

ABSTRACT

Rail surface defect detection is essential for maintaining railway operational safety. This study applies YOLOv11 to recognize five rail-surface defect classes: breaks, cracks, lightband, rails, and scars. The dataset consists of 1,596 images with more than 30,000 annotations, and the model is trained for 100 epochs using transfer learning. Systematic experiments are conducted by varying input resolution (640×640 and 1280×1280), applying data augmentation, and introducing an architectural modification using Efficient Multi-scale Attention (EMA). The best configuration, using 1280×1280 resolution with augmentation, achieves precision 0.828, recall 0.882, mAP@50 0.917, and mAP@50–95 0.688. Per-class evaluation indicates that breaks, cracks, rails, and lightband are detected with high accuracy, while scars remain the most challenging class with mAP@50 0.649 and recall 0.451. These findings show that detection performance is strongly influenced by visual separability and intra-class variance, particularly for scars, which exhibit irregular shapes, subtle textures, and visual similarity to the rail surface. This study demonstrates that increasing input resolution provides the most significant contribution to performance improvement, especially for small and visually subtle defects.

Keywords: *YOLOv11, object detection, rail surface defects, high-resolution training*

ABSTRAK

Deteksi cacat permukaan rel merupakan komponen penting dalam menjaga keselamatan operasional kereta api. Penelitian ini menerapkan YOLOv11 untuk mengenali lima kelas cacat permukaan rel, yaitu *breaks*, *cracks*, *lightband*, *rails*, dan *scars*. Dataset yang digunakan terdiri dari 1.596 citra dengan lebih dari 30.000 anotasi, dan model dilatih selama 100 epoch menggunakan pendekatan *transfer learning*. Eksperimen dilakukan secara sistematis melalui variasi resolusi input (640×640 dan 1280×1280), penerapan augmentasi data, serta modifikasi arsitektur menggunakan *Efficient Multi-scale Attention* (EMA). Konfigurasi terbaik, yaitu resolusi 1280×1280 dengan augmentasi, menghasilkan *precision* 0.828, *recall* 0.882, mAP@50 sebesar 0.917, dan mAP@50–95 sebesar 0.688. Evaluasi per-kelas menunjukkan bahwa *breaks*, *cracks*, *rails*, dan *lightband* terdeteksi dengan akurasi tinggi, sedangkan *scars* menjadi kelas paling menantang dengan mAP@50 sebesar 0.649 dan *recall* 0.451. Temuan ini menunjukkan bahwa performa deteksi sangat dipengaruhi oleh keterpisahan visual (*visual separability*) dan variasi intra-kelas (*intra-class variance*), khususnya pada *scars* yang memiliki bentuk tidak beraturan, tekstur halus, serta kemiripan visual dengan permukaan rel. Penelitian menunjukkan bahwa peningkatan resolusi input memberikan kontribusi paling signifikan terhadap peningkatan performa, terutama pada objek kecil dan cacat dengan karakteristik visual yang halus.

Kata Kunci: YOLOv11, deteksi objek, cacat permukaan rel, resolusi tinggi.