Supplementary Materials

Application of Object Detection and Action Recognition Toward Automated Recognition of Chemical Experiments

Ryosuke Sasaki,[a] Mikito Fujinami[b] and Hiromi Nakai*[a,b]

[a]Department of Chemistry and Biochemistry, School of Advanced Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

[b]Waseda Research Institute for Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan.

S-1. Computational time of object detection

Table S1 lists the prediction times for YOLOv8n and YOLOv8x. The total calculation time was 3.6 and 9.6 ms/image (278 and 104 fps) for YOLOv8n and YOLOv8x, respectively. YOLOv8n exhibited predictions approximately 2.6 times faster than those of YOLOv8x. The variance in calculation time is primarily caused by the inference time using the neural network model, which was 0.9 and 7.1 ms/image for YOLOv8n and YOLOv8x, respectively. The differences in the time taken for image pre-processing and post-processing were small.

Table S1. Computational time required for object detection using YOLOv8n and YOLOv8x.

Model	Total (ms/image)	Preprocess (ms/image)	Inference (ms/image)	Postprocess (ms/image)
YOLOv8n	3.6	0.1	0.9	2.6
YOLOv8x	9.6	0.1	7.1	2.4

S-2. Object detection for external data

LabPics dataset V2[1,2] was examined as the generality of the trained model for external data detection. The mAPs of Erlenmeyer flask and separatory funnel, which are included in both LabPics dataset and our dataset, are listed in Talbe S-2. The object detection model was directly applied to the LabPics dataset without any additional training. The mAP values by YOLOv8x model for Erlenmeyer flask and separatory funnel are 0.611 and 0.422, respectively. The values were lower than the test data in our dataset. The LabPics dataset includes many closed-up or small objects that are out of our dataset and trained model's range. When the whole objects were captured in an image, part of the detection works reasonably.

Table S2. Object detection accuracy for LabPics database.

Model	Number of images	Number of objects	mAP	
			YOLOv8n	YOLOv8x
Erlenmeyer flask	911	1262	0.563	0.611
Separatory funnel	451	528	0.252	0.422

References

[1] S. Eppel, H. Xu, M. Bismuth and A. Aspuru-Guzik, *ACS Cent. Sci.*, 2022, **6**(10), 1743–1752.

[2] https://zenodo.org/records/4736111