

## INTISARI

Ortofoto UAV beresolusi tinggi semakin banyak dimanfaatkan untuk ekstraksi data tapak bangunan (*building footprint*) karena mampu menampilkan secara detail objek bangunan. Akan tetapi, proses ekstraksi tapak bangunan secara otomatis/semi otomatis masih menjadi tantangan terutama pada wilayah dengan karakteristik bangunan yang beragam dan kompleks. Penelitian ini dilakukan untuk menghasilkan tapak bangunan secara otomatis melalui integrasi *deep learning* YOLOv12 dan *Segment Anything Model* (SAM) khususnya untuk objek bangunan kompleks, serta memanfaatkan hasilnya sebagai dasar pemodelan bangunan 3D *Level of Detail 2* (LOD-2). Fokus penelitian ini adalah mengintegrasikan deteksi objek dan segmentasi untuk memperoleh batas tapak bangunan yang lebih akurat, melakukan regularisasi poligon agar geometri bangunan lebih rapi, serta mengevaluasi kualitas tapak bangunan dan model 3D yang dihasilkan.

Penelitian dilakukan pada tiga klasifikasi wilayah yang mewakili kondisi bangunan beragam dan kompleks pada area urban (Kelurahan Embong, Surabaya), semi-urban (Kelurahan Klitren, Yogyakarta), dan *rural* (Desa Mnelalete, Timor Tengah Selatan, NTT) menggunakan data ortofoto UAV. Tahapan pengolahan meliputi pelatihan YOLOv12 untuk deteksi bangunan, pembentukan *bounding box* sebagai *prompt* untuk SAM, segmentasi tapak bangunan pada area terdeteksi, serta penggabungan hasil segmentasi multi-skala dan regularisasi poligon. Evaluasi ekstraksi dilakukan menggunakan metrik *precision*, *recall*, *F1-score*, dan IoU, serta analisis efisiensi waktu dibandingkan digitasi manual. Selanjutnya, tapak bangunan hasil regularisasi digunakan sebagai data *input* rekonstruksi bangunan 3D tingkat LOD2 pada perangkat lunak DREAM 3D, dan kualitas model 3D dievaluasi menggunakan RMSE yang mengacu pada KAK ATR/BPN tentang Pemetaan Foto Udara dan Lidar dalam Rangka Peningkatan Data Dasar 3D.

Hasil penelitian menunjukkan bahwa integrasi YOLOv12 dan SAM menghasilkan ekstraksi tapak bangunan yang lebih efisien dan akurat dibandingkan digitasi manual pada ketiga klasifikasi wilayah. Efisiensi waktu meningkat sebesar 87% (urban), 98% (semi-urban), dan 78% (*rural*). Secara kuantitatif, metode ini menghasilkan IoU rata-rata 86,34%, *precision* 93,83%, *recall* 91,55%, dan *F1-score* 92,66%. Tapak bangunan yang telah diregularisasi selanjutnya digunakan sebagai input rekonstruksi bangunan 3D tingkat LOD2 pada perangkat lunak DREAM 3D, dengan tipe atap yang teridentifikasi meliputi *complex*, *gable*, *flat*, *mansard*, dan *pyramid*. Evaluasi model 3D menunjukkan rata-rata RMSE 0,48 m, memenuhi standar KAK ATR/BPN ( $\leq 1,5$  m), sehingga pendekatan ini dinilai layak untuk mendukung pemodelan bangunan 3D serta pengembangan basis data 3D pada berbagai konteks wilayah.

**Kata kunci:** Tapak bangunan, Ortofoto UAV, YOLOv12, *Segment Anything Model* (SAM), Regularisasi poligon, LOD-2, Pemodelan 3D

## ABSTRACT

High-resolution UAV orthophotos are increasingly being used for building footprint extraction because they can display building objects in detail. However, automatic or semi-automatic building footprint extraction remains a challenge, especially in areas with diverse and complex building characteristics. This research was conducted to automatically generate building footprints through the integration of YOLOv12 deep learning and the Segment Anything Model (SAM), specifically for complex building objects, and to utilize the results as the basis for Level of Detail 2 (LOD-2) 3D building modeling. The focus of this research is to integrate object detection and segmentation to obtain more accurate building footprint boundaries, perform polygon regularization to make the building geometry neater, and evaluate the quality of the resulting building footprints and 3D models.

The research was conducted on three classifications of areas representing diverse and complex building conditions in urban (Embong Village, Surabaya), semi-urban (Klitren Village, Yogyakarta), and rural (Mnelalete Village, South Central Timor, NTT) areas using UAV orthophoto data. The processing stages included training YOLOv12 for building detection, forming bounding boxes as prompts for SAM, segmenting building footprints in the detected areas, and combining multi-scale segmentation results and polygon regularization. The extraction evaluation was performed using precision, recall, F1-score, and IoU metrics, as well as an analysis of time efficiency compared to manual digitization. Furthermore, the regularized building footprints were used as input data for LOD2-level 3D building reconstruction in DREAM 3D software, and the quality of the 3D models was evaluated using RMSE in accordance with KAK ATR/BPN on Aerial Photo and Lidar Mapping for the Improvement of 3D Basic Data.

The results show that integrating YOLOv12 and SAM produces more efficient and accurate building footprint extraction than manual digitization across all three area classifications. Time efficiency increased by 87% (urban), 98% (semi-urban), and 78% (rural). Quantitatively, this method produced an average IoU of 86.34%, precision of 93.83%, recall of 91.55%, and F1-score of 92.66%. The regularized building footprints were then used as input for LOD2-level 3D building reconstruction in DREAM 3D software, with identified roof types including complex, gable, flat, mansard, and pyramid. Evaluation of the 3D model showed an average RMSE of 0.48 m, meeting the KAK ATR/BPN standard ( $\leq 1.5$  m). These results confirm that the proposed approach not only improves the quantitative accuracy of building footprint extraction but is also feasible for 3D building modeling, the development of urban 3D databases, and other spatial activities.

**Keywords:** Building footprint, UAV orthophoto, YOLOv12, Segment Anything Model (SAM), Polygon regularization, LOD-2, 3D modeling