UNIVERSITAS
GADJAH MADA
**Perbandingan Hadil Ekstraksi Bangunan Metode SVM, Mask R-CNN, dan Interpretasi Visual pada Foto
Udara Kawasan The Park Sawangan, Kota Depok**
Camila Fauzia, Dr. Barandi Sapta Widartono, S.Si., M.Si., M.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

PERBANDINGAN HASIL EKSTRAKSI BANGUNAN METODE SVM, MASK R-CNN, DAN INTERPRETASI VISUAL PADA FOTO UDARA KAWASAN THE PARK SAWANGAN, KOTA DEPOK

Oleh:

Camila Fauzia

22/497339/SV/21119

INTISARI

Foto udara tegak resolusi tinggi merupakan basis data utama dalam pemetaan bangunan perkotaan, namun pemilihan metode ekstraksi menentukan kualitas dan efisiensi data yang dihasilkan. Penelitian ini membandingkan tiga pendekatan ekstraksi atap bangunan di Kawasan The Park Sawangan, Kota Depok menggunakan foto udara tegak tahun 2024 dengan resolusi spasial 0,047 meter, yaitu metode *Support Vector Machine* (SVM), *deep learning Mask Region-based Convolutional Neural Network* (Mask R-CNN), dan interpretasi visual. Evaluasi akurasi dilakukan menggunakan metrik *Intersection over Union* (IoU) dengan data referensi hasil interpretasi visual sebagai *ground truth*.

Metode SVM menghasilkan klasifikasi berbasis nilai spektral piksel RGB dengan waktu pemrosesan sekitar 30 menit tanpa GPU khusus, dan menghasilkan nilai IoU rata-rata sebesar 0,5687 (56,87%). Performa terbaik SVM ditemukan pada atap berbahan tanah liat berwarna oranye atau cokelat, namun metode ini mengalami banyak kesalahan klasifikasi pada objek non-bangunan yang memiliki kemiripan warna dengan atap seperti area parkir semen. Metode Mask R-CNN dilatih menggunakan 64 poligon sampel yang mencakup variasi bahan dan warna atap serta kelas non-bangunan, dan menghasilkan nilai IoU rata-rata sebesar 0,7801 (78,01%). Mask R-CNN terbukti lebih konsisten dalam mendeteksi berbagai variasi bahan atap berkat kemampuannya mengenali objek berdasarkan bentuk dan konteks spasial, meskipun memerlukan dukungan GPU dan waktu pelatihan sekitar 5–6 jam. Interpretasi visual menghasilkan delineasi yang paling presisi pada seluruh kondisi atap, namun memerlukan sumber daya waktu dan tenaga yang lebih besar. Meskipun Mask R-CNN menghasilkan akurasi tertinggi di antara kedua metode otomatis, nilai tersebut belum mencapai ambang performa optimal untuk ekstraksi bangunan berbasis citra resolusi tinggi (≥80%).

Kata kunci: Ekstraksi Bangunan, Foto Udara Tegak, SVM, Mask R-CNN, Interpretasi Visual, IoU.

UNIVERSITAS
GADJAH MADA
**Perbandingan Hadil Ekstraksi Bangunan Metode SVM, Mask R-CNN, dan Interpretasi Visual pada Foto
Udara Kawasan The Park Sawangan, Kota Depok**
Camila Fauzia, Dr. Barandi Sapta Widartono, S.Si., M.Si., M.Sc.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

A COMPARATIVE STUDY OF BUILDING EXTRACTION USING SVM, MASK R-CNN, AND VISUAL INTERPRETATION ON AERIAL IMAGERY OF THE PARK SAWANGAN AREA, DEPOK CITY

By:

Camila Fauzia

22/497339/SV/21119

ABSTRACT

High-resolution orthophoto is the primary data source for urban building mapping, yet the choice of extraction method determines the quality and efficiency of the resulting data. This study compares three building rooftop extraction approaches in The Park Sawangan area, Depok City, using 2024 orthophoto imagery with a spatial resolution of 0.047 meters: the Support Vector Machine (SVM) method, the Mask Region-based Convolutional Neural Network (Mask R-CNN) deep learning model, and visual interpretation. Accuracy was assessed using the Intersection over Union (IoU) metric with visual interpretation results as ground truth.

SVM performed pixel-based classification using RGB spectral values, completing the entire process in approximately 30 minutes without dedicated GPU hardware, and achieved an average IoU of 0.5687 (56.87%). SVM performed best on clay tile roofs with distinctive orange or brown coloring, but produced significant misclassification on non-building objects with similar spectral characteristics such as cement parking areas. Mask R-CNN was trained using 64 annotated polygons covering various rooftop material types and non-building classes, achieving an average IoU of 0.7801 (78.01%). Mask R-CNN demonstrated greater consistency across rooftop material variations due to its ability to recognize objects based on shape and spatial context, though it required GPU support and approximately 5–6 hours of training. Visual interpretation produced the most precise delineation across all rooftop conditions but required considerably greater time and human resources. Although Mask R-CNN achieved the highest accuracy among the two automated methods, the value has not yet reached the optimal performance threshold for building extraction from high-resolution imagery (≥80%).

Keywords: *Building Extraction, Orthophoto, SVM, Mask R-CNN, Visual Interpretation, IoU.*