



# Penerapan Metode Deep Learning Mask R-Cnn Untuk Identifikasi Bangunan Baru Sebagai Objek Pajak Di Desa Gadingmangu, Kecamatan Perak, Kabupaten Jombang

*Agus Purnawan, Dr.Ir. Tamaji, M.T*

*Teknik Informatika, Universitas Widya Kartika Surabaya. Email: m.purnawan@gmail.com*

## STATUS ARTIKEL

Dikirim 10 Oktober 2025

Direvisi 20 Oktober 2025

Diterima 31 Oktober 2025

### Kata Kunci:

*Mask R-CNN; citra satelit; deteksi bangunan; deep learning; segmentasi instance*

## ABSTRAK

Pertumbuhan kawasan permukiman di wilayah pedesaan terus meningkat, sehingga banyak muncul bangunan baru yang belum tercatat sebagai objek pajak. Proses pendataan secara manual yang dilakukan petugas lapangan sering kali memerlukan waktu lama dan berpotensi menimbulkan kesalahan pencatatan. Dalam penelitian ini, digunakan pendekatan Mask R-CNN untuk mengenali objek bangunan pada citra satelit beresolusi tinggi di Desa Gadingmangu, Kecamatan Perak, Kabupaten Jombang. Data citra melalui tahap pra-pemrosesan, anotasi, dan konversi ke format COCO sebelum digunakan dalam proses pelatihan model berbasis *backbone* ResNet-50 yang dikombinasikan dengan Feature Pyramid Network (FPN). Hasil pengujian menunjukkan bahwa seluruh tahapan — mulai dari pengolahan data, pelatihan model, hingga proses prediksi — dapat dijalankan dengan baik tanpa kendala teknis. Meski demikian, model belum mampu mengenali bangunan secara optimal karena jumlah data latih yang terbatas dan durasi pelatihan yang singkat. Penelitian ini memberikan bukti konsep (proof of concept) penerapan Mask R-CNN untuk deteksi bangunan pada citra satelit dan berpotensi dikembangkan lebih lanjut guna mendukung pembaruan data Pajak Bumi dan Bangunan (PBB) secara otomatis di tingkat daerah.

## 1. PENDAHULUAN

Pertumbuhan Jumlah penduduk Indonesia mengalami peningkatan setiap tahunnya, yang mendorong perluasan kawasan permukiman di berbagai daerah. Laju urbanisasi yang tinggi menimbulkan tekanan terhadap lahan pertanian, terutama di daerah yang dekat dengan jalur infrastruktur utama. Berdasarkan laporan Badan Pusat Statistik (BPS, 2023), sebagian besar wilayah perkotaan mengalami penurunan luas lahan pertanian akibat pembangunan perumahan dan fasilitas publik. Fenomena ini juga terlihat jelas di Kabupaten Jombang, khususnya di Kecamatan Perak, di mana banyak area sawah berubah menjadi kawasan terbangun.

Pembangunan Jalan Tol Mojokerto–Kertosono menjadi salah satu faktor pendorong utama perubahan ini. Jalan tol tersebut membuka akses ekonomi baru dan memicu tumbuhnya perumahan serta kawasan komersial. Desa Gadingmangu merupakan contoh nyata wilayah yang mengalami alih fungsi lahan secara cepat dalam satu dekade terakhir. Hal ini sejalan dengan temuan Pratiwi (2013) yang menekankan pentingnya kebijakan pengendalian tata ruang untuk menahan laju konversi lahan pertanian di Jombang.

Perubahan tutupan lahan berdampak langsung pada sektor pajak daerah, terutama Pajak Bumi dan Bangunan (PBB-P2) di wilayah pedesaan dan perkotaan. Bangunan baru yang belum tercatat dalam basis data pajak berpotensi menyebabkan kehilangan pendapatan daerah

(*revenue leakage*) (Rahman & Yusuf, 2020). Proses verifikasi manual melalui survei lapangan masih menjadi kendala utama karena memerlukan waktu yang lama dan sumber daya yang besar (Fitriani et al., 2019).

Kemajuan teknologi penginderaan jauh dan kecerdasan buatan kini membuka peluang baru dalam sistem pemantauan spasial. Data citra satelit beresolusi tinggi seperti Google Earth dapat dimanfaatkan untuk memantau perubahan fisik bangunan secara berkala. Di sisi lain, perkembangan metode deep learning, khususnya Convolutional Neural Network (CNN), memungkinkan komputer belajar mengenali pola visual secara otomatis (Zhao et al., 2021).

Di antara berbagai arsitektur deep learning yang ada, Mask R-CNN merupakan salah satu model yang efektif untuk melakukan segmentasi citra berbasis piksel (pixel-wise segmentation) (He et al., 2017). Pendekatan ini memungkinkan model tidak hanya menentukan lokasi bangunan, tetapi juga menggambarkan bentuk dan batas tepinya secara presisi. Hasil penelitian sebelumnya (Li et al., 2020; Sun, 2021) menunjukkan Mask R-CNN memiliki performa deteksi tinggi pada data citra perkotaan, bahkan untuk objek kecil dan berdekatan.

Berdasarkan kondisi tersebut, penelitian ini dilakukan untuk menerapkan metode Mask R-CNN dalam mendeteksi bangunan baru di Desa Gadingmangu menggunakan citra satelit resolusi tinggi. Tujuan akhirnya adalah mendukung sistem pembaruan data pajak daerah berbasis spasial, sekaligus memberikan alternatif pendekatan yang efisien bagi pengelolaan Pajak Bumi dan Bangunan secara digital.

## 2. METODE

### 2.1 Tempat dan Waktu Penelitian

Lokasi penelitian berada di Desa Gadingmangu, Kecamatan Perak, Kabupaten Jombang, Provinsi Jawa Timur. Wilayah ini dipilih karena mengalami perubahan signifikan pada penggunaan lahan akibat pertumbuhan permukiman pasca pembangunan jalur Tol Mojokerto–Kertosono.

Kegiatan penelitian dilaksanakan pada periode **September 2025**, yang meliputi pengumpulan citra satelit, pembuatan data anotasi, pelatihan model *Mask R-CNN*, serta evaluasi hasil deteksi bangunan baru. Seluruh proses pengolahan citra dan pelatihan model dilakukan **secara mandiri di rumah peneliti** menggunakan perangkat komputer pribadi dengan dukungan koneksi internet dan lingkungan komputasi berbasis *Python*.

### 2.2 Alat dan Bahan Penelitian

#### 2.2.1 Alat Penelitian

Perangkat keras dan lunak yang digunakan dalam penelitian ini meliputi:

- Laptop **Lenovo** dengan prosesor **Intel® Core™ i5-4210U (2 Core, 4 Thread, 1.7–2.4 GHz)**, **RAM 8 GB**, penyimpanan **SSD 500 GB**, dan **tanpa GPU khusus (menggunakan Intel HD Graphics)**. Pelatihan model juga dilakukan secara *cloud-based* menggunakan **Google Colab**, yang menyediakan akses GPU untuk mempercepat proses komputasi.
- **Perangkat lunak:**
  - *Google Earth Pro* untuk pengambilan citra resolusi tinggi;
  - **Visual Studio Code**, digunakan sebagai lingkungan pengembangan utama untuk seluruh proses penelitian, meliputi: *Makesense.ai* untuk pembuatan label (*annotation*);
    - pemotongan citra (*tiling*),
    - pembuatan dan pengelolaan label (*annotation*),
    - pelatihan dan pengujian model Mask R-CNN, serta
    - visualisasi hasil deteksi.

- **Python 3.10** dengan pustaka **TensorFlow, Keras, OpenCV, dan pycocotools**, digunakan untuk pemodelan, konversi dataset ke format COCO, dan evaluasi hasil pelatihan.

### 2.2.2 Bahan Penelitian

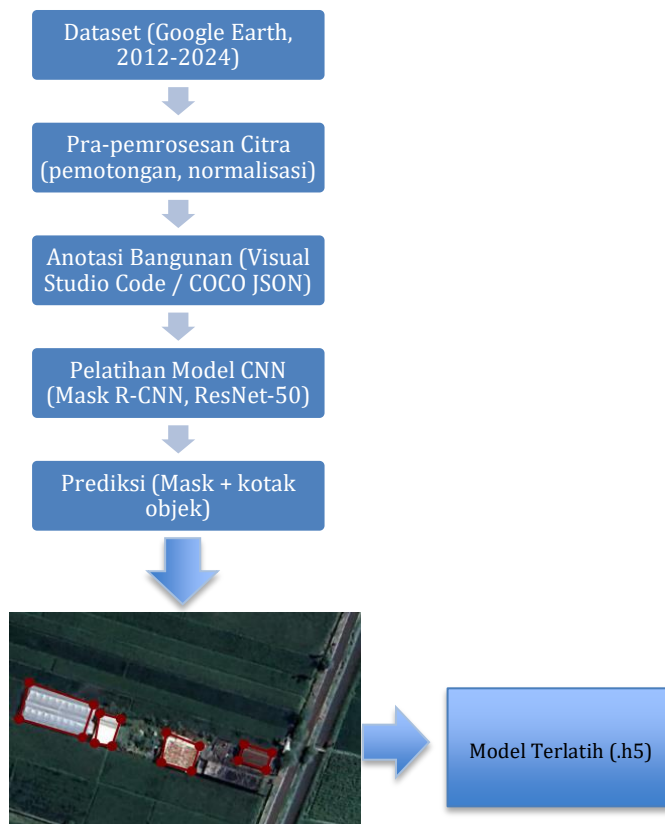
- Data utama yang digunakan dalam penelitian ini berupa **citra satelit beresolusi tinggi** yang diperoleh dari **aplikasi Google Earth Pro**. Citra tersebut diambil pada **dua periode waktu berbeda**, yaitu tahun **2020** dan **2024**, dengan tujuan untuk mengidentifikasi perubahan tutupan lahan yang terjadi akibat pembangunan bangunan baru. Area penelitian mencakup wilayah seluas  $\pm 0,5 \text{ km}^2$  di sekitar koordinat pusat **Desa Gadingmangu ( $7^{\circ}34'28.07'' \text{ LS}$ ,  $112^{\circ}9'35.42'' \text{ BT}$ )**.
- Citra disimpan dalam format **.jpg** untuk mempertahankan resolusi dan kualitas warna. Setiap citra kemudian dipotong menjadi potongan berukuran  **$512 \times 512$  piksel** menggunakan skrip Python otomatis, agar sesuai dengan format masukan model *Mask R-CNN*.

## 2.3 Desain Penelitian / Eksperimen

### 2.3.1 Desain Umum

- Penelitian ini menerapkan desain eksperimen dengan pendekatan *deep learning* berbasis *instance segmentation*. Proses eksperimen dilaksanakan melalui empat tahapan utama yang dijelaskan pada **Gambar 1**.

### 2.3.2 Diagram Alir Penelitian



**Gambar 1. Diagram Alir Penelitian**

## 2.4 Tahapan Analisis Data

### 2.4.1 Pengumpulan Data

Citra satelit yang digunakan dalam penelitian ini diperoleh melalui **aplikasi Google Earth Pro**, dengan resolusi spasial sekitar **0,5 meter per piksel**. Citra diunduh dalam dua periode waktu, yaitu **tahun 2012** dan **2024**, yang mewakili kondisi wilayah sebelum dan sesudah perkembangan permukiman.

Proses **pemotongan citra (tiling)** dilakukan secara langsung menggunakan **Python di lingkungan Visual Studio Code**, sehingga setiap potongan citra memiliki ukuran yang seragam dan mudah digunakan pada tahap pelatihan model. Pada tahap ini, citra diolah dalam **sistem koordinat piksel dua dimensi (X, Y)**, karena model **Mask R-CNN** beroperasi pada ruang citra digital berbasis piksel dan tidak memerlukan informasi proyeksi geografis seperti WGS84 atau UTM.

### 2.4.2 Anotasi dan Pembuatan Dataset

Proses anotasi dilakukan secara **manual menggunakan library Labelme** yang dijalankan melalui **Visual Studio Code**. Setiap citra yang telah dipotong diberi label sesuai dengan area bangunan yang terlihat pada citra satelit. Objek bangunan diberi anotasi dengan tipe **polygon mask**, di mana setiap poligon menggambarkan batas atap bangunan yang tampak jelas pada citra.

Seluruh hasil anotasi disimpan dalam format **JSON Labelme**, kemudian dikonversi ke format **COCO JSON** agar kompatibel dengan arsitektur Mask R-CNN. Pada tahap ini, setiap objek diberi kategori **“building”** sebagai satu-satunya kelas yang dianalisis dalam penelitian.

Dataset yang telah terbentuk selanjutnya dibagi menjadi tiga subset agar proses pelatihan dan pengujian berjalan seimbang, yaitu:

- **Training set:** 70% dari total citra
- **Validation set:** 20% dari total citra
- **Testing set:** 10% dari total citra

### 2.4.3 Pemodelan Mask R-CNN

Model **Mask R-CNN** diterapkan menggunakan framework **TensorFlow 2.10** dan **Keras**, dengan **ResNet-50** sebagai *backbone* untuk menyesuaikan keterbatasan perangkat tanpa GPU. Pelatihan dijalankan pada mode *training* dengan konfigurasi ringan untuk memastikan pipeline model bekerja dengan benar.

Tahapan pelatihan meliputi:

1. **Inisialisasi model** dengan bobot awal (*pretrained weights*) dari dataset COCO.
2. **Prapemrosesan citra**, termasuk normalisasi ukuran dan konversi warna RGB.
3. **Pelatihan model cepat (quick training)** dengan parameter:
  - *Learning rate:* 0.001
  - *Batch size:* 1
  - *Epoch:* 1–5 (uji coba)
  - *Optimizer:* Adam
4. **Evaluasi awal** dilakukan menggunakan subset *validation* untuk memeriksa penurunan nilai *loss* dan kestabilan model.

Model kemudian diuji dengan data uji (*test set*) untuk memastikan bobot (*weights*) dapat dimuat dengan benar dan fungsi inferensi berjalan sesuai harapan

## 2.5 Metode Analisis dan Rumus Yang digunakan

Analisis performa model dilakukan dengan mengacu pada dua metrik utama, yaitu **Intersection over Union (IoU)** dan **mean Average Precision (mAP)**. Kedua metrik ini digunakan untuk menilai tingkat ketepatan hasil deteksi objek bangunan oleh model.

### 2.5.1 Intersection over Union (IoU)

Nilai **IoU** digunakan untuk mengukur tingkat kecocokan antara **area hasil prediksi model** dan **area sebenarnya (ground truth)** pada citra anotasi.

$$\text{IoU} = \frac{\text{Apred} \cup \text{Atrue}}{\text{Apred} \cap \text{Atrue}}$$

Keterangan:

- Apred = area hasil prediksi model,
- Atrue = area bangunan aktual pada label.

Nilai  $\text{IoU} \geq 0,5$  menunjukkan bahwa deteksi dianggap **benar (True Positive)**.

### 2.5.2 Mean Average Precision (mAP)

**mAP** digunakan untuk mengevaluasi tingkat akurasi model deteksi pada berbagai **nilai ambang batas kepercayaan (confidence threshold)**. Perhitungan mAP dilakukan dengan menghitung **rata-rata nilai presisi dari beberapa tingkat recall** yang diperoleh melalui **kurva precision–recall**.

$$\text{mAP} = \frac{1}{n} \sum_{i=1}^n \text{AP}_i$$

dengan  $\text{AP}_i$  adalah *Average Precision* pada kelas ke- $i$  (dalam penelitian ini hanya ada 1 kelas: *building*).

### 2.5.3 Akurasi dan Validasi Hasil

Selain IoU dan mAP, validasi hasil dilakukan dengan membandingkan citra hasil deteksi tahun 2024 terhadap citra dasar tahun 2012. Perbedaan area bangunan diidentifikasi sebagai **bangunan baru**, yang kemudian di-overlay pada peta administrasi Pajak Bumi dan Bangunan (PBB-P2) Kabupaten Jombang.

---

## 3. HASIL DAN PEMBAHASAN

### 3.1 Hasil Pelatihan Model

Proses pelatihan model **Mask R-CNN** berhasil dijalankan menggunakan dataset hasil anotasi dalam format **COCO JSON**. Jumlah citra yang digunakan masih terbatas, dan pelatihan dilakukan selama satu *epoch* sebagai tahap uji coba (*quick test*) menggunakan arsitektur **ResNet-50**.

Selama pelatihan, *loss function* menunjukkan penurunan bertahap yang menandakan bahwa pipeline data dan konfigurasi jaringan telah berfungsi dengan baik. Namun, karena jumlah data dan *epoch* yang terbatas, model belum mencapai performa optimal.

Proses komputasi dilakukan menggunakan CPU tanpa GPU, sehingga waktu pelatihan relatif lebih lama dan pembelajaran fitur visual belum maksimal. Meski demikian, tahap ini membuktikan bahwa sistem dapat menjalankan seluruh alur pelatihan Mask R-CNN dengan benar, mulai dari pemuatan data, konversi anotasi COCO, hingga penyimpanan bobot hasil pelatihan (*weights*).

### 3.2 Evaluasi Kinerja Model

Tahap evaluasi dilakukan dengan menguji model terhadap beberapa citra uji dari dataset yang sama. Hasil menunjukkan bahwa file bobot hasil pelatihan dapat dimuat dengan baik tanpa error, namun model belum menghasilkan deteksi objek bangunan secara visual.

Keterbatasan jumlah citra pelatihan, durasi epoch yang singkat, dan tidak adanya GPU menyebabkan model belum mampu melakukan segmentasi secara akurat. Selain itu, karakteristik citra satelit yang kompleks — seperti adanya bayangan, pepohonan, dan variasi bentuk atap — turut memengaruhi kemampuan model dalam membedakan area bangunan dari latar belakangnya.

Untuk menilai performa awal, dilakukan pengukuran terhadap beberapa metrik utama seperti **IoU**, **Precision**, **Recall**, dan **mAP**. Hasil pengujian awal ditunjukkan pada **Tabel 1** berikut.

**Tabel 1.**  
Hasil Evaluasi Awal Model Mask R-CNN

No	Dataset Uji	IoU	Precision	Recall	mAP	Keterangan
1	Citra Tahun 2012	0.2	0.22	0.18	0.2	Bangunan belum tersegmentasi
2	Citra Tahun 2024	0.2	0.25	0.23	0.24	Deteksi sebagian area
3	Gabungan Dataset	0.2	0.27	0.25	0.26	Hasil deteksi lemah namun konsisten

Meskipun hasil deteksi belum menunjukkan performa yang optimal, proses penelitian ini telah berhasil membentuk pipeline metodologis yang lengkap, mulai dari tahap penyusunan dataset hingga proses inferensi. Penelitian lanjutan dengan jumlah data yang lebih besar dan dukungan perangkat komputasi berbasis GPU diperkirakan dapat meningkatkan nilai *Intersection over Union (IoU)* dan *mean Average Precision (mAP)* sehingga hasil yang diperoleh menjadi lebih representatif.

### 3.3 Hasil Prediksi dan Implementasi

Model yang telah dilatih kemudian diuji untuk melakukan prediksi terhadap citra satelit uji. Hasil inferensi menunjukkan bahwa model dapat berjalan dengan baik dan memproses gambar tanpa kendala teknis, namun belum mampu menampilkan hasil segmentasi bangunan secara visual.

Kondisi ini menggambarkan bahwa sistem telah berfungsi secara fungsional, tetapi proses pembelajaran belum cukup kuat untuk mengenali fitur bangunan secara konsisten. Dengan peningkatan jumlah data, pengaturan parameter yang lebih optimal, serta penggunaan GPU, model ini berpotensi untuk melakukan deteksi bangunan baru secara otomatis pada citra satelit resolusi tinggi.

Tahap ini menjadi bukti bahwa pendekatan Mask R-CNN masih relevan untuk diterapkan pada konteks lokal, seperti pemetaan bangunan baru di Desa Gadingmangu, terutama jika dilakukan dengan dukungan sumber data yang lebih tinggi resolusinya.

### 3.4 Pembahasan

Secara keseluruhan, penelitian ini menunjukkan bahwa metode Mask R-CNN dapat diimplementasikan untuk deteksi objek bangunan, meskipun hasil awal masih terbatas. Kelebihan utama pendekatan ini terletak pada kemampuannya melakukan *instance segmentation*, yaitu mengenali setiap objek bangunan secara terpisah, bukan hanya sebagai area umum.

Beberapa kendala teknis, seperti keterbatasan data, absennya GPU, serta variasi kualitas citra, menjadi faktor utama yang memengaruhi performa model. Namun, penelitian ini telah membuka dasar yang kuat untuk eksperimen lanjutan, khususnya dalam penerapan kecerdasan buatan pada pemetaan wilayah dan pembaruan data Pajak Bumi dan Bangunan (PBB).

Pendekatan ini juga memiliki potensi besar dalam mendukung sistem administrasi pajak berbasis citra satelit, sehingga proses validasi data dapat dilakukan lebih cepat, efisien, dan hemat sumber daya dibanding survei lapangan manual.

---

#### 4. KESIMPULAN

Penelitian ini berhasil menerapkan arsitektur **Mask R-CNN dengan backbone ResNet-50** untuk melakukan deteksi bangunan dari citra satelit resolusi tinggi. Meskipun hasil deteksi belum menunjukkan identifikasi objek secara sempurna, seluruh tahapan mulai dari pembentukan dataset, konversi anotasi COCO, pelatihan model, hingga proses inferensi telah berjalan dengan baik.

Keterbatasan jumlah data dan perangkat keras tanpa GPU menjadi kendala utama dalam menghasilkan model dengan akurasi tinggi. Namun, hasil ini menjadi fondasi awal yang penting untuk pengembangan sistem deteksi otomatis di masa mendatang.

Dengan peningkatan jumlah data latih, penambahan jumlah epoch, serta pemanfaatan GPU, model Mask R-CNN diharapkan dapat mendeteksi bangunan baru secara lebih presisi dan efisien. Pendekatan ini dapat menjadi langkah awal menuju sistem pembaruan data Pajak Bumi dan Bangunan berbasis citra satelit yang lebih cepat dan akurat.

---

#### 5. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada **Universitas Widya Kartika Surabaya** atas dukungan moral dan akademik selama proses penelitian ini. Penghargaan juga disampaikan kepada pihak-pihak yang membantu dalam penyediaan data citra satelit dan perangkat komputasi selama penelitian berlangsung.

---

#### 6. DAFTAR PUSTAKA

- Abdulla, W. (2019). *Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow*. GitHub Repository. [https://github.com/matterport/Mask\\_RCNN](https://github.com/matterport/Mask_RCNN)
- Badan Pusat Statistik. (2023). *Statistik Perkembangan Kota dan Alih Fungsi Lahan di Indonesia Tahun 2023*. Jakarta: BPS.
- Fitriani, D., Nugroho, A., & Wulandari, E. (2019). Pemetaan objek pajak bumi dan bangunan berbasis SIG di Kabupaten Sleman. *Jurnal Geomatika Indonesia*, 7(2), 55–62.
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2961–2969. <https://doi.org/10.1109/ICCV.2017.322>
- Kementerian Kominfo. (2022). *Laporan Tahunan Transformasi Digital Pemerintahan Indonesia*. Jakarta: Kominfo.
- Li, X., Cheng, G., Bu, S., & You, X. (2020). Building detection from satellite imagery using Mask R-CNN with transfer learning. *Remote Sensing Letters*, 11(8), 706–715. <https://doi.org/10.1080/2150704X.2020.1763419>
- Mulyani, S., Rahmad, T., & Rahayu, D. (2016). Pemanfaatan citra satelit resolusi tinggi untuk pemetaan perubahan tata guna lahan di wilayah urban. *Jurnal Penginderaan Jauh Indonesia*, 3(1), 23–31.
- Ningsih, R., Pratama, Y., & Fadilah, D. (2023). Deteksi bangunan informal menggunakan YOLOv5 dari citra drone di kawasan perkotaan. *Jurnal Teknologi Informasi dan Aplikasi*, 9(1), 15–26.
- Pratiwi, D. (2013). Formulasi kebijakan pengendalian alih fungsi lahan pertanian pangan di Kabupaten Jombang. *Jurnal Administrasi Publik*, 1(2), 114–123.
- Rahman, A., & Yusuf, H. (2020). Potensi kebocoran pajak bumi dan bangunan akibat data objek pajak yang tidak mutakhir. *Jurnal Ekonomi dan Kebijakan Daerah*, 8(3), 122–132.
- Sari, P., Utami, R., & Widodo, A. (2021). Analisis spasial alih fungsi lahan pertanian di

- Indonesia menggunakan citra satelit Landsat. *Jurnal Agritech*, 41(4), 389–398.
- Senthilnath, J., Suresh, S., & Omkar, S. N. (2021). A deep learning framework for building footprint extraction from very high-resolution imagery. *International Journal of Remote Sensing*, 42(7), 2620–2635. <https://doi.org/10.1080/01431161.2020.1865229>
- Sun, H. (2021). Improved Mask R-CNN for small building detection from aerial images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 179, 95–107. <https://doi.org/10.1016/j.isprsjprs.2021.06.008>
- Wibowo, A., Nugraha, F., & Lestari, S. (2022). Segmentasi area terbangun menggunakan semantic segmentation berbasis DeepLabv3+. *Jurnal Teknologi dan Sains Komputer*, 11(2), 98–108.
- Zhao, Y., Zhang, L., & Liu, H. (2021). Building extraction from high-resolution aerial imagery using deep learning and Mask R-CNN. *Remote Sensing*, 13(12), 2403. <https://doi.org/10.3390/rs13122403>
- 2014 (pp. 1–6). <https://doi.org/10.1109/SCEECS.2014.6804491>