Object Detection in Paddy Field for Robotic Combine Harvester
Based on Semantic Segmentation
（セマンティックセグメンテーションに基づく
ロボットコンバインのための物体検出）

ZHU JIAJUN

Abstract

For decades, agricultural vehicles such as robotic tractor and robotic combine harvester have been utilizing commercially available auto-steering and auto-driving systems for automatic and efficient operations. However, human supervision is still required for safe operations. To achieve fully autonomous farming, sensor technologies need to match or surpass human performance to detect all objects. Additionally, detection algorithms must operate in real-time to ensure the safety and efficient operations of the robotic vehicles. In this thesis, a semantic segmentation (SS) method was applied to detect all significant objects in paddy fields for robotic combine harvester.

In Chapter 1, the background of the research was introduced at first, then the research related to field objects detection was summarized. And then, the objective and overview of this research were described.

Chapter 2 introduced the experiment apparatus used in this thesis, including the Kubota WRH1200A, a commercialized robotic combine harvester equipped with an RTK-GNSS for navigation. Additionally, two RGB cameras and one depth camera installed on the harvester's cabin roof were introduced for obtaining paddy field images.

Chapter 3 described the utilization of a deep learning-based SS method, image cascade network (ICNet), for detecting objects in paddy fields. Six ICNet models and one fully convolution networks (FCN) model were developed for training and testing. The results showed that ICNet-VGG11 achieved the best performance in segmenting paddy field images, with high accuracy in pixel, class mean accuracy, and mean intersection over union. However, ridge detection was only successful when the harvester was close to the ridge, and the segmentation of unharvested and lodging areas was unstable. Nevertheless, the best model successfully detected lodging existence with high accuracy, which is crucial for the harvester's operation. Overall, the study concluded that the SS method effectively detected harvested rice, unharvested rice, lodging rice, humans, and paddy field ridges for the robotic combine harvester, despite the slow prediction speed and low segmentation accuracy in the lodging area.

To improve the segmentation accuracy and prediction speed, a new SS model, the robotic combine network (TRCNet), was designed specifically for the robotic combine harvester in Chapter 4. In TRCNet's design, context information extraction was enhanced while spatial information extraction was weakened. Five different lightweight CNNs were applied as the backbone of TRCNet to improve prediction speed. TensorRT was then used to accelerate the prediction speed of all models for real-time detection. The models were evaluated for detection accuracy and prediction speed, which were tested on Jetson TX2 after acceleration. The highest mean intersection over union (mIoU) model, TRCNet-MobileNetV3-Small ×4, achieved relatively higher intersection over unions (IoUs) for some classes but lower IoUs for others due to proximity limitations and mistake detection. A threshold value of 1000 pixels was set to improve detection accuracy of lodging existence, achieving an accuracy of 0.914 even with a relatively low IoU for the lodging area. Overall, this study concluded that TRCNet had higher segmentation accuracy and faster prediction speed than ICNet, with a fastest model achieving an FPS of 47.48, which is sufficient for real-time detection.

One issue with model training in Chapter 3 and Chapter 4 is the inclusion of far objects in the labeled data. These far objects are difficult to classify, particularly when distinguishing between unharvested and lodging areas. To address this problem, some far objects were labeled as the background class, but this approach can lead to misclassification of near objects. To overcome this challenge, a depth camera was used in Chapter 5 to extract only near objects for training. This approach eliminated the influence of far objects and increases labeling speed, resulting in higher quality and larger quantity of training data. Deep dual-resolution networks (DDRNets) were used as the SS models to detect objects in paddy fields. To improve prediction speed, an upgrade was made to obtain the Up-DDRNets. Two different datasets, original and filtered, were created to train and test the SS models for object detection in paddy fields. The filtered dataset provided higher quality and larger quantity of training data, which improved the accuracy of the SS models. Among the tested models, Up-DDRNet-re1 performed the best in terms of accuracy and speed, achieving real-time detection on the Jetson TX2. It had high pixel accuracy, mean accuracy, and mean IoU, and successfully detected the background, harvested area, human, and header classes. However, lower IoUs were observed for the ridge classes due to mistake detection, and for the lodging area and unharvested area classes due to similarity between these classes. The detection accuracy of the existence of ridge, human, and lodging rice were successfully improved by setting different threshold values.

Finally, conclusions and future perspectives of the research were described in Chapter 6.