BAB 1 INTRODUCTION

Improper disposal of waste highly pollutes the water bodies, thereby affecting communities and ecosystems living around it. Floating trash represent the increasing amount of waste problem in waterways. Between 4.8 and 12.7 million out of 215 million tons of plastic entered the ocean from 192 coastal countries in 2010 [1]. Indonesia also faces major water pollution challenges due to poor waste management (ranked 55th) [2]. Annually, 3.2 million tons of plastic waste generated in Indonesia result in 1.29 million tons reaching the ocean [3]. This pollution reduces light penetration and oxygen levels in marine habitats [4]. Raising public awareness and encouraging collaborative efforts in waste management are essential for minimizing these adverse effects.

Pollution of aquatic environments by trash was first monitored using manual inspections, which were inefficient and not scalable [5]. These methods were labor-intensive and made widespread monitoring difficult. Cameras have been used to assist, but they lack automatic detection and quantification. In Indonesia, current practices use satellite-tracked GPS beacons and manual methods [6], but GPS beacons and costly and lack real-time accuracy, while manual monitoring is inconsistent and resource-heavy [5]. Reseaerch on the detection of floating trash has been carried out in different water surfaces [7], [8], [9], [10], but it primarily focuses on general trash detection scenarios and does not consider fine-tuned scenarios for specific river conditions.

For this research on fine-tuning the trash detection model on specific scenarios, YOLOv11 is used due to its fast and accurate detection [8], [9], [11]. It uses the latest techniques, such as the C3k2, SPPF, and C2PSA block to improve the fusion of multi-scale features, which improves the accuracy as well as robustness [12], [13], [14]. The trained YOLOv11 models can recognize floating waste from multiple angles in relation to the water surfaces, which enormously eases the process of measuring and estimating the presence and amount of floating waste.

The YOLOv11 architecture is trained on a diverse dataset to ensure the robustness. First, the model learns from public dataset that were captured from different perspectives, as shown in Figure Ia. An ablation test is done to identified the combination of datasets that serves as the best, as illustrated in Figure Ib, which then used to evaluate different YOLOv11 sizes to see the impact on performance and accuracy depicted in Figure Ic. Additionally, new temporal data called BojongTrash is collected from local waterway near Telkom University at one frame per second, shown in Figure Id. The data is used to fine-tune the pre-trained model by training it with different sizes of data. This method, which is shown in Figure Ie is necessary for ascertainting the optimal amount of data needed for effective fine-tuning. The refined model is assessed using metrics like mAP50, precision, and recall, determining the optimal amount of data and epochs required to fine-tune the model in a given river scenario.

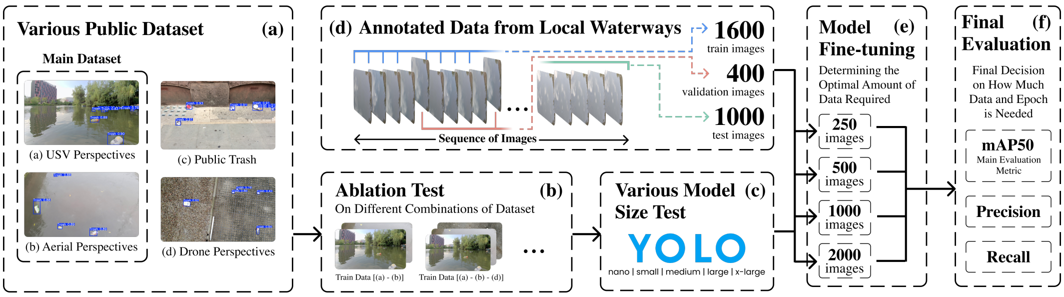

Figure I. Overview of the research workflow. (a) Public dataset used for training from diverse perspectives. (b) Ablation testing to identify the optimal dataset combination. (c) Testing different YOLOv11 model sizes. (d) Annotated data collected from local waterway near Telkom University. (e) Fine-tuning the model with different training sizes.(f) Final evaluation using various evaluation metrics.