ABSTRAK

Indonesia menempati peringkat ketiga sebagai negara penyumbang polusi plastik terbesar di dunia, dengan sampah plastik mendominasi komposisi limbah perkotaan. Model YOLO telah banyak digunakan untuk deteksi sampah plastik berbasis citra, namun sebagian besar penelitian terdahulu hanya berfokus pada peningkatan akurasi rata-rata tanpa memperhatikan ketimpangan performa antar kelas (*class bias*) akibat distribusi data yang tidak seimbang. Penelitian ini bertujuan menganalisis pengaruh model CLIP (*Contrastive Language-Image Pre-training*) sebagai tahap *post-processing* terhadap reduksi bias kelas pada hasil deteksi sampah plastik berbasis YOLOv8n, melalui dua pendekatan integrasi, yaitu *Selective Zero-Shot* dan prompt learning berbasis CoOp (*Context Optimization*).

Penelitian menggunakan dataset ML2 WCN UKM yang terdiri atas 8.978 citra dengan 8 kelas sampah plastik setelah proses penggabungan kelas. Model YOLOv8n dilatih sebagai *baseline*, kemudian hasil deteksinya diproses ulang oleh CLIP ViT-B/32 secara selektif berdasarkan zona ambiguitas skor kepercayaan. Pada pendekatan CoOp, *context vector* dioptimasi menggunakan *Class-Balanced Loss* dan *Logit Adjustment* untuk menangani ketidakseimbangan kelas, sementara parameter utama CLIP tetap dibekukan. Kalibrasi skor kepercayaan antara YOLO dan CLIP dilakukan menggunakan *Isotonic Regression* agar keputusan pembaruan label bersifat adil. Evaluasi dilakukan menggunakan metrik performa deteksi (*precision*, *recall*, F1-*score*, mAP@50) dan metrik bias kelas (*Gap AP*, *Standard Deviation AP*, *Min-Class AP*).

Hasil penelitian menunjukkan bahwa model *baseline* YOLOv8n mencapai mAP@50 sebesar 0,5802 dengan *Gap AP* sebesar 0,3367, mengindikasikan adanya ketimpangan performa antar kelas yang signifikan. Integrasi CLIP melalui *Selective Zero-Shot* menghasilkan mAP@50 sebesar 0,5665, sedangkan pendekatan CoOp menghasilkan mAP@50 sebesar 0,5190, keduanya lebih rendah dibandingkan *baseline*. Kedua pendekatan CLIP juga belum berhasil menurunkan *Gap AP* maupun meningkatkan *Min-Class AP* secara konsisten. Meskipun demikian, pendekatan CoOp menunjukkan nilai *Standard Deviation AP* terendah (0,1056) di antara ketiga konfigurasi, mengindikasikan sebaran performa antar kelas yang relatif lebih merata. Penurunan performa ini diduga disebabkan oleh *domain gap* antara data pelatihan CLIP dengan karakteristik citra lapangan, hilangnya detail visual akibat *resize* objek berukuran kecil, serta sifat fungsi *softmax* CLIP yang *forced-choice*. Penelitian ini menyimpulkan bahwa CLIP sebagai mekanisme *post-processing* belum efektif mereduksi *class bias* pada konfigurasi dan dataset yang digunakan, sekaligus memberikan pemahaman mengenai keterbatasan penerapan model *vision-language* pada domain deteksi objek dengan kondisi visual kompleks.

Kata Kunci: *CLIP, YOLOv8n, Class Bias, Post-Processing, CoOp,* Deteksi Sampah Plastik

ABSTRACT

Indonesia ranks third among the world's largest contributors to plastic pollution, with plastic waste dominating urban waste composition. YOLO-based models have been widely used for image-based plastic waste detection; however, most prior studies have focused primarily on improving average accuracy without addressing class bias, the performance imbalance across categories caused by uneven data distribution. This study aims to analyze the influence of CLIP (Contrastive Language-Image Pre-training) as a post-processing stage on reducing class bias in YOLOv8n-based plastic waste detection, through two integration approaches: Selective Zero-Shot and CoOp (Context Optimization) prompt learning.

The study uses the ML2 WCN UKM dataset, consisting of 8,978 images across 8 plastic waste classes after class merging. YOLOv8n was trained as the baseline detector, and its detection results were selectively re-evaluated by CLIP ViT-B/32 based on a confidence-score ambiguity zone. In the CoOp approach, a learnable context vector was optimized using Class-Balanced Loss and Logit Adjustment to address class imbalance, while the core CLIP parameters remained frozen. Confidence calibration between YOLO and CLIP was performed using Isotonic Regression to ensure fair label-override decisions. Evaluation was conducted using detection performance metrics (precision, recall, F1-score, mAP@50) and class bias metrics (Gap AP, Standard Deviation AP, Min-Class AP).

The results show that the baseline YOLOv8n model achieved an mAP@50 of 0.5802 with a Gap AP of 0.3367, indicating substantial performance imbalance across classes. CLIP integration via Selective Zero-Shot yielded an mAP@50 of 0.5665, while the CoOp approach yielded 0.5190, both lower than the baseline. Neither CLIP approach consistently reduced Gap AP or improved Min-Class AP. However, the CoOp approach recorded the lowest Standard Deviation AP (0.1056) among the three configurations, suggesting a relatively more even distribution of performance across classes. This performance degradation is attributed to the domain gap between CLIP's pretraining data and field-captured litter images, loss of visual detail from resizing small objects, and the forced-choice nature of CLIP's softmax function. This study concludes that CLIP as a post-processing mechanism has not yet proven effective in reducing class bias under the dataset and configuration used, while providing insight into the limitations of applying vision-language models to object detection domains with visually complex conditions.

Keywords: *CLIP, YOLOv8n, Class Bias, Post-Processing, CoOp, Plastic Waste Detection*