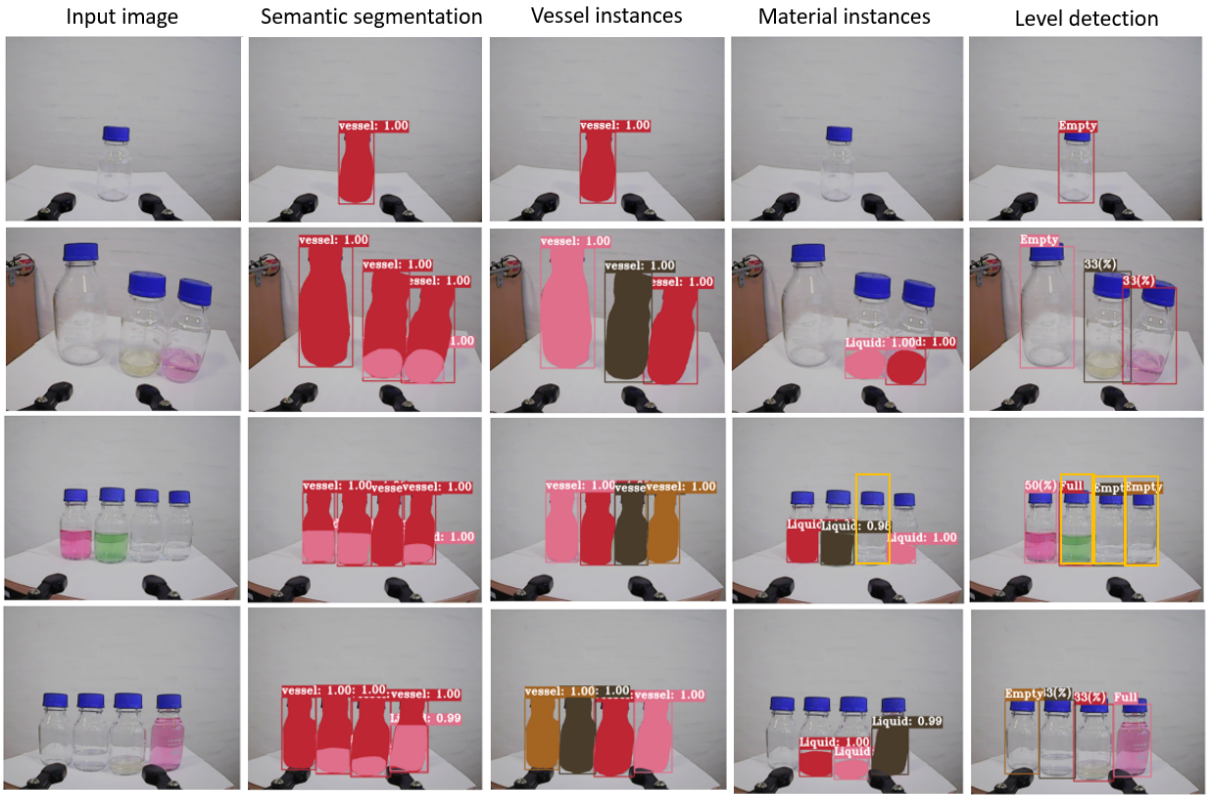


Fig. 5: Segmentation and level detection results from the proposed architecture. MobileNet-V2 was used as a backbone, and ViT-b was implemented as the level detector. The network can predict single and multiple vessels within various colored liquids inside. Yellow boxes indicate wrong or missing predictions. Both segmentation and level estimation face difficulties with non-colored transparent liquids.

of the vision transformer and its data-hungry characteristics.

TABLE III: Mean Average Precision and mean Average Recall for the liquid-level detection task tested on the custom dataset using Mask R-CNN to detect vessels and crop inputs for the level detector.

Method	mAP	mAR	F1
Level detection	39.7	38.8	0.393
Level detection w/ ‘Empty’, ‘half full’, ‘full’	66.0	38.7	0.487

Next, the entire proposed architecture was tested on the custom dataset. The results show that the Mask R-CNN can provide useful predictions to crop the image as input of the level detector. Some indicative examples are shown in Figure 5. The results illustrate that in most cases, the network can predict the vessels, liquid, and liquid levels well, but there is a tricky case: non-colored transparent liquids are often missed. However, the prediction of the liquid segments and the liquid level are independent, which gives us the idea of using the liquid mask to calculate the level in post-processing as a supplement.

TABLE IV: Prediction results for vessel instance segmentation, liquid volume (percentage) estimation based on segmentation, liquid level detection by proposed architecture, and color identification, based on an example of real use case from Figure 6. The containers are ordered from left to right.

Vessel No.	Vessel segment -ation	Estimated volume	Level detection	Color [R,G,B]
1	Yes	0.3962	50%	[80, 99, 125]
2	Yes	0.3166 (failed)	66%	[119 ,94, 10]
3	Yes	0.5260	50%	[32, 24, 51]
4	Yes	0.5491	50%	[17, 46, 28]
5	Yes	0.4901	50%	[40, 47, 66](failed)

Table III lists values of mean average precision and mean average recall of the entire architecture when combining Mask R-CNN with liquid level detector. The model is evaluated on our custom dataset. Model performance is slightly worse than that what demonstrates in II, due to the scene in COQE being different and the model using segmentation results as input to level detection instead ground truth. Specifically, the values of average precision of classes ‘empty’, ‘33% full’ and ‘full’ are higher than those of classes ‘50% full’ and ‘66% full’, which indicates that the model has

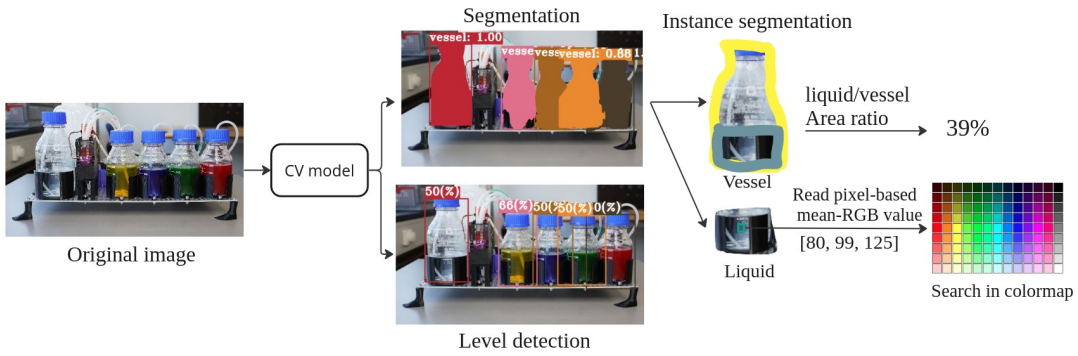


Fig. 6: An example to demonstrate the real use case: implementing the vision-based monitoring system on the self-driving platform. The picture shows that the proposed model can predict vessel instances and liquid levels at the same time. In addition, two applications after segmentation are implemented. One is to estimate the liquid level by comparing the area of the vessel mask and the corresponding liquid mask. Another is to identify the liquid color inside the containers.

difficulty to distinguish those two close cases and this needs to be addressed in future work. Meanwhile, we modified the level detection in 3 categories: 'empty', '50%full', and 'full' which as expected highly increased the prediction precision.

B. Real use case demonstration

In this section, an example is considered to demonstrate the applicability of the proposed system in a real use case. We combined the monitoring module with the self-driving platform and captured the scene shown in Figure 6. The result indicates that, similar to testing on previous custom data, the computer vision model is compatible with the self-driving platform, which can segment individual vessels and liquid instances as well as detect the liquid level in each vessel simultaneously. Additionally, two further applications are considered in the segmentation branch. The first is to use the liquid-vessel area percentage to estimate liquid volume, which can provide a specific volume percentage. The liquid volume is computed by $(A_{MaskL} \times A_{MaskV})/A_{MaskV}$, where A_{MaskV} is the area of the individual vessel mask and the numerator is the area of liquid mask belonging to the corresponding vessel. However, this method relies on the precision of the previous segmentation step, so we only consider it as a complement to the model-based level detection branch. The second application is liquid color identification. Instead of relying on a predefined color for each container, we provide an additional color check mechanism. This is done by extracting the central pixels of the liquid segments, reading the median RGB values, and searching in the color map, which is a useful function in material identification. Table IV shows the prediction of 5 containers shown in Figure 6. The containers are ordered from left to right. The results of the estimated liquid volume percentage from segmentation, liquid level detection by the proposed architecture, and color identification results are listed in the table. Note that the color identification result of container 5, which contains red liquid, failed due to the black container holder partially occluding the liquid. In the future, we plan to modify the configuration of the holder, making it thinner and shorter to enhance vision detection. We also noticed that the color detection by the camera is sensitive to the background lighting condition. However, our objective is not to estimate the absolute RGB values of the liquid, but rather to identify the relative color components that indicate the material held within the container which is accessible in our work.

VII. Conclusion and Future Work

Lab automation is a popular topic in chemistry, materials science and life science research, where many materials acceleration platforms and self-driving labs have gained prominence. To make SDLs part of lab education, we introduced a small-scale self-driving platform for university students, which can conduct class-level materials discovery experiments. To monitor the status of individual vessels on the self-driving platform, a Mask R-CNN-based computer vision module was built for simultaneous vessel-liquid instance segmentation and liquid level detection. The model architecture is designed to be easily trained on different datasets for different tasks, which enlarges the utilization of public datasets and saves efforts on collecting huge data on the specific project. Our experiments demonstrate the precision and real-time performance of the proposed architecture, as well as its compatibility with the liquid handling platform. Both the monitoring system and self-driving platform support our proposed SDL concept: to implement a mobile robot as the lab supervisor to monitor the experiment status for several platforms in a room, and manipulate lab equipment. The demonstration of the working platform and the computer vision system is in the attached video.

During experiments, we noticed that the model had difficulties with transparent liquids. Therefore we introduced the use of segmentation results to estimate liquid level as a complement of the model-based liquid level detection. In the future, we will continue improving the configuration of the self-driving platform to make it more friendly to the computer vision system and the robot manipulator. It is worth noting that the holders used to fix the vessels on the platform, seen as black plastic modules may mislead the

model detection and post-processing in some cases, which will also be modified in future work. The project underscores the significance of co-creation among researchers from diverse disciplines. Robotics experts are tasked with designing an autonomous system tailored to meet the requirements of chemical experiments, while chemists must ensure that the experiment platform configuration is conducive to the operation of the robotics system. This ensures the successful integration of technology and scientific research. Robot manipulation will also be implemented based on the detection from the monitoring system. For example, if the liquid level falls below the acceptable range, the robot should replace the container with a new one. Conversely, if the liquid level increases in the raw material container, the robot should issue a warning to human operators to inspect the operation. We believe that the combination of the mobile manipulator and small-scale self-driving platforms will enable the realization of fully autonomous educational SDLs in the future.

REFERENCES

[1] Milad Abolhasani and Eugenia Kumacheva. "The rise of self-driving labs in chemical and materials sciences". In: *Nature Synthesis* 2.6 (2023), pp. 483–492.

[2] Seoin Back et al. "Accelerated chemical science with AI". In: *Digital Discovery* 3.1 (2024), pp. 23–33.

[3] Benjamin Burger et al. "A mobile robotic chemist". In: *Nature* 583.7815 (2020), pp. 237–241.

[4] Miriam Cobo et al. "Artificial intelligence to estimate wine volume from single-view images". In: *Heliyon* 8.9 (2022).

[5] Alexey Dosovitskiy. "An image is worth 16x16 words: Transformers for image recognition at scale". In: *arXiv preprint arXiv:2010.11929* (2020).

[6] Sagi Eppel, Haoping Xu, Mor Bismuth, and Alan Aspuru-Guzik. "Computer vision for recognition of materials and vessels in chemistry lab settings and the vector-labpics data set". In: *ACS central science* 6.10 (2020), pp. 1743–1752.

[7] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "Mask r-cnn". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2961–2969.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[9] Jiaqi Jiang, Guanqun Cao, Jiankang Deng, Thanh-Toan Do, and Shan Luo. "Robotic perception of transparent objects: A review". In: *IEEE Transactions on Artificial Intelligence* (2023).

[10] Diederik P Kingma. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[11] Tsung-Yi Lin et al. "Feature pyramid networks for object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 2117–2125.

[12] Stanley Lo et al. "Review of Low-cost Self-driving Laboratories: The" Frugal Twin" Concept". In: (2023).

[13] Roozbeh Mottaghi, Connor Schenck, Dieter Fox, and Ali Farhadi. "See the glass half full: Reasoning about liquid containers, their volume and content". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 1871–1880.

[14] Gautham Narasimhan, Kai Zhang, Ben Eisner, Xingyu Lin, and David Held. "Self-supervised transparent liquid segmentation for robotic pouring". In: *2022 International Conference on Robotics and Automation (ICRA)*. IEEE. 2022, pp. 4555–4561.

[15] Morgan Quigley et al. "ROS: an open-source Robot Operating System". In: *ICRA workshop on open source software*. Vol. 3. 3.2. Kobe, Japan. 2009, p. 5.

[16] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. "Mobilenetv2: Inverted residuals and linear bottlenecks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 4510–4520.

[17] Martin Seifrid et al. "Autonomous chemical experiments: Challenges and perspectives on establishing a self-driving lab". In: *Accounts of Chemical Research* 55.17 (2022), pp. 2454–2466.

[18] Sebastian Steiner et al. "Organic synthesis in a modular robotic system driven by a chemical programming language". In: *Science* 363.6423 (2019), eaav2211.

[19] Mingxing Tan and Quoc Le. "Efficientnetv2: Smaller models and faster training". In: *International conference on machine learning*. PMLR. 2021, pp. 10096–10106.

[20] Yi Ru Wang et al. "Mvtrans: Multi-view perception of transparent objects". In: *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2023, pp. 3771–3778.

[21] Enze Xie et al. "Segmenting transparent objects in the wild". In: *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*. Springer. 2020, pp. 696–711.

[22] Tara Zepel, Veronica Lai, Lars PE Yunker, and Jason E Hein. "Automated liquid-level monitoring and control using computer vision". In: (2020).

[23] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. "Object detection in 20 years: A survey". In: *Proceedings of the IEEE* 111.3 (2023), pp. 257–276.