DAFTAR GAMBAR

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. 2

Figure II. Architecture blocks of YOLOv11 [25]. 4

Figure III. Example of images from the public dataset used, showcasing different perspectives including USV, aerial, pedestrian, and low-altitude drone point of views. 6

Figure IV. (a) Example image from the BojongTrash dataset. (b) Surrounding condition of the waterways. 7

Figure V. The mAP50 scores at every 10-epoch interval fromepoch 50 to 300 on the WaterTrash and FloW-Img test dataset. 10

Figure VI. mAP50 and F1-scores across different fine-tuning scenarios. 12

Figure VII. Comparison of detection results from different models of the ablation test. (a) Original annotated FloW-Img test dataset. (b) Detection by YOLOv11-s showing false positive on reflection and waves. (c) Detection by YOLOv11-x showing false positive only on reflection. 13

Figure VIII. Sequential detection inconsistently from the fine-tuned model with 500 train images. (a) Trash detected confidently. (b) The same trash not detected. (c) The same trash is detected again. 14