ABSTRAK

Kerusakan jalan berlapis aspal masih menjadi persoalan utama dalam pengelolaan infrastruktur transportasi di Indonesia, dengan panjang jaringan jalan nasional mencapai 47.604,36 km. Pemantauan kondisi jalan masih banyak dilakukan melalui survei visual manual, sehingga membutuhkan waktu dan tenaga besar. Seiring perkembangan *deep learning*, metode deteksi otomatis berbasis citra mulai diterapkan. Namun, detektor berbasis *Convolutional Neural Network* (CNN) seperti YOLO mengalami penurunan kinerja pada kondisi visual yang bervariasi dan ketika data pelatihan terbatas. Oleh karena itu, penelitian ini merancang sistem deteksi kerusakan jalan melalui integrasi YOLOv8m dengan *Vision Transformer* (ViT) sebagai mekanisme verifikasi, serta pemanfaatan data sintetis berbasis *StyleGAN3-ADA* untuk meningkatkan variasi data pelatihan.

Dataset penelitian terdiri atas 1.758 citra kerusakan jalan dari data primer dan sekunder, serta 1.831 citra sintetis hasil augmentasi menggunakan *StyleGAN3-ADA*, sehingga total berjumlah 3.589 citra. Dataset dibagi menjadi 3.202 citra pelatihan, 259 citra validasi, dan 128 citra pengujian, dengan data validasi dan pengujian berasal dari dataset Indonesia. Kualitas citra sintetis dievaluasi menggunakan metrik *Fréchet Inception Distance* (FID), dengan nilai terbaik sebesar 42,5395 pada kelas *Pothole* dan 42,3484 pada kelas *Patch*.

Deteksi objek dilakukan menggunakan YOLOv8m, sedangkan hasil deteksi diverifikasi menggunakan ViT melalui klasifikasi ulang pada *region of interest* (ROI). Keputusan akhir diperoleh melalui metode *weighted score fusion* dan diimplementasikan dalam aplikasi berbasis Streamlit.

Hasil penelitian menunjukkan bahwa penggunaan dataset campuran (real + sintetis) meningkatkan performa YOLOv8m dengan mAP@0.5 dari 0,613 menjadi 0,781 (kenaikan 27,41%). Selain itu, akurasi ViT meningkat dari 0,9060 menjadi 0,9829 (kenaikan 8,49%). Pada pengujian akhir metode fusion, konfigurasi terbaik dengan *weight data asli* menghasilkan mAP@0.5 sebesar 0,4304, sedangkan *weight data campuran* menghasilkan mAP@0.5 sebesar 0,6022, sehingga terjadi peningkatan sebesar 39,92%. Dengan demikian, integrasi YOLOv8m–ViT serta augmentasi data sintetis *StyleGAN3-ADA* efektif meningkatkan kinerja deteksi kerusakan jalan pada kondisi dataset terbatas.

Kata Kunci: Deteksi kerusakan jalan, YOLOv8m, Vision Transformer, *fusion*, StyleGAN3-ADA, data sintetis, *weighted score fusion*

ABSTRACT

Asphalt-paved road damage remains a major issue in transportation infrastructure management in Indonesia, with the national road network reaching 47,604.36 km. Road condition monitoring is still largely conducted through manual visual surveys, requiring significant time and labor. Along with the rapid development of deep learning, image-based automatic detection methods have been widely applied. However, Convolutional Neural Network (CNN)-based detectors such as YOLO often experience performance degradation under diverse visual conditions and limited training data. Therefore, this study proposes a road damage detection system through the integration of YOLOv8m with a Vision Transformer (ViT) as a verification mechanism, as well as the utilization of synthetic data generated by StyleGAN3-ADA to increase training data diversity.

The dataset consists of 1,758 road damage images collected from primary and secondary sources, and 1,831 synthetic images produced through augmentation using StyleGAN3-ADA, resulting in a total of 3,589 images. The dataset was divided into 3,202 training images, 259 validation images, and 128 testing images, where the validation and testing sets were derived from an Indonesian dataset. The quality of synthetic image generation was evaluated using the Fréchet Inception Distance (FID), achieving the best scores of 42.5395 for the Pothole class and 42.3484 for the Patch class.

Object detection was performed using YOLOv8m, while detection results were verified by ViT through re-classification on the region of interest (ROI). Final decisions were obtained using a weighted score fusion method and implemented in a Streamlit-based application.

The results show that the mixed dataset (real + synthetic) improved YOLOv8m performance, increasing mAP@0.5 from 0.613 to 0.781 (a 27.41% improvement). In addition, ViT accuracy increased from 0.9060 to 0.9829 (an 8.49% improvement). In the final fusion evaluation, the best configuration with original-data weights achieved an mAP@0.5 of 0.4304, while mixed-data weights achieved 0.6022, resulting in a 39.92% performance improvement. Thus, the integration of YOLOv8m–ViT and synthetic data augmentation using StyleGAN3-ADA is effective in improving road damage detection performance under limited dataset conditions.

.

Keywords: *road damage detection, YOLOv8m, Vision Transformer, fusion, StyleGAN3-ADA, synthetic data, weighted score fusion*