

ABSTRAK

Klasifikasi sampah yang efisien merupakan landasan penting dalam sistem pengelolaan limbah otomatis, terutama dalam menghadapi tantangan lingkungan dan operasional yang kompleks. Metode berbasis Vision Transformer (ViT) yang ada sering mengutamakan mekanisme atensi global dengan mengabaikan representasi fitur lokal sehingga membatasi adaptabilitasnya terhadap variasi visual dunia nyata, seperti pencahayaan, tekstur, dan orientasi objek. Penelitian ini memperkenalkan HR-ViT, model hybrid yang mengintegrasikan ResNet50 sebagai *feature extractor* dengan Vision Transformer untuk mengatasi keterbatasan ini. Dengan menggabungkan ekstraksi fitur konvolusional dan mekanisme *self-attention*, HR-ViT meningkatkan generalisasi pada enam kategori sampah: metal, kaca, organik, kertas, baterai, dan plastik. *Pipeline* pelatihan mencakup augmentasi data, seperti rotasi, *shear*, *zoom*, dan *flipping*, yang mengurangi risiko *overfitting* dan meningkatkan robustitas model. Diuji pada dataset sampah berisi 4.650 gambar dengan keragaman visual, HR-ViT mencapai akurasi validasi sebesar 99.13%, melampaui metode konvensional, seperti VGG19, SVM, ResNet50, dan Inception. Hasil ini membuktikan efektivitas model dalam mengatasi keterbatasan pendekatan ViT tradisional dan meningkatkan presisi klasifikasi. Penelitian lanjutan akan mengeksplorasi robustitas model dalam skenario multi-objek dan dataset dengan karakteristik visual lebih kompleks, mendukung sistem pengelolaan limbah cerdas yang skalabel.

Kata Kunci: Vision Transformer, ResNet50, Klasifikasi Sampah, Augmentasi Data, Model Hybrid

ABSTRACT

Efficient waste classification is a fundamental component of automated waste management systems, addressing critical environmental and operational challenges. Existing Vision Transformer (ViT)-based methods often prioritize global attention mechanisms while neglecting local feature representation, thus limiting their adaptability to real-world visual variations such as lighting, texture, and object orientation. This investigation introduces HR-ViT, a hybrid model integrating ResNet50 as a feature extractor and Vision Transformer to overcome these limitations. By combining convolutional feature extraction and self-attention mechanisms, HR-ViT enhances generalization across six waste categories: metal, glass, organic, paper, battery, and plastic. The training protocol incorporates comprehensive data augmentation techniques, including rotation, shear, zoom, and flipping, mitigating overfitting risks and improving model robustness. Evaluated on a diverse waste dataset comprising 4,650 images, HR-ViT achieves a validation accuracy of 99.13%, surpassing conventional approaches such as VGG19, SVM, ResNet50, and Inception. These results demonstrate the model's efficacy in addressing ViT's limitations while improving classification precision. Future research will explore its robustness in multi-object classification scenarios and diverse visual datasets, contributing to scalable, intelligent waste management systems.

Keywords: Vision Transformer, ResNet50, Waste Classification, Data Augmentation, Hybrid Model