Deep learning object detection for tracing the plasma portion of whole blood from images of medical sample containers

Dan-Sebastian Bacea[1,3], Volker von Einem[2], Jaiganesh Srinivasan[2]
D.bacea@stratec.com,
V.vonEinem@stratec.com,
J.Srinivasan@stratec.com

2023 IEEE 19th International Conference on Intelligent Computer Communication and Processing (ICCP) | 979-8-3503-7035-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCP60212.2023.10398632

***Abstract*— In this work, we explore the use of a deep learning neural network architecture for doing object detection of medical fluid samples, from color images. The ability of reliable identification and localization of liquid phases inside sample containers can help in automating significant parts of the work carried out in medical laboratories and hospitals. Our focus is on tracing the liquid levels of the plasma portion of whole blood, which may be further used in aliquoting, platelet rich plasma treatments automation, complete blood count analysis and others. The majority of sample containers are transparent and may contain or may not, one or more barcodes attached on the external surface, in addition to the manufacturer's labels. Consequently, the available clearance window to view the liquid contents is of variable size and shape, and may be sufficiently large only from few view angles of the sample container. To address the clearance window challenge, we introduce a new input representation technique, named vertical image stitching, that takes multiple images of a rotating sample container, extracts a region of interest from each of them and then stitches them together vertically in a single output image, called vertically stitched image. Based on the vertical stitched image, we introduce a new data augmentation technique named vertical stitch permutation, which enables obtaining 10 times more sample variations in the dataset. Finally, we convert the 2D bounding box localization computation into 1D points localization by removing from the loss function the bounding box components corresponding to the box height. We performed experiments on a dataset of 976 unique samples, containing 17 sample containers from 11 manufacturers. We trained YoloV4-tiny with the above mentioned techniques and achieved a mean average precision of 98.09% (at intersection over union threshold of 0.90) with an average intersection over union of 95.18%.**

Keywords—blood serum, plasma quantity tracing, convolutional neural network, object detection, lightweight object detector, laboratory sample containers

I. INTRODUCTION

Handling liquid samples, such as blood, urine, IV fluids, is required for numerous tasks in medical laboratories and hospitals. Automating these tasks can significantly increase the throughput of the medical facility, while reducing the workload on the staff. Significant effort has been dedicated in automating the systems [1] [2] [3] [4] [5], but still the majority of them rely on human vision for most of the liquid handling tasks. A couple research works have integrated some image processing algorithms [6], [7], [8], [9], [10], which were tested in controlled environments and in simple scenarios, so their deployment in real-world scenarios might require additional optimizations. AlexNet [11] was the publication that highlighted the feasibility of using deep convolutional neural networks (DCNN) for achieving state-of-the-art (SOTA) performance on computer vision tasks. Ever since, the focus in the computer vision research has started to shift towards deep neural networks.

Figure 1. Examples of medical laboratory sample containers placed in racks

There have been some efforts on using CNN based algorithms for tracing liquid boundary levels [12], [13], [14], [15], [16], [17]. Most of them tackle the image segmentation computer vision task. In this work we are using object detection for finding the liquid boundaries. Object detection is widely adopted for tasks related to robot vision, industrial quality inspection, object tracking etc. Object detection handles both localization and classification of objects of interest in images.

CNN based object detectors are either two-stage or single-stage. The two-stage detectors use the first stage for identifying the regions in the image that have a high probability of containing objects, while in the second stage the earlier proposed regions are classified and localized. In contrast, single-stage detectors do all the above mentioned computation in a single processing step. Typically, two-stage detectors achieve better detection performance at the cost of significantly higher computation required for doing the detections, while single-stage detectors need far less computation, but their detection performance is usually lower. The current trend shows significant improvements in the detection performance of the lightweight single stage detectors [18], [19], [20], [21], [22], [23], [24].

[1] Stratec Biomedical, Cluj-Napoca, Romania
[2] Stratec SE, Birkenfeld, Germany
[3] Technical University of Cluj-Napoca, Cluj-Napoca, Romania
979-8-3503-7035-5/23/$31.00 ©2023 IEEE

Training deep neural networks for computer vision tasks, such as classification, object detection, semantic segmentation, requires large datasets of annotated images. The broader the range of conditions that are contained in the training dataset, the better generalization capabilities of the achieved models, often surpassing human level accuracy on particular tasks. The trained DCNNs are as good as the quality of the data that they were trained on, so particular care needs to be taken when using such datasets. Collecting and annotating high quality images can lead to high costs (i.e. time and money). There are several publicly available datasets [9], [13], but the number of real blood samples and the diversity of sample containers is limited. For ensuring highest quality of the measurements and the diversity of sample containers, in this work we produced our own dataset.

This paper introduces a new method for tracing plasma portion boundaries of centrifuged whole blood by using object detection neural networks. The proposed method introduces a new input representation, data augmentation and updates to the bounding box coordinates computation. Our method achieves improved detection performance and real-time performance.

Our main contributions are:

- Vertical image stitching, an input representation that takes as input multiple images of a rotating sample container, extracts a region of interest (ROI) and then stitches them together vertically
- Random stitch permutation data augmentation technique, which takes as input a vertical stitched image and randomly permutes it's horizontal slices
- We convert the 2D bounding box localization computation into 1D points localization by removing from the loss function the bounding box components corresponding to the box height
- We perform experiments on a dataset of 976 stitched images (created from 9,760 individual images) containing 17 sample containers from 11 different manufacturers

The paper is organized as follows. Related work about tracing liquid levels in medical images and object detection are introduced in Section 2. Vertical image stitching, random stitch permutation and bounding box localization computation adjustments are presented in Section 3. In Section 4 we present experimental results and discuss them, focus being on mean average precision (mAP) and Intersection over Union (IoU) at different IoU thresholds. Finally, we draw the conclusions in Section 5.

II. Related Work

A. Fill level and phase boundary determination

Early approaches used for liquid level recognition in industrial bottles for filling determination were using capacitors or laser beams. These devices identify the change in the amount of dielectric material (i.e. capacitance) or reflectance in the liquid-to-air phase transition. Newer approaches switched to using cameras and computer vision. Some approaches use low-level image processing algorithms that identify edges or lines by detecting strong intensity gradients in the image [6], [25], [26], [27], [9], [28], [29], [30]. Structured light [31], [32] and colored floating beads [33] were used in some approaches for making the computer vision task easier to solve and more precise. These approaches are capable of recognizing single lines as phase boundaries, expecting the view angles need to be flat and the liquid surfaces to be flat.

Eppel et al. [9] propose a method that allows the recognition of liquid surfaces from various angles, achieved through scanning the image of the observed vessel line by line and finding the best correlated parabolic curve with the actual liquid interface. This approach is still limited to working reliably on liquid with flat surfaces. Dijkstra's algorithm [34] was also considered for obtaining the optimal curve between two pixels on a vessels contour, and using this curve as the phase boundary [35]. These approaches are limited to detecting phase boundaries that take the shape of a curve that starts and ends on the vessel contour. In [10] the author investigates the feasibility of using the graph cut method for detecting the phase boundaries of the contents of vessels. This approach leads to improved results, but with the limitation that small amount of materials within the vessel can't be detected reliably, and it's impacted by the surface reflection coming from the vessel.

One of the main limitations of the above mentioned methods is the limited number of tested scenarios (a couple of hundreds). Larger datasets became publicly available recently [13], [17], a couple of years after several publications that refer to these datasets by the authors [15] [16]. In [16] a semantic segmentation CNN is used for identifying pixel level categories for the materials found within vessels. An attention map that helps focus the network only on specific segments in the image is built and used. However, the performance of the network was significantly different from one material to the other, such as two phased liquid, suspensions, emulsion. In [15] the author proposes serially connecting two semantic segmentation CNNs, the first one segments the image into background and vessel, while the second one uses as input the original image along with the output segmentation map of the first CNN, which acts as an attention map. This approach shows promising results, improving performance on liquid phases and suspensions, but there is still room for improvement.

We haven't found any publication about using object detection for determining fill levels or phase boundaries in medical tubes.

B. Datasets

There are few publicly available datasets. The Materials-in-Vessels dataset was introduced with [16] and it contains 949 laboratory setting images of different materials in different phases involved in chemical processes. All the images were extracted from several YouTube channels dedicated to chemical experiments. Pixel level annotations are available for several classification layers, such as vessel/background, filled/empty, phase type and fine-grained physical phase type. The number of images in the dataset is small, the images are quite heterogeneous, view angles differ from image to image and real blood samples are

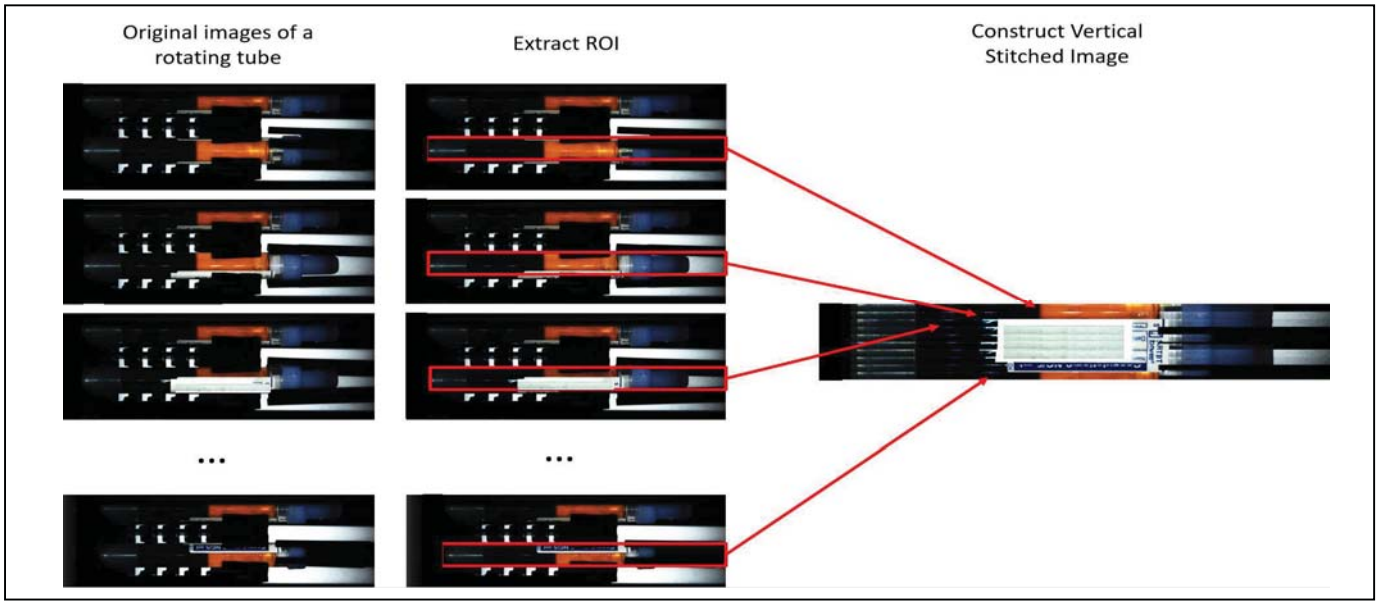

Figure 2. How a vertical stitched image is constructed. A region of interest (ROI) of fixed size is extracted from each initial input image of the rotating tube and then vertically concatenated to form a single output image

underrepresented in this dataset. Out of the available real blood samples, only few of them contain centrifuged real blood.

Vector-LabPics was introduced in [17] and it contains 2,187 laboratory setting images, but also in everyday situations like beverage handling. The annotations are pixel level, each material phase and its type being annotated, along with the vessel region, container labels, and corks. This dataset contains more samples than Materials-in-Vessels dataset and a bigger variety of conditions, but still the number of real blood samples is small, while the number of centrifuged real blood is even smaller.

LabPicsMedical dataset [13] is an extension to Vector-LabPics and it contains 1,300 images of medical samples of vessels containing liquids and solid materials. Pixel level annotations are available for different types of vessels (e.g. syringe, tube), liquids (e.g. urine, blood), phases of material (e.g. liquid, solid) and for optical properties of the vessel (e.g. transparent, opaque). The main limitation of this dataset is its small size and limited number of samples with real blood (403). The number of centrifuged real blood samples is also very low.

We weren't able to identify any public dataset containing more than 403 real blood annotated samples in tubes from different manufacturers. The impact of labels and barcodes on the clearance window inside the tube is also quite underrepresented in the public datasets. As our focus in this work is on tracing the plasma portion of centrifuged real blood, the number of such images in public datasets is significantly smaller. These limitations made us create our own dataset.

C. *Object detectors*

The computer vision task that addresses object category identification and localization of objects in images is named object detection. Several traditional image processing algorithms were studied in the literature for doing object detection, such as histograms of oriented gradients [36], the deformable part-based model proposed in [37] and the Viola Jones detectors [38]. These methods use manually engineered features, but their performance on large datasets and real-time performance make them challenging for deployment on edge devices. The publication of the results achieved with AlexNet [11] in 2012 made a huge change in the direction of research effort in the computer vision domain. This change was boosted by the ability of the convolutional neural networks (CNNs) to learn the best features from the data itself.

Current existing CNN based object detectors can be grouped in single-stage and two-stage detectors. Typically, two-stage detectors achieve higher accuracies, while requiring significantly more computation at inference time. The majority of lightweight object detectors are single-stage and their target is to work in real-time on resource-constrained devices, while achieving a sufficiently high accuracy. YoloV4-tiny [22] was the first lightweight object detector that achieved an mean average precision above 20% on the COCO dataset [39], while achieving over 10 FPS on a very low compute device such as NVIDIA Jetson Nano. YoloV7-tiny [23] is one of the latest lightweight models in the YOLO series, it significantly outperforms YoloV4-tiny while requiring only around 4 additional FPS on NVIDIA Jetson Nano. NanoDet [40] is one of the anchor-free detectors that also achieves real-time performance on edge devices, by using the Generalized Focal Loss [41] as classification and regression loss. YoloV8 [42] was released by Ultralytics in five scaled versions. It is also an anchor-free model with a decoupled head which processes objectness, classification and regression tasks independently. It also introduced the C2f module (cross-stage partial bottleneck with two convolutions) which enables better contextual information understanding for the model.

III. Proposed Solution

A. *Motivation*

We weren't able to find any public dataset with a sufficient number of centrifuged whole blood with plasma portion

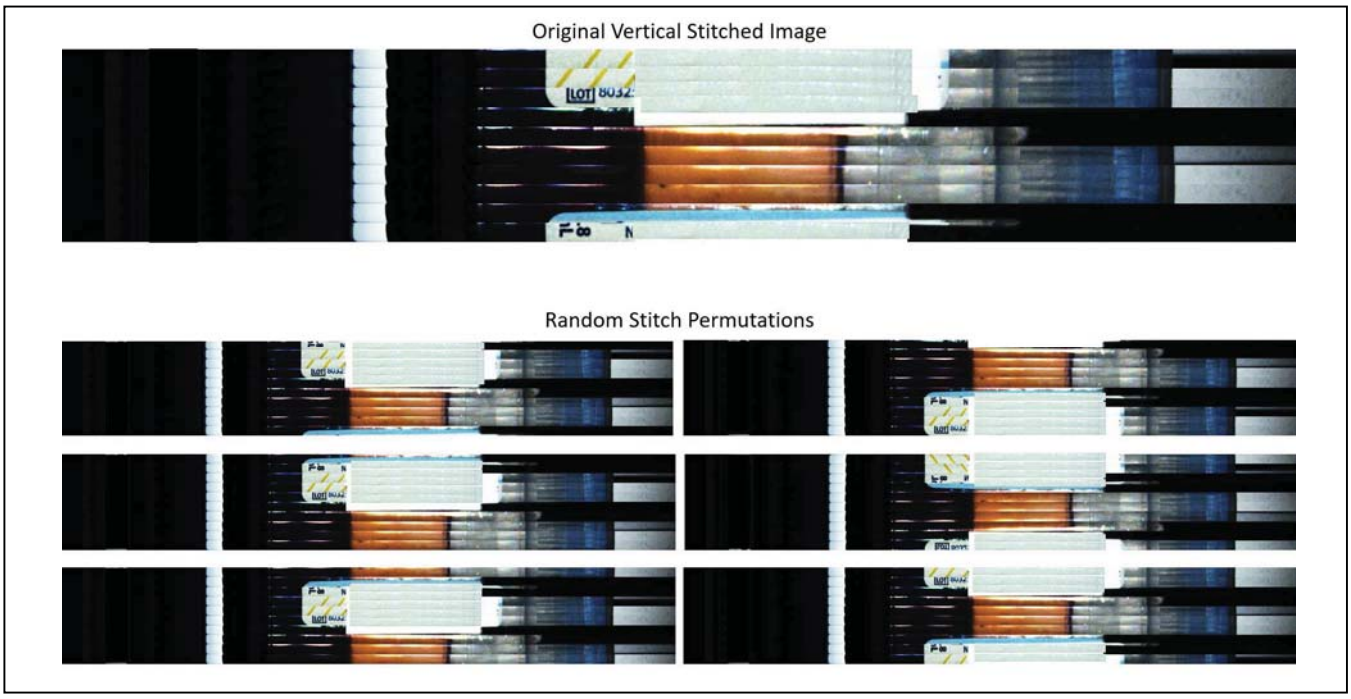

Figure 3. Random stitch permutations. From each initial vertically stitched image, multiple variations are generated by virtually simulating the scan of the tube from different initial positions with respect to the camera

annotated samples in tubes from different manufacturers. The sample containers that hold the real blood may contain or not one or more labels glued on the container walls (such as manufacturer label, barcodes etc.). The CLSI AUTO02-A2 [43] standard provides specifications for using barcodes on tubes in the clinical laboratory, which ensures that a clearance window is available for each tube after sticking the barcodes on the tubes. The size of the clearance window depends on several factors, such as the tube width, manufacturer label width and positioning, barcode label width and positioning. If the tube is imaged only from one side, it might be the case that its contents are visually obstructed. Being able to develop and test an algorithm that detects the plasma portion of centrifuged real blood samples that works in real world high throughput laboratory scenarios made us acquire our own dataset.

B. *Image acquisition setup*

Patient samples comprising whole blood are stored in consumables (e.g. tubes) and then placed in racks, which can accommodate several consumables at a time. Before inserting whole blood patient samples into the tubes, they are centrifuged. The camera is positioned at the same level with the racks which are presented in front of it (side view). The camera resolution is 2,048x1,536 pixels. Each tube from a rack is presented in front of the camera. The tubes are frontally illuminated with two white LEDs positioned above and below the rack, and also back illuminated with vertical white LED strips. The back illumination LED strips improve the image quality by eliminating shadows and better highlighting the tube edges. The front illumination LEDs enable to better visualize the contents of the tubes, but also introduce reflections on the tube center, which represent a challenge that the proposed algorithm should be robust against. Each tube is gripped with a two finger gripper and approximately 360 degrees rotated within the rack. While the tube is rotating, 10 color images are acquired and stored on the cameras storage device. When storing the images on the disk, they are 90 degrees rotated to the right, due to camera mount particularities.

C. *Vertical image stitching*

In the acquired 10 color images of a tube, there is a variable number of images in which the liquid content is visible through the clearance window. The number of suitable images for doing liquids level detection depends on the size of the clearance window. While the tube is rotating, the clearance window visible in the image depends on position of the tube at the moment when the image is acquired and it has also a variable size. Running a liquid levels detection algorithm on an image that has no clearance window or a too small clearance window would result in false positive results. At this point, there are two options: develop an algorithm that determines if a particular image has a sufficient clearance window, or find an input representation that is inherently robust against the clearance window variation. Instead of constructing an algorithm that detects if the image is suitable for doing liquid levels detections, we introduce the vertical image stitching algorithm, which extracts a region of interest (ROI) of 32 pixels from each of the 10 original images. The ROI is extracted relative to the tube center (which can be detected with classical computer vision algorithms, such as edge detectors). The extracted 10 ROIs are then vertically stitched (concatenated) together and produce the vertically stitched image. Liquid levels detection can then be run on the vertically stitched image, which ensures having a clearance window to the liquid contents of the tube (as long as the tube labeling follows the CLSI AUTO02-A2 standard).

D. *Random stitch permutation*

The vertical image stitching technique preserves the sequential order in which the original images were acquired when constructing the stitched image (i.e. ROI from first trigger image is located at the top, ROI from second trigger image is right below the top ROI and so on). The clearance window in the stitched image is not always a continuous

region, as it might be the case that the clearance window is split between top ROI and bottom ROI. The order in which the ROIs are used for creating the stitched image can be used as a data augmentation technique that can significantly enhance the sample visual variability in the dataset. A random integer number can be picked from 0 to 9. Based on this number, the final position in the vertical stitched image for each of the 10 original ROIs will be shifted with the random number of positions.

E. Bounding box computation transformation

When doing object detection with YOLO models, the network predicts 4 coordinates, 2 of them for the bounding box center location in the image, while the other 2 represent the height and width of the bounding box. When using the vertical image stitching input representation, the 2D localization of the bounding box can be simplified into 1D by fixing the Y-axis coordinate of the bounding box to value 0 and the Y-axis coordinate of the bounding box height to value 1. By doing this modification, the complexity of the function to learn by the network through training becomes much easier.

IV. Experimental Results

In this section, we present the experimental settings and datasets. Then the performance comparison under the metrics of mAP, BFLOP/s, model size and FPS are carried out.

A. Experimental settings

In our experiments we use the Darknet framework, which is open-source and written in C and CUDA, and which supports multiple CPUs and GPUs. We use Stochastic Gradient Descent (SGD) with momentum 0.9 as optimizer, and we set the initial learning rate to 0.00261. We set the batch size to 64 with 2 sub-divisions, as these values lead to a stable convergence of the learning process. The model iterates 9,000 times on the dataset that we constructed. The YOLO layer parameter settings are consistent with those of YoloV4-tiny with some exceptions. We use only one single object class, we recalculated to use 4 anchors for YoloV4-tiny on our dataset, we've fine-tuned the data augmentation parameters and we set the IoU threshold to 0.9. We use the pre-trained weights yolov4-tiny.conv.29. We train the networks on two NVIDIA Quadro RTX 6,000 GPUs. The input images are rescaled to 256x160 pixels resolution for training and inference. The detection speed is tested on NVIDIA Jetson Nano. In Table 1 we show the hardware specifications of NVIDIA Jetson Nano.

TABLE I. Hardware configuration of NVIDIA Jetson Nano

	NVIDIA Jetson Nano
Accelerator	128-core NVIDIA Maxwell GPU
TFLOPs (FP16)	0.5 TFLOP
CPU	Quad-core ARM Cortex-A57 @ 1.43 GHz
Memory Size	2 GB
Memory Type	LPDDR4
Interface	PCIe
Power	5W-10W
Price	121$

TABLE II. Plasma detection dataset details

	Plasma Detection Dataset
Number of classes	1
Samples in training set	678
Samples in validation set	298
Total number of ground-truth boxes	976
Average number of boxes per image	1

B. Experimental Datasets

The dataset that we constructed contains 9,760 original centrifuged real blood images. The real blood samples are stored in 17 different tube types coming from 8 different manufacturers (e.g. Sarstedt, Sekisui, Terumo, Nipro, Becton Dickinson, Evergreen, Greiner Bio-One and Corning). The tubes capacities range from 1.0 milliliters to 5.0 milliliters, and the tubes were filled from 70% up to 110% of their nominal fill volumes. From the original images, we extract a region of interest (ROI) of 32 pixels from each of the 10 original images of a scanned tube and vertical stitch them together, resulting in 976 unique vertical stitched samples of resolution 2,048x320 pixels. The number of samples per tube type is unequal, the average number of samples per tube type being 90, with a minimum of 23 samples for one of the tube types. We use only one object class. The annotations are bounding boxes done on a per vertical stitched image basis, in the YOLO format. The bounding boxes always span through all the height of the image (Y-axis point 0 to Y-axis maximum point). When calculating the AP and mAP values, we set the default value of 0.9 for the intersection over union (IOU) threshold and 0.95, because we are interested only in very well localized liquid levels. Table 2 shows the configuration that we used for the dataset.

C. Ablation Experiment on bounding box computation

To better understand the impact of the number of coordinates used in the bounding box coordinates regression, we train Yolov4-tiny with all 4 coordinates enabled, and then with only the X-axis coordinates enabled (center and width of the bounding box). The comparison can be observed in Table 3. Our first observation was about the convergence of the loss function, which is happening faster and it is more stable (less negative spikes, less variation in the trend) when using only 2 bounding box coordinates. The performance at IoU threshold 0.90 shows a +1.08% improvement for the average IoU and +0.73% improvement for the mAP. The performance at IoU threshold 0.95 shows a significant improvement of +5.73% for the average IoU and +7.25% for the mAP. This bounding box computation change proves to be most effective for high fidelity localization of the liquid levels.

Figure 4. Some example detections of the plasma portion for different sample containers and blood specimens. The green vertical lines represent the ground truth annotations, while the red vertical lines represent the predicted liquid levels. The absolute pixel index difference between the predicted liquid levels and ground truth levels is at most several pixels across the dataset

TABLE III. Ablation study on bounding box computation

Model	**Average IoU @ 0.90 (%)**	**mAP @ 0.90 (%)**	**Average IoU @ 0.95 (%)**	**mAP @ 0.95 (%)**
YoloV4-tiny trained with all 4 bounding box coordinates	93.74	97.17	63.62	47.73
YoloV4-tiny trained with only 2 bounding box coordinates	**94.82**	**97.90**	**69.35**	**54.98**

D. Random stitch permutation data augmentation

In the first step, we did a small scale experiment for testing the robustness against random stitch permutations of the Yolov4-tiny model trained with the vertically stitched images. We selected 10 samples for one tube type and offline augmented them to generate all the 10 possible stitch permutations, thus obtaining a total of 100 samples.

One false negative was observed on the 100 samples, the average confidence score was at 85.16% with a standard deviation of 19. Next we wanted to look at the confidence score range for each original sample along with its permutations. The average scan confidence range was 22.0% with a standard deviation of 23, while the maximum scan confidence range was 75% for one of the samples. These results strongly indicate the sensitivity of the model to such permutations, leading to the conclusion that the generalization capabilities of the model are limited. For mitigating these effects, we introduced into the training procedure of the model the random stitch permutation data augmentation. The technique is applied online while training the model.

The achieved results on the validation set are shown in Table 4. The performance at IoU threshold 0.90 shows a +0.36% improvement for the average IoU and +0.19% improvement for the mAP. The performance at IoU threshold 0.95 shows a significant improvement of +12.44% for the average IoU and +21.55% for the mAP.

The achieved performance on the small scale experiment mentioned in the first paragraph of this sub-chapter are shown in Table 5. The average confidence on the 100 samples increases with 6.65% (91.81%), while the standard deviation decreases with 8 (11.0). The average scan confidence range decreases with 6% (16%) and the standard deviation decreases with 8 (15.0). The results confirm that the robustness and stability of the model significantly improve.

The random stitch permutation data augmentation seems to be very effective in improving the robustness of the model, such a technique being of great value especially when the number of samples in the dataset is limited.

Next, we evaluated the real-time performance of the obtained model on a low-end GPU. When tested on NVIDIA Jetson Nano, the YoloV4-tiny model achieves 19.7 FPS, so it's able to perform detections in real time.

TABLE IV. Random stitch permutation data augmentation effect on performance of YoloV4-tiny

Model	**Average IoU @ 0.90 (%)**	**mAP @ 0.90 (%)**	**Average IoU @ 0.95 (%)**	**mAP @ 0.95 (%)**
YoloV4-tiny trained with only 2 bounding box coordinates	94.82	97.90	69.35	54.98
YoloV4-tiny trained with only 2 bounding box coordinates, trained with random stitch permutation	**95.18**	**98.09**	**81.79**	**76.53**

TABLE V. Performance of YoloV4-tiny on 10 samples offline augmented into 100 re-stitched samples

Model	Average confidence (%)	Standard deviation	Average scan confidence (%)	Standard deviation
YoloV4-tiny trained with only 2 bounding box coordinates	85.16	19.0	22.0	23.0
YoloV4-tiny trained with only 2 bounding box coordinates, trained with random stitch permutation	**91.81**	**11**	**16.0**	**15.0**

E. Performance comparison with other detectors

For comparing the performance of the YoloV4-tiny network that we fine-tuned and trained with 2 bounding box coordinates and with the random stitch permutation, we've selected the original implementations YoloV4-tiny and YoloV7-tiny. We've trained both models on our experimental dataset, trained for 9,000 iterations and kept every other parameter and hyper parameters identical to the official implementations. The achieved results are shown in Table 6.

YoloV7-tiny performs better than YoloV4-tiny, but our fine-tuned YoloV4-tiny significantly outperforms both. The average IoU @ 0.90 is with 4.71% better, mAP @ 0.90 is with 4.21% better, average IoU @ 0.95 improves with 15.81% and mAP @ 0.95 improves with 18.94%.

F. Qualitative performance analysis

In Figure 4 we present three scenarios in which the detection performance is very precise. Even though the plasma portion has very different visual characteristics, the model is able to accurately trace the true liquid levels.

G. Conclusions

This work demonstrated that tracing medical fluid samples with deep learning object detection is feasible. Our investigations were focused on whole blood plasma portion tracing. The experimental dataset contained 976 unique samples stored 17 different types of sample containers from 11 manufacturers. One of the main challenges was due to variability of the available clearance window size for viewing inside the sample containers. We developed and proved the efficacy of a new input representation method called vertical image stitching. Based on the proposed vertical image stitching we introduced a new data augmentation technique named random stitch permutation and we also changed the 2D bounding box computation to 1D points localization for the liquid levels tracing. We trained and fine-tuned YoloV4-tiny with the proposed input representation and augmentation techniques and achieved an average precision of 98.09% at intersection over union threshold of 0.90, with an average intersection over union of 95.18%. Our method proves high localization precision and robustness across container types and clearance

TABLE VI. Performance comparison with other detectors

Model	Average IoU @ 0.90 (%)	mAP @ 0.90 (%)	Average IoU @ 0.95 (%)	mAP @ 0.95 (%)
YoloV4-tiny [22]	89.35	90.63	57.14	40.98
YoloV7-tiny [23]	90.47	93.88	65.98	57.59
YoloV4-tiny trained with only 2 bounding box coordinates, trained with random stitch permutation (**ours**)	**95.18**	**98.09**	**81.79**	**76.53**

window sizes. The proposed model significantly outperforms the original YoloV4-tiny and YoloV7-tiny on the experimental dataset. Future work could be done for improving the localization precision of the model by switching to a newer state-of-the-art architectures for the object detector. Although we've generated and used a dataset of almost 1,000 unique samples, more samples could be added to the dataset to ensure sufficient coverage for all possible variations of real blood (e.g. presence of interferences such as high amounts of hemoglobin, lipids, white cells etc.)

V. REFERENCES

[1] C. Naugler and D. L. Church, "Automation and artificial intelligence in the clinical," *Critical reviews in clinical laboratory,* vol. 2, no. 56, pp. 98-110, 2019.

[2] J. R. Genzen, C.-A. D. Burnham, R. A. Felder, C. D. Hawker, G. Lippi and O. M. Peck Palmer, "Challenges and opportunities in implementing total laboratory automation," *Clinical chemistry,* vol. 2, no. 64, pp. 259-264, 2018.

[3] D. Oros, M. Pencic, J. Sulc, M. Cavic, S. Stankovski, G. Ostojic and O. Ivanov, "Smart Intravenous Infusion Dosing," *Applied Sciences,* vol. 2, no. 11, p. 513, 2021.

[4] Y. Melanie L., L. William, M. Allison R., A. Neil W. and B. Carey-Ann D., "Impact of total laboratory automation on workflow and specimen processing time for culture of urine specimens," *European Journal of Clinical Microbiology & Infectious Diseases,* vol. 12, no. 37, pp. 2405-2411, 2018.

[5] K. Patel, J. M. El-Khoury, C. W. Farnsworth, F. Broell, J. R. Genzen and T. K. Amukele, "Evolution of Blood Sample Transportation and Monitoring Technologies," *Clinical Chemistry,* vol. 67, no. 6, pp. 812-819, 2021.

[6] H. Zhu, "New algorithm of liquid level of infusion bottle based on image processing," in *International Conference on Information Engineering and Computer Science*, Wuhan, China, 2009.

[7] D. Pathak, R. N. Sarangapani, D. Katare, A. S. Gaikwad and M. El-Sharkawy, "IOT based solution for level detection using CNN and OpenCV," in *International Conference on Internet Computing (ICOMP)*, Athens, 2018.

[8] T. Zepel, V. Lai, L. P. E. Yunker and J. E. Hein, "Automated Liquid-Level Monitoring and Control using Computer Vision," ChemRxiv, Cambridge, 2020.

[9] E. Sagi and K. Tal, "Computer vision-based recognition of liquid surfaces and phase boundaries in transparent vessels, with emphasis on chemistry applications," arXiv preprint arXiv:1404.7174, 2014.

[10] E. Sagi, "Tracing liquid level and material boundaries in transparent vessels using the graph cut computer vision approach," arXiv preprint arXiv:1602.00177, 2016.

[11] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, Lake Tahoe, Nevada, United States, 2012.

[12] S. Khalapyan, L. Rybak, V. Nebolsin, D. Malyshev, A. Nozdracheva, T. Semenenko and D. Gavrilov, "Robotic System for Blood Serum Aliquoting Based on a Neural Network Model of Machine Vision," *Machines,* vol. 11, no. 3, p. 349, 2023.

[13] S. Eppel, H. Xu and A. Aspuru-Guzik, "Computer vision for liquid samples in hospitals and medical labs using hierarchical image segmentation and relations prediction," arXiv preprint arXiv:2105.01456, 2021.

[14] R. Dhaka, R. Singh, A. K. Dubey and S. Shah, "An Efficient Method for Identification of Liquid Samples in Hospitals and Medical Labs," in *International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, Greater Noida, India, 2022.

[15] S. Eppel, "Hierarchical semantic segmentation using modular convolutional neural networks," arXiv preprint arXiv:1710.05126, 2017.

[16] S. Eppel, "Setting an attention region for convolutional neural networks using region selective features, for recognition of materials within glass vessels," arXiv preprint arXiv:1708.08711, 2017.

[17] 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.

[18] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, Nevada, USA, 2016.

[19] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, Hi, USA, 2017.

[20] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[21] A. Bochkovskiy, C.-Y. Wang and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.

[22] C.-Y. Wang, A. Bochkovskiy and H.-Y. M. Liao, "Scaled-yolov4: Scaling cross stage partial network," in *Proceedings of the IEEE/cvf conference on computer vision and pattern recognition*, 2021.

[23] C.-Y. Wang, A. Bochkovskiy and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, Canada, 2023.

[24] D.-S. Bacea and F. Oniga, "Single stage architecture for improved accuracy real-time object detection on mobile devices," *Image and Vision Computing,* vol. 130, p. 104613, 2023.

[25] V. Karathanassi, C. Iossifidis and D. Rokos, "Application of machine vision techniques in the quality control of pharmaceutical solutions," *Computers in Industry,* vol. 32, no. 2, pp. 169-179, 1996.

[26] K. J. Pithadiya, C. K. Modi and J. D. Chauhan, "Comparison of optimal edge detection algorithms for liquid level inspection in bottles," in *Second international conference on emerging trends in engineering & technology*, Nagpur, Maharashtra, India, 2009.

[27] S. Schinwald, D. Engel and M. Seidler, "Efficient automated liquid detection in microplates," in *IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, Rome, Italy, 2012.

[28] K. J. Pithadiya, C. K. Modi and J. D. Chauhan, "Machine vision based liquid level inspection system using ISEF edge detection technique," in *Proceedings of the International Conference and Workshop on Emerging Trends in Technology*, Mumbai, India, 2010.

[29] W. Yuan and D. Li, "Measurement of liquid interface based on vision," in *Intelligent Control and Automation*, 2004.

[30] C. Zhao and Y. Chen, "A new liquid level measuring system of standard metal tank based on sub-pixel edge detection," in *IEEE International Conference on Control and Automation*, 2007.

[31] M. Kurahashi, H. Tomikawa and Y. Doi, "Method and device for detecting average liquid level in a bottle". United States of America Patent 4,733,095, 22 March 1988.

[32] S. Chakravarthy, R. Sharma and R. Kasturi, "Noncontact level sensing technique using computer vision," *IEEE Transactions on Instrumentation and Measurement,* vol. 51, no. 2, pp. 353-361, 2002.

[33] T.-H. Wang, M.-C. Lu, C.-C. Hsu, C.-C. Chen and J.-D. Tan, "Liquid-level measurement using a single digital camera," *Measurement,* vol. 42, no. 4, pp. 604-610, 2009.

[34] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische mathematik,* vol. 1, no. 1, pp. 269-271, 1959.

[35] S. Eppel, "Tracing the boundaries of materials in transparent vessels using computer vision," arXiv preprint arXiv:1501.04691, 2015.

[36] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE computer society conference on computer vision and pattern recognition*, San Diego, CA, USA, 2005.

[37] P. F. Felzenswalb, R. B. Girshick, D. McAllester and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE transactions on pattern analysis and machine intelligence,* vol. 32, no. 9, pp. 1627-1645, 2010.

[38] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE computer society conference on computer vision and pattern recognition*, Kauai, HI, USA, 2001.

[39] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *European conference on computer vision*, Zurich, Switzerland, 2014.

[40] N. Authors, "NanoDet," https://github.com/RangiLyu/nanodet, 2021.

[41] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang and J. Yang, "Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection," in *Advances in Neural Information Processing Systems 33*, 2020.

[42] G. Jocher, A. Chaurasia and J. Qiu, "YOLO by Ultralytics," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics. [Accessed 5 July 2023].

[43] P. J. Mountain, AUTO02 Laboratory Automation: Bar Codes for Specimen Container Identification, 2nd Edition, CLSI, 2006.