

## INTISARI

Tapak atap bangunan merupakan informasi spasial yang merepresentasikan batas terluar dari atap suatu bangunan. Tapak atap bangunan merupakan dasar dalam pembuatan model 3D bangunan. Informasi tapak atap dapat diperoleh dari ortofoto beresolusi tinggi dengan teknik segmentasi. Seiring berkembangnya kecerdasan buatan (AI), pendekatan berbasis *deep learning* mulai banyak digunakan untuk menggantikan metode manual dalam ekstraksi tapak atap bangunan. Salah satu model segmentasi yang populer saat ini adalah *Segment Anything Model* (SAM), yang mampu melakukan segmentasi presisi hingga tingkat piksel. Meskipun demikian, untuk mendapatkan fitur tapak atap bangunan dalam ortofoto, SAM membutuhkan input berupa petunjuk awal (*prompt*) seperti titik atau kotak (*bounding box*) agar tahu bagian objek mana yang harus disegmentasi. Oleh karena itu, dibutuhkan bantuan model pendeteksi objek (*object detection*) untuk memberi petunjuk awal bagi SAM agar proses segmentasi bisa berjalan otomatis dalam satu *deep learning pipeline* pada ortofoto. Salah satu model *deep learning* unggulan yang mutakhir digunakan untuk mendeteksi objek adalah *You Only Look Once* (YOLO), yaitu model *object detection* yang mampu mengenali dan menandai objek yang ada di gambar dalam satu kali proses (*single-shot object detection*). YOLO menghasilkan *bounding box* yang berperan sebagai penunjuk posisi objek dalam ortofoto lalu disegmentasi menggunakan SAM. Rangkaian alur kerja kedua algoritma tersebut disebut sebagai *pipeline* YOLO-SAM.

Tahapan awal penelitian dimulai dengan membuat *dataset* dari ortofoto resolusi tinggi daerah gedung dan perumahan yang dipotong menjadi 400 gambar berukuran 1280 piksel, dengan potongan tumpang tindih sebesar 50 persen. Setiap gambar diberi label (anotasi) bangunan secara manual menggunakan Roboflow, lalu *dataset* dibagi menjadi data pelatihan dan validasi. Untuk memperbanyak variasi data, dilakukan augmentasi atau modifikasi gambar pelatihan. Model YOLOv12 dilatih dalam tiga varian model, yaitu kecil (s), sedang (m), dan sangat besar (x), untuk dibandingkan performanya dalam proses deteksi objek bangunan. Hasil deteksi berupa *bounding box* kemudian digunakan sebagai masukan untuk SAM, yang bertugas mensegmentasi atap bangunan. SAM menghasilkan tiga hasil segmentasi (*mask*) untuk setiap objek, lalu dipilih satu *mask* dengan skor tertinggi sebagai hasil akhir. Hasil ini kemudian diubah dari posisi piksel menjadi koordinat peta dan disimpan dalam format GeoJSON. Terakhir, bentuk atap yang telah disegmentasi disederhanakan menggunakan algoritma *Ramer–Douglas–Peucker* di QGIS agar lebih ringan namun tetap mewakili bentuk aslinya. Untuk menilai kinerja model, deteksi objek dievaluasi menggunakan metrik mAP50, *precision*, dan *recall*, sedangkan hasil segmentasi SAM dinilai menggunakan *Intersection over Union* (IoU) dan *Dice Similarity Coefficient* (DSC).

Dari tiga varian model deteksi objek yang diuji, YOLOv12-x menunjukkan performa terbaik dengan mAP50 sebesar 92,11 persen, *precision* 93,05 persen, dan *recall* 83,20 persen. Dipadukan dengan SAM, model ini menghasilkan segmentasi tapak atap dengan akurasi IoU 86,50 persen dan DSC 92,80 persen, serta tetap efisien setelah penyederhanaan geometri. Penerapan di area penelitian seluas 149 hektare mampu mendeteksi 544 atap bangunan, setara 80,71 persen area penelitian. Area dibagi menjadi gedung dan perumahan, dengan akurasi deteksi benar (*True Positive*) masing-masing 82,35 persen dan 76,88 persen. Performa lebih tinggi dicapai pada area gedung yang memiliki bangunan terpisah dan atap seragam, dibanding perumahan yang padat dan beragam.

**Kata kunci:** Segmentasi Tapak Atap, Ortofoto, YOLO, Segment Anything Model (SAM), *Deep Learning Pipeline*

## ABSTRACT

The building roof outline is spatial information that represents the outer boundary of a building's roof. The roof footprint of a building is the basis for creating a 3D model of the building. Roof edge information can be obtained from high-resolution orthophotos using segmentation techniques. With the advancement of artificial intelligence (AI), deep learning-based approaches are increasingly being used to replace manual methods in extracting building roof edges. One of the most popular segmentation models today is the Segment Anything Model (SAM), which can perform segmentation with pixel-level precision. However, to obtain building roof footprint features in orthophotos, SAM requires input in the form of initial prompts, such as points or bounding boxes, to determine which parts of the object should be segmented. Therefore, an object detection model is needed to provide initial prompts for SAM so that the segmentation process can run automatically within a single deep learning pipeline on orthophotos. One of the most advanced deep learning models currently used for object detection is You Only Look Once (YOLO), an object detection model capable of recognizing and marking objects in an image in a single process (single-shot object detection). YOLO generates bounding boxes that serve as position indicators for objects in orthophotos, which are then segmented using SAM. The workflow sequence of these two algorithms is referred to as the YOLO-SAM pipeline.

The initial stages of the research began with creating a dataset from high-resolution orthophotos of building and residential areas that were cropped into 400 images measuring 1280 pixels, with a 50 percent overlap. Each image was manually labeled (annotated) with buildings using Roboflow, then the dataset was divided into training and validation data. To increase data variation, augmentation or modification of the training images was performed. The YOLOv12 model was trained in three model variants, namely small (s), medium (m), and very large (x), to compare their performance in the building object detection process. The detection results in the form of bounding boxes were then used as input for SAM, which was tasked with segmenting building roofs. SAM produced three segmentation results (*masks*) for each object, and the *mask* with the highest score was selected as the final result. These results are then converted from pixel positions to map coordinates and saved in GeoJSON format. Finally, the segmented roof shapes are simplified using the Ramer–Douglas–Peucker algorithm in QGIS to make them lighter while still representing their original shapes. To assess the model's performance, object detection is evaluated using the mAP50 metric, precision, and recall, while SAM segmentation results are evaluated using Intersection over Union (IoU) and Dice Similarity Coefficient (DSC).

From the three object detection models tested, YOLOv12-x showed the best performance with a mAP50 of 92.11 percent, precision of 93.05 percent, and recall of 83.20 percent. When combined with SAM, this model produces roof edge segmentation with an IoU accuracy of 86.50 percent and a DSC of 92.80 percent, while remaining efficient after geometric simplification. When applied to the 149-hectare area of research, it detected 544 building roofs, equivalent to 80.71 percent of the study area. The area was divided into buildings and residential areas, with true positive detection accuracy of 82.35% and 76.88%, respectively. Higher performance was achieved in building areas with separate structures and uniform roofs compared to densely populated and diverse residential areas.

**Keywords:** Building Rooftop Outline Segmentation, Orthophoto, YOLO, Segment Anything Model (SAM), Deep Learning Pipeline