

ABSTRAK

Pemilahan sampah plastik pada fasilitas daur ulang menghadapi tantangan besar pada tahap klasifikasi jenis polimer seperti PET, HDPE, PVC, dan PP akibat kemiripan visual yang tinggi dari segi warna, bentuk, dan tekstur permukaan objek. Kondisi tersebut menyebabkan objek sulit dibedakan secara konsisten oleh mata manusia maupun model deteksi berbasis citra RGB konvensional. Arsitektur deteksi standar seperti YOLOv8 belum secara eksplisit memanfaatkan mekanisme atensi untuk memfokuskan representasi fitur pada area yang relevan, sehingga rentan terhadap kesalahan klasifikasi pada polimer sejenis. Permasalahan tersebut mendorong penelitian mengenai penerapan fusi berbasis atensi pada YOLOv8 dalam mengintegrasikan data multimodal RGB dan *Near-Infrared* (NIR) untuk meningkatkan kinerja klasifikasi sampah plastik.

Penelitian eksperimental ini menggunakan data sekunder berupa pasangan citra RGB dan NIR sampah plastik yang mencakup sepuluh kategori polimer dengan anotasi bounding box. Tahap pra-pemrosesan data meliputi konversi format anotasi COCO ke YOLO, penyeragaman dimensi citra melalui interpolasi bilinear, serta augmentasi data menggunakan teknik *mosaic* dan *mixup* secara daring selama pelatihan. Metode pengembangan sistem dilakukan dengan merancang arsitektur FusedYOLOv8, yaitu modifikasi YOLOv8 menggunakan *dual-stream backbone* secara paralel. Fitur kedua modalitas digabungkan melalui fusi pada level P3, P4, dan P5 menggunakan modul *Convolutional Block Attention Module* (CBAM) untuk memperkuat selektivitas fitur secara kanal dan spasial. Model dilatih selama 50 epoch menggunakan *optimizer* AdamW, kemudian dievaluasi menggunakan metrik precision, recall, F1-score, dan mAP, disertai studi ablatasi.

Hasil pengujian menunjukkan model FusedYOLOv8 memperoleh nilai mAP@0.5 sebesar 0,9148, mAP@0.5:0.95 sebesar 0,8930, dan F1-score sebesar 0,9093. Hasil ini lebih tinggi dibanding model berbasis RGB saja dengan mAP@0.5 tertinggi 0,7531 maupun NIR saja dengan mAP@0.5 tertinggi 0,6783. Temuan ini menjawab permasalahan penelitian bahwa integrasi data multimodal RGB dan NIR memberikan informasi komplementer, di mana karakteristik spasial RGB dan spektral NIR secara bersama-sama meminimalkan kesalahan klasifikasi pada polimer yang identik secara visual. Kontribusi penelitian ini adalah desain modifikasi arsitektur fusi *dual-stream* YOLOv8 berbasis atensi CBAM untuk mengoptimalkan ekstraksi fitur multimodal. Evaluasi model menunjukkan keandalan yang cukup baik dengan nilai precision sebesar 0,8593 dan recall sebesar 0,9196, meskipun masih terdapat batasan berupa kesalahan klasifikasi minor pada kategori Trans_bottle dengan PPC dan PS.

Kata Kunci: Deteksi Objek, YOLOv8, Fusi Multimodal, RGB-NIR, CBAM

ABSTRACT

Plastic waste sorting in recycling facilities faces a significant challenge in classifying polymer types such as PET, HDPE, PVC, and PP due to high visual similarities in color, shape, and surface texture. Consequently, these objects are difficult to distinguish consistently by both the human eye and conventional RGB image-based detection models. Standard detection architectures like YOLOv8 do not explicitly utilize attention mechanisms to focus feature representation on relevant areas, rendering them susceptible to misclassifying similar polymers. This problem prompts research into the application of attention-based fusion in YOLOv8 to integrate multimodal RGB and Near-Infrared (NIR) data for improving plastic waste classification performance.

This experimental study utilized secondary data consisting of paired RGB and NIR plastic waste images spanning ten polymer categories with bounding box annotations. The data preprocessing stage included converting the annotation format from COCO to YOLO, unifying image dimensions via bilinear interpolation, and performing online data augmentation using mosaic and mixup techniques during training. The system development method involved designing the FusedYOLOv8 architecture, which modifies YOLOv8 using parallel dual-stream backbones. The features of both modalities were integrated through fusion at the P3, P4, and P5 levels using a Convolutional Block Attention Module (CBAM) to enhance channel and spatial feature selectivity. The model was trained for 50 epochs using the AdamW optimizer and evaluated using precision, recall, F1-score, and mAP metrics, along with an ablation study.

The test results demonstrated that the FusedYOLOv8 model achieved an mAP@0.5 of 0.9148, an mAP@0.5:0.95 of 0.8930, and an F1-score of 0.9093. These results were higher than those of the RGB-only model with a peak mAP@0.5 of 0.7531 and the NIR-only model with a peak mAP@0.5 of 0.6783. This finding addresses the research problem by confirming that the integration of multimodal RGB and NIR data provides complementary information, in which spatial RGB and spectral NIR characteristics collectively minimize misclassification among visually identical polymers. The contribution of this research lies in the structural modification of the dual-stream YOLOv8 fusion architecture based on CBAM attention to optimize multimodal feature extraction. The model evaluation indicates reasonable reliability with a precision of 0.8593 and a recall of 0.9196, although a limitation remains in the form of minor misclassifications between the Trans_bottle, PPC, and PS categories.

Keywords: Object Detection, YOLOv8, Multimodal Fusion, RGB-NIR, CBAM