

Object detection yang akurat sangat penting bagi sistem transportasi cerdas, seperti *autonomous driving* dan *advanced driver-assistance systems*. Penelitian ini bertujuan untuk meningkatkan kinerja metode *object detection* berbasis Transformer, yaitu SAM-DETR, dalam mendeteksi objek-objek terkait lalu lintas pada skenario berkendara di Yogyakarta, Indonesia. Metode yang diusulkan adalah integrasi fitur-fitur multilevel dari Feature Pyramid Network (FPN), dengan melakukan konkatenasi pada seluruh level output FPN yang mana ukuran spasialnya telah disesuaikan agar sama dengan level terakhir, kemudian fitur-fitur tersebut dikonkatenasi sepanjang dimensi kanal (*channel*). *Feature maps* hasil konkatenasi ini selanjutnya ditransformasikan menjadi himpunan fitur yang digunakan sebagai input ke *encoder layer*. Metode konkatenasi fitur ini disebut sebagai Concatenated FPN (CatFPN). Hasil eksperimen menunjukkan bahwa SAM-DETR CatFPN yang diusulkan mengungguli SAM-DETR orisinal pada dataset primer, dengan peningkatan sebesar 3,7% mAP, 4,7% mAP_{0,5}, dan 3,2% mAP_{0,75}. Selain itu, kinerja metode yang diusulkan melampaui metode-metode deteksi lain yang berbasis DETR pada dataset COCO 2017, mencapai nilai mAP tertinggi sebesar 34,3%. Pada deteksi objek kecil, kinerja metode SAM-DETR CatFPN yang diusulkan sebanding dengan Deformable DETR, dengan skor mAP_s masing-masing sebesar 14,8% dan 15,0%.

Kata kunci : *object detection*, Transformer, Feature Pyramid Network, konkatenasi fitur, skenario berkendara.

ABSTRACT

Accurate object detection is crucial for intelligent transportation systems such as autonomous driving and advanced driver-assistance systems. This study aims to enhance the detection performance of the Transformer-based object detector SAM-DETR to detect traffic-related objects in real-world driving scenarios in Yogyakarta. The improvement is achieved by incorporating multi-level features from Feature Pyramid Network (FPN), combining all levels of FPN outputs by spatially resizing them to the same size as the final level and then concatenating them along the channel dimension. The concatenated feature map is then transformed into a set of features, which serve as the input to the encoder layer. We refer this feature concatenation method to as Concatenated FPN (CatFPN). The experiments shows that the proposed SAM-DETR CatFPN outperforms the original SAM-DETR on custom dataset, achieving improvements of 3.7% mAP, 4.7% mAP_{0.5}, and 3.2% mAP_{0.75}. Furthermore, the performance of the proposed method surpasses that of other DETR-like methods on COCO 2017 dataset, reaching the highest mAP of 34.3%. On small objects, the performance of the proposed SAM-DETR CatFPN is comparable to that of Deformable DETR, with mAP_s scores of 14.8% and 15.0%, respectively.

Keywords : object detection, Transformer, Feature Pyramid Network, feature concatenation, driving scenarios.