

ABSTRAK

Industri *e-commerce* menghadapi tantangan dalam pencarian produk yang masih bergantung pada input teks, sehingga menyulitkan pengguna yang hanya memiliki referensi visual. Fitur pencarian berbasis gambar (*visual search*) yang didukung oleh teknologi deteksi objek (*object detection*) muncul sebagai solusi, namun pemilihan model yang tepat menjadi krusial untuk menyeimbangkan akurasi dan kecepatan. Penelitian ini bertujuan untuk melakukan studi perbandingan komprehensif antara dua arsitektur *single-stage detector* populer, yaitu RetinaNet (dengan *backbone* ResNet-50 FPN) dan YOLOv8n, dalam konteks studi kasus deteksi empat kelas sepatu NIKE untuk aplikasi *e-commerce*. Proses penelitian mencakup tahap eksperimen pelatihan untuk kedua model guna menemukan konfigurasi *hyperparameter* optimal sebelum dievaluasi pada *dataset testing* yang sama. Evaluasi dilakukan berdasarkan metrik akurasi (mAP, *Precision*, *Recall*), kecepatan (latensi inferensi), dan efisiensi komputasi (jumlah parameter, GFLOPs, ukuran model). Hasil evaluasi menunjukkan keunggulan absolut dari YOLOv8n di semua aspek. YOLOv8n tidak hanya lebih akurat secara signifikan, dengan $mAP@50-95$ mencapai 82.6% dibandingkan RetinaNet (75.1%), tetapi juga jauh lebih andal dengan *precision* 96.7% berbanding 19.9%. Selain itu, YOLOv8n secara dramatis lebih efisien dengan latensi inferensi 5.03 ms (~15 kali lebih cepat) dan ukuran model yang 20 kali lebih kecil. Penelitian ini menyimpulkan bahwa YOLOv8n adalah model yang secara signifikan lebih unggul dan lebih sesuai untuk diimplementasikan pada sistem pencarian visual di *e-commerce*, yang tidak hanya merepresentasikan *trade-off* yang lebih baik, tetapi juga sebuah lompatan generasi dalam arsitektur deteksi objek.

Kata Kunci: Deteksi Objek, Pencarian Visual, E-commerce, YOLOv8, RetinaNet, Deep Learning, Studi Perbandingan.

ABSTRACT

The e-commerce industry faces challenges in product discovery, which still relies heavily on text-based input, creating difficulties for users who only have visual references. Image-based search (visual search), powered by object detection technology, has emerged as a solution, but selecting the right model is crucial to balance accuracy and speed. This study aims to conduct a comprehensive comparative study between two popular single-stage detector architectures, RetinaNet (with a ResNet-50 FPN backbone) and YOLOv8n, in the context of a case study on detecting four classes of NIKE shoes for e-commerce applications. The research process involved an experimental training phase for both models to find optimal hyperparameter configurations before evaluating them on the same testing dataset. The evaluation was based on accuracy metrics (mAP, Precision, Recall), speed (inference latency), and computational efficiency (parameter count, GFLOPs, model size). The evaluation results demonstrated the absolute superiority of YOLOv8n in all aspects. YOLOv8n was not only significantly more accurate, achieving an $mAP@50-95$ of 82.6% compared to RetinaNet's 75.1%, but also far more reliable with a precision of 96.7% versus 19.9%. Furthermore, YOLOv8n was dramatically more efficient, with an inference latency of 5.03 ms (~15 times faster) and a model size that was 20 times smaller. This research concludes that YOLOv8n is a significantly superior and more suitable model for implementation in a visual search system for e-commerce, representing not just a better trade-off but a generational leap in object detection architecture.

Keywords: Object Detection, Visual Search, E-commerce, YOLOv8, RetinaNet, Deep Learning, Comparative Study.