Improving YOLO-based Recognition Accuracy for Custom Object Set through TR3D Point Cloud Verification

Ki Hong Im [a*], Young Kyun Kim [a] and Pil Geun Jang [a]
[a]*Korea Atomic Energy Research Insititute, 111, Daedeok-Daero 989Beon-Gil, Yuseong-Gu, Daejeon, KOREA*
[*]*Corresponding author: khim@kaeri.re.kr*

***Keywords : Object Recognition, 3D Point Cloud, YOLO, TR3D**

1. Introduction

Recent advancements in deep learning have notably increased the demand for efficient object recognition systems in industrial settings. Models like YOLO, customized for specific application domains using tailored datasets, are prevalent due to their ability to detect predefined sets of objects with high accuracy. Despite these capabilities, YOLO often misidentifies or falsely recognizes objects not included in its training, which could pose significant risks in dynamic environments where working conditions and target objects frequently change.

To overcome these challenges, we introduce an extended object recognition method designed to enhance the reliability of YOLO in identifying objects accurately when it needs to handle other objects together which are not trained before. Our solution combines YOLO with the TR3D model (Towards Real-Time Indoor 3D Object Detection), which leverages 3D point cloud data for additional object verification [1]. This integration allows for real-time processing of images through both models simultaneously. The outputs are then merged in a post-processing stage, ensuring a more accurate final recognition result.

2. YOLO-TR3D Combined Object Recognition

In this study, even distinct object classes are used for the object recognition task. Fig. 1 shows the shape of the target objects used in this study. The object set contains some fruits, devices and items that could be found around us.

For this recognition task, images and point cloud data are taken using RGBD camera and they are used as source data to be preprocessed and establish the training data for both YOLO and TR3D models.

Fig. 1. Seven distinct object classes used for YOLO and TR3D training.

2.1. Image and 3D Point Cloud Data Acquisition

For the preparation of training data, labeling and segmentation of the extracted 2D images were carried out for the YOLO model, and ground truth bounding boxes were labeled for the corresponding 3D point cloud data used in the TR3D model [2]. Subsequently, to enhance model robustness, data augmentation was applied, resulting in datasets comprising approximately 7,000 images for YOLO and 4,000 images for TR3D. Fig. 2 shows an example of labeled training data for TR3D.

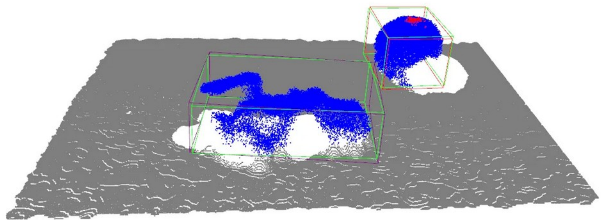

Fig. 2. Example of bounding box predictions generated by the TR3D model.

2.2 Training the TR3D Model on Custom Objects

The preprocessed 3D point cloud data were converted into the input format required by the TR3D model. Training was performed in a PyTorch and CUDA environment. Model performance was evaluated using Average Precision (AP) and Average Recall (AR) metrics. As shown in Table I, the model achieved high recognition performance at an Intersection over Union (IoU) threshold of 0.80.

Table I: Validation performance of the TR3D model

Classes	**AP_0.80**	**AR_0.80**
Apple	0.97	0.98
Grape	0.99	0.99
Glass	0.95	0.97
Elephant	0.99	1.00
Bulb	1.00	1.00
IPhone	1.00	1.00
Arduino	1.00	1.00
Overall	**0.99**	**0.99**

2.3. Test results

As mentioned above, to address the misrecognition issues observed when using the YOLO model alone, the TR3D model was incorporated. For the misrecognition case shown in Fig. 3, the corresponding 3D point cloud data from the same scene were processed using the TR3D model. As illustrated in Fig. 4, the TR3D model effectively utilizes the shape information learned from the 3D point cloud to largely eliminate erroneous detections of untrained objects. This significantly reduces the misrecognition rate observed with the YOLO-only model and improves the overall recognition accuracy.

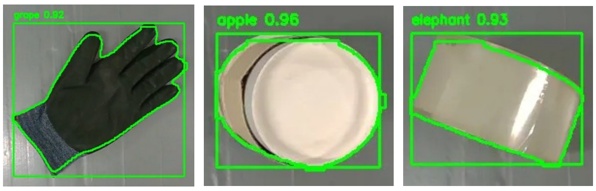

Fig. 3. Example of YOLO misrecognition.

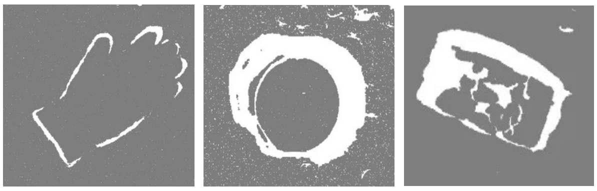

Fig. 4. Recognition results of the TR3D model for the YOLO misrecognition case shown in Fig. 3.

Fig. 5. Confusion matrix of the YOLO-only model.

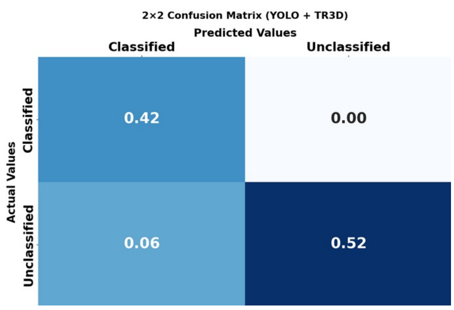

Fig. 6. Confusion matrix of the YOLO and TR3D extended model.

Fig. 5 shows the confusion matrix for the YOLO-only model, while Fig. 6 presents the results of the extended YOLO and TR3D approach based on 200 test trials. A comparative analysis indicates that the extended approach achieves considerably higher accuracy and precision. In contrast, the YOLO-only model exhibits a relatively elevated false positive rate that adversely affects its overall performance. The integrated method minimizes false positives while maintaining a high true positive rate, thereby substantially improving both accuracy and precision. As a result, the extended YOLO and TR3D approach demonstrates significantly enhanced reliability and robustness in object detection.

3. Conclusion

In this study, we presented an effective approach to mitigate misrecognition of untrained objects frequently encountered with the YOLO model. By extending the YOLO model with the TR3D model which leverages 3D point cloud data we substantially reduced false positives and improved recognition accuracy for custom object sets. Experimental results confirm that the extended YOLO and TR3D approach significantly alleviates misrecognition compared to the YOLO-only model. In future work, we will develop an early fusion pipeline that combines both models at the data level to further reduce real-time misrecognition, enabling more robust and reliable object detection in dynamic industrial environments.

ACKNOWLEGEMENT

This work was supported by Robot Industry Core Technology Development Programs of the Ministry of Trade, Industry & Energy of KOREA(20018270)

REFERENCES

[1] D. Rukhovich, A. Vorontsova, and A. Konushin, "TR3D: Towards Real-Time Indoor 3D Object Detection," arXiv:2302.02858v3 [cs.CV], Dec. 2023.

[2] Im, Ki Hong, Park, Jongwon, Lee, Jinyi and Kwak, Siwoo, "Automation and Post-Processing of 3D Object Point Cloud Model Generation Using Robotic Manipulators," Proc. of the Korean Society of Mechanical Engineers Autumn Meeting, 2024, Jeju, Korea.