El-Jughrafiyah
Mohammad Misbahuddin, dkk
Volume 05, Issue 02, Tahun 2025

Integrasi Deep Learning Mask R-CNN dalam Pemodelan 3D LOD1 Berbasis Citra Foto Udara UAV

Mohammad Misbahuddin[1], Septa Erik Prabawa[2], Slamet Riadi[3], Yahya Alfrid Koroh[4] dan Shilvy Choiriyatun Navisa[5]

[1,2] *Program Studi Teknik Geomatika, Universitas Dr. Soetomo, Surabaya*
[3,5] *Infomap Geo Survey, Indonesia*
[4] *Teknik Geodesi, Institut Teknologi Nasional Bandung, Indonesia*
e-mail: misbahhud21@gmail.com

ABSTRAK. Pemodelan 3D Level of Detail 1 (LOD1) merupakan representasi geometris dasar berupa bentuk prismatik dengan atap datar, yang umum digunakan dalam perencanaan wilayah dan manajemen aset karena efisiensi komputasi serta kesederhanaan strukturnya. Namun, proses pembentukan model ini kerap menghadapi kendala dalam aspek efisiensi waktu, akurasi spasial, serta skalabilitas, khususnya pada tahap segmentasi dan identifikasi objek dari data citra 2D. Penelitian ini bertujuan untuk mengimplementasikan algoritma deep learning Mask R-CNN dalam rangka meningkatkan efisiensi dan akurasi proses ekstraksi jejak bangunan (building footprint) dari citra udara beresolusi tinggi yang dihasilkan oleh wahana UAV. Mask R-CNN memiliki keunggulan dalam mendeteksi objek serta menghasilkan segmentasi berbasis pixel (pixel-wise mask), sehingga memungkinkan delineasi batas objek secara presisi. Alur kerja penelitian meliputi akuisisi data UAV, segmentasi citra secara otomatis menggunakan Mask R-CNN, dan rekonstruksi 3D berbasis Sistem Informasi Geografis (SIG). Studi kasus dilakukan di Kelurahan Jawa, Kota Samarinda, dengan resolusi spasial citra sebesar 0,023 meter, akurasi horizontal (CE90) 0,139 meter, dan vertikal (LE90) 0,282 meter. Tiga skenario pelatihan menunjukkan tingkat deteksi bangunan masing-masing sebesar 33%, 42%, dan 55% dengan nilai presisi 0,589; 0,746; dan 0,794. Skenario ketiga (80% data pelatihan) menghasilkan visualisasi model 3D paling mendekati hasil digitasi manual. Namun, skenario kedua (70% data pelatihan) direkomendasikan karena waktu pemrosesan tercepat (2 jam 27 menit 58 detik) dan kebutuhan penyimpanan terkecil (1,82 GB). Secara ekonomi, metode ini mengurangi biaya sebesar 13% dibandingkan digitasi manual, dengan total biaya Rp13.367.504.

Kata kunci: deep learning; LOD1; Mask R-CNN; Pemodelan 3D; UAV

PENDAHULUAN

Pemodelan 3D LOD1 (Level of Detail 1) merupakan salah satu pendekatan fundamental dalam rekonstruksi objek bangunan dan infrastruktur, yang menyediakan representasi geometris sederhana dengan bentuk atap datar dan struktur dasar (Templin, 2023). Pendekatan ini banyak digunakan dalam berbagai aplikasi, termasuk perencanaan kota, manajemen aset, dan pemantauan lingkungan, karena kemampuannya dalam menyajikan informasi spasial secara cepat dan efisien. Meskipun memiliki tingkat detail yang lebih rendah dibandingkan level LOD yang lebih tinggi, LOD1 tetap menjadi pilihan utama dalam skala besar karena biaya dan waktu pemrosesan yang lebih rendah, sambil tetap mempertahankan akurasi yang memadai untuk analisis makro (Templin, 2023; Nys dkk., 2020).

Seiring dengan perkembangan teknologi yang pesat, pemanfaatan UAV (Unmanned Aerial Vehicle) dalam pemodelan tiga dimensi (3D) mengalami kemajuan yang signifikan. UAV telah menjadi instrumen yang sangat efisien untuk akuisisi data fotogrametri berkualitas tinggi, yang selanjutnya dapat digunakan untuk menghasilkan model 3D dengan presisi dan akurasi yang tinggi

(Nex & Remondino, 2014; Jiang dkk., 2021). Didukung oleh perangkat lunak pengolahan citra yang canggih serta algoritma komputasi yang mutakhir, UAV telah memberikan kontribusi inovatif dalam bidang pemodelan 3D, mencakup berbagai aplikasi mulai dari arsitektur, perencanaan tata kota, hingga pemantauan lingkungan (Remondino, 2011). Namun, proses pembangunan model 3D seringkali menghadapi tantangan dalam hal efisiensi, akurasi, dan skalabilitas, terutama pada tahap segmentasi dan identifikasi objek dari data citra 2D. Selain itu, dibutuhkan waktu yang relatif lama dan sumberdaya manusia yang semakin banyak apabila data yang diolah semakin besar terutama pada proses digitasi bangunan yang dilakukan secara manual. Hasil digitasi yang dilakukan juga tergantung kepada keahlian operator yang melakukan interpretasi sehingga bersifat tidak konsisten untuk operator yang berbeda – beda (Kraff dkk., 2020). Untuk mengatasi hal ini, diperlukan pendekatan yang mampu mengotomatisasi proses segmentasi objek untuk melakukan ekstraksi bangunan secara cepat dengan cakupan wilayah yang luas (Anurogo dkk., 2017).

Algoritma deep learning, khususnya Mask Region-Based Convolutional Neural Network (Mask R-CNN), menawarkan solusi yang efektif untuk melakukan segmentasi dan klasifikasi objek pada citra foto udara (Abdulla, 2017; Danielczuk, 2019). Mask R-CNN merupakan pengembangan dari Faster R-CNN yang dilengkapi dengan cabang (branch) tambahan untuk melakukan segmentasi instance (instance segmentation) (Le dkk., 2018; Xavier dkk., 2022). Dengan memanfaatkan arsitektur ini, Mask R-CNN tidak hanya mampu mendeteksi objek dalam citra, tetapi juga menghasilkan masker pixel-level yang memungkinkan identifikasi batas objek secara presisi. Model Mask R-CNN banyak digunakan karena kesederhanaan model, fleksibilitas, dan juga kemampuannya dalam mendeteksi objek yang rapat (He dkk., 2018).

Penerapan algoritma Mask R-CNN dengan data UAV seperti LiDAR sudah pernah diterapkan oleh Jayaprakash (2024). Penelitian tersebut menerapakan segmentasi instan berbasis Convolutional Neural Network (CNN) yang ditingkatkan dengan menggabungkan arsitektur ResNet50 untuk mengekstrak fitur topografi lereng galian dari Digital Terrain Model (DTM) hasil LiDAR. Namun, penelitian ini masih sebatas integrasi Mask R-CNN dengan data UAV untuk pemetaan otomatis. Penelitian lain oleh Wu dkk (2025) yang melakukan deteksi objek berbasis Mask R-CNN dengan modifikasi ResNet Group Cascade (RGC) untuk mengatasi overfitting. Namun kelemahan metode ini jika diterapkan untuk segmentasi objek akan membutuhkan waktu yang lama dan kesalahan yang bertingkat karena prediksi awal dari suatu tahap digunakan dan ditingkatkan pada tahap berikutnya. Metode Mask R-CNN lain oleh Makungo dkk (2022) juga diterapkan untuk ekstraksi building footprint di area perumahan dan industri yang terstruktur. Hasil penelitian ini menunjukkan bahwa integrasi foto udara dan LiDAR efektif untuk segmetasi tapak bangunan di area perumahan dan industri, namun kurang efektif dalam mendeteksi tapak bangunan di area padat pemukiman yang tidak terstruktur. Selain itu pemetaan LiDAR membutuhkan biaya yang tinggi yang kurang efektif dalam cakupan wilayah relatif kecil seperti area perumahan terstruktur. Oleh karena itu, novelty penelitian ini yakni menerapakn algortima Mask R-CNN dengan ResNet backbone 101 yang efektif untuk pemukimanan padat yang tidak terstruktur pada foto udara UAV yang lebih minim biaya daripada data LiDAR. Selain itu penelitian ini juga membangun model 3D LOD1 dari hasil segmentasi otomatis Mask R-CNN yang menunjukkan model realistis sederhana. Melalui pendekatan ini, diharapkan model 3D LOD1 bangunan dapat dikembangkan secara untuk skala besar sehingga mendukung berbagai aplikasi seperti pemetaan kota, simulasi lingkungan, dan rekonstruksi bencana. Integrasi Mask R-CNN dalam alur kerja pembangunan model 3D diharapkan dapat mempercepat proses, mengurangi biaya, dan meminimalkan kesalahan manusia, sehingga memberikan kontribusi signifikan dalam bidang geospasial dan pemodelan 3D.

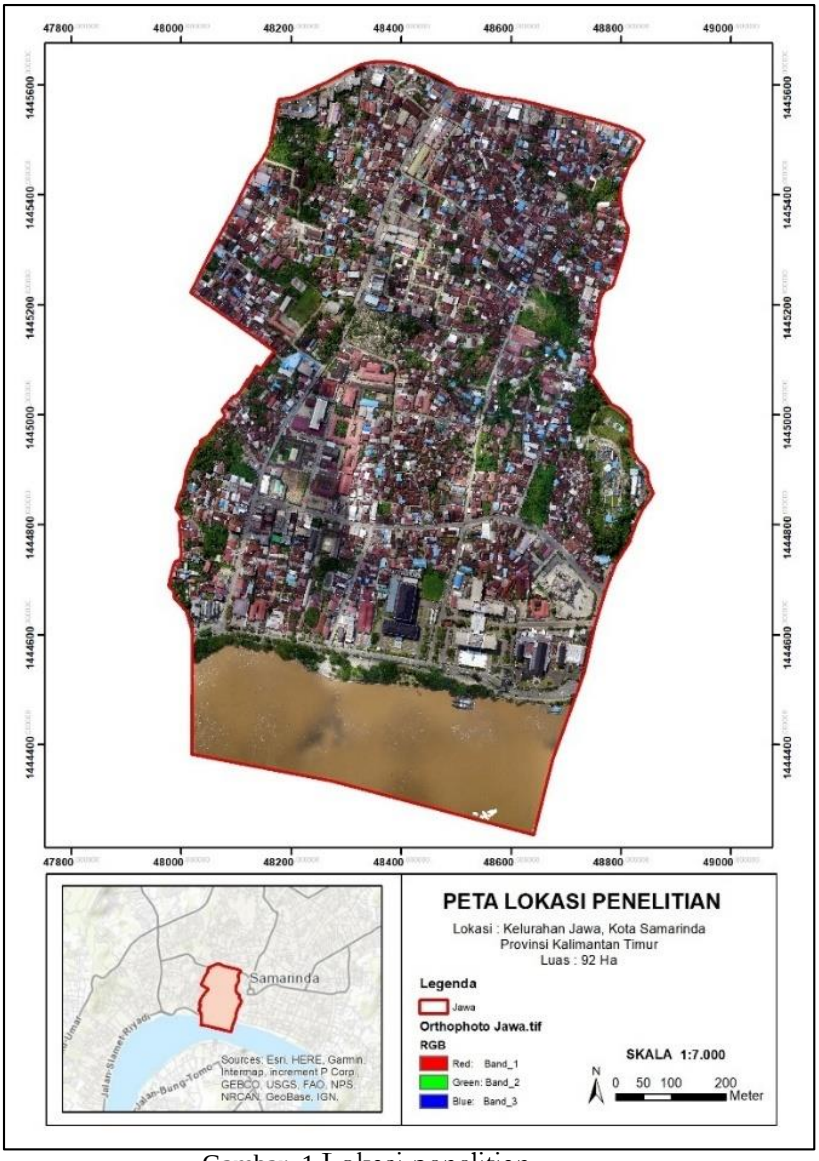
Sedangkan untuk konten pada bagian pendahuluan, harus memuat: *Pertama*, pemaparan topik utama penenelitian. *kedua*, memuat literatur terbaru terkait dengan mensitasi literatur penelitian terbaru terkait dengan artikel yang dikaji. *Ketiga*, harus menunjukkan kesenjangan yang belum terisi oleh penelitian/literatur, ketidakkonsistenan dan kotrovrsi yang muncul diantara literatur yang ada.

Keempat, memuat permasalahan, tujuan kajian, kontek kajian, dan unit analisis yang digunakan, dan *Kelima*, menampilkan apa yang dibahas dalam struktur artikel.

METODE

Data dan Lokasi

Penelitian ini mengambil lokasi studi yang terletak di Kota Samarinda tepatnya di Kelurahan Jawa seluas 92 Hektar. Kelurahan ini berbatasan dengan Kelurahan Dadi Mulya dan Kelurahan Sidodadi di bagian Utara, Kelurahan Bugis di bagian Timur, Sungai Mahakam di bagian Selatan, dan Kelurahan Teluk Lerong Ilir di bagian Barat. Kondisi geografis lokasi penelitian memiliki ketinggian tanah 500 meter dengan curah hujan 250 mm/tahun. Topografi lokasi penelitian termasuk ke dalam topografi rendah dengan suhu udara rata-rata 20°C – 30°C. Penelitian ini memakai study area di Kelurahan Jawa karena daerah ini merupakan daerah padat pemukiman yang tidak terstruktur dan berada di tengah pusat kota. Hal ini dibuktikan dengan jarak Kelurahan Jawa dari pusat Pemerintahan Kota yakni sejauh 3 km dan jarak dari pusat pemerintahan Kabupaten hanya sejauh 0.5 km. Pengambilan data foto udara dilakukan menggunakan wahana drone tipe fixed-wing. Illustrasi lokasi penelitian disajikan pada Gambar 1 berikut.



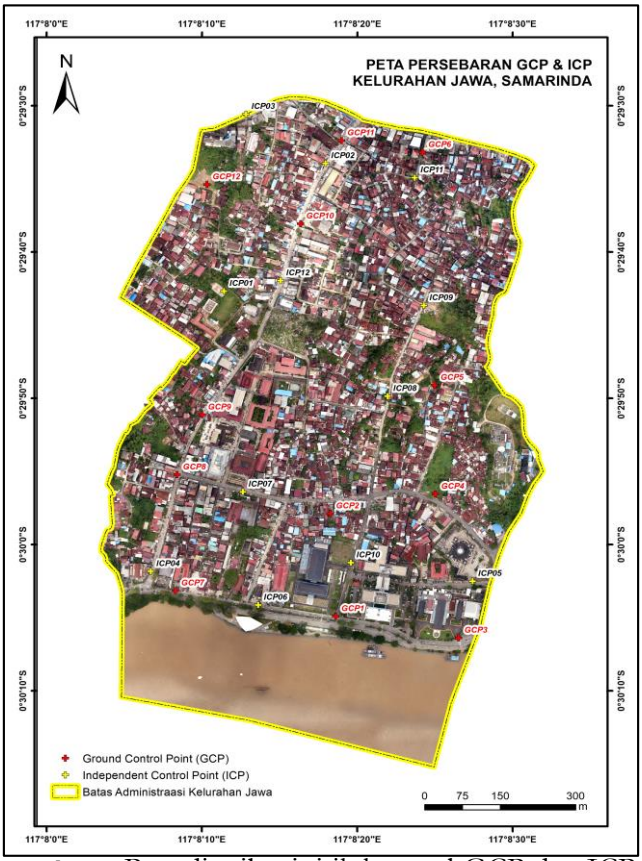
Gambar 1 Lokasi penelitian

Penelitian ini tidak perlu melakukan kalibrasi kamera sebelum pengukuran (pre-survey). Kalibrasi kamera dilakukan saat pengolahan data pada software Agisoft dengan hasil focal length sebesar 16 mm. Hasil kalibrasi kamera ini sudah memenuhi standar ideal focal length untuk fotogrametri vertikal (top-down) <24-35 mm. Spesifikasi data foto udara yang diakuisisi menggunakan survey foto udara UAV disajikan pada Tabel 1 berikut ini.

Tabel 1 Spesifikasi data foto udara

Spesifikasi	Nilai
GSD (*Ground Sampling Distance*)	6,46 cm/px
Luas citra foto udara	92 Hektar
Pertampalan ke muka (*Endlap*)	85%
Pertampalan ke samping (*Sidelap*)	70%
Tinggi terbang	± 400 meter
Sudut kamera	0°
Jumlah GCP & ICP	12 & 12
Referensi koordinat (XY & Z)	UTM 50S & Ellipsoid WGS84
Tanggal Akuisisi	1-3 November 2024

Akuisisi data titik kontrol GCP dan ICP menggunakan receiver GNSS geodetik. Pengamatan GCP dan ICP dilakukan tersebar merata ke seluruh area pengukuran dengan metode GNSS differential static–radial dengan waktu pengamatan per titik selama 1 jam. Penggunaan titik kontrol secara merata di area penelitian ini bertujuan untuk kontrol kualitas hasil orthophoto yang menyajikan nilai akurasi horisontal (CE90) dan vertikal (LE90) yang diperoleh dari nilai RMSE. Rumus perhitungan CE90 dan LE90 dijabarkan pada persamaan (1) dan (2) bab Metodologi. Gambar 2 di bawah ini merupakan peta sebaran GCP dan ICP pada akuisisi foto udara kelurahan Jawa.



Gambar 2 Peta distribusi titik kontrol GCP dan ICP

Metodologi

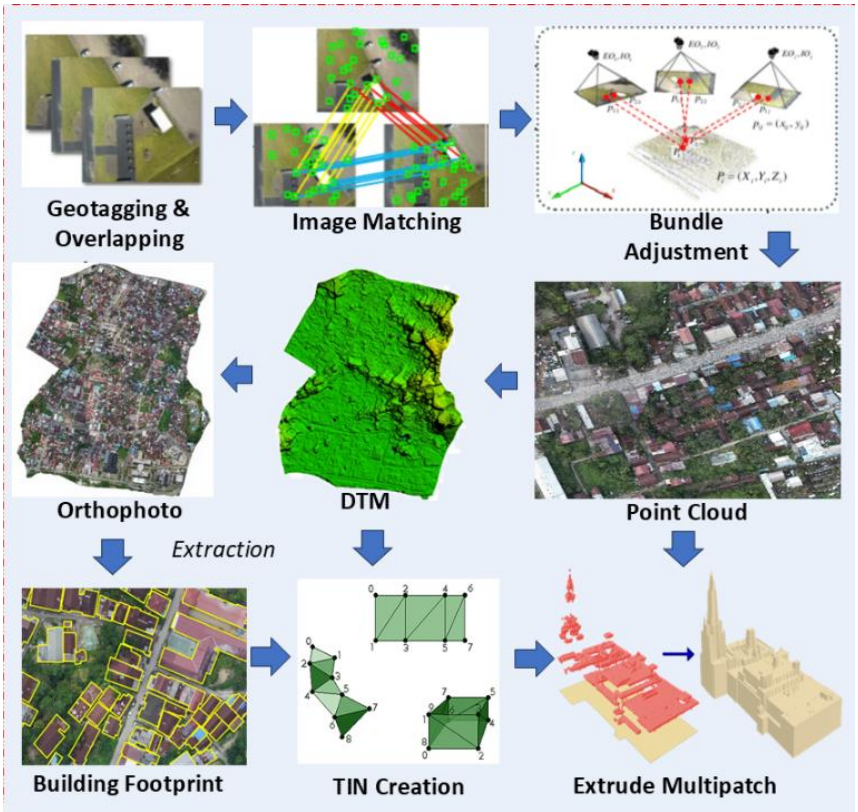
Metode pengambilan data dalam penelitian ini melibatkan pemotretan foto udara yang dilakukan dengan menggunakan drone tipe *fixed-wing* dan dilengkapi dengan kamera resolusi tinggi 61 MP untuk menangkap citra udara yang akan diproses menjadi model 3D dari area studi di Kelurahan Jawa, Kota Samarinda. Gambar wahana drone yang digunakan pada penelitian ini diilustrasikan pada Gambar 3.

Gambar 3 Foto unit drone *fixed-wing*
(Dokumentasi pribadi)

Hasil ukuran titik kontrol GCP dan ICP diproses menggunakan *software Trimble Business Center* (TBC) dengan metode statik radial. Namun sebelum itu, titik *Benchmark* (BM) yang digunakan untuk mengikat titik kontrol ini perlu diikatkan terlebih dahulu dengan CORS Samarinda (CSAM) menggunakan metode jaring. Setelah hasil BM *fixed* dengan standar eror fraksi milimeter, maka bisa diproses ke langkah selanjutnya untuk memperoleh koordinat GCP dan ICP. Validitas koordinat hasil GCP dan ICP yang bisa digunakan untuk proses SfM yakni koordinat yang sudah *fixed* (tidak *float*) dengan standar eror dibawah 1 cm (fraksi mm).

Metode pengolahan data dilakukan menggunakan teknik SfM. SfM adalah teknik dalam bidang fotogrametri dan visi komputer yang digunakan untuk merekonstruksi struktur tiga dimensi dari serangkaian gambar dua dimensi. Ini melibatkan pengambilan gambar dari berbagai sudut dan menggunakan informasi tersebut untuk membangun model 3D dari objek atau pemandangan (Ullman, 1979). Tahapan proses SfM pada penelitian ini yakni dimulai dari *georeferencing* foto udara untuk pengikatan foto udara terhadap titik kontrol atau *benchmark* di permukaan bumi dengan metode *Post Processing Kinematik* GNSS yang selanjutnya dilakukan proses *image matching* antara foto yang bertampalan untuk membangun *point clouds*. *Point cloud* diperoleh dari data *depth maps* dengan kualitas *high* dan *filtering mode mild* yang diproses selama 2 jam 6 menit. Proses selanjutnya yakni pembuatan DEM yang berasal dari data *point cloud* yang sudah diproses sebelumnya. Proses ini mengubah titik-titik individual menjadi model raster yang kontinu. Pembentukan DEM ini akan dilakukan proses untuk menghasilkan data DTM dan DSM yang selanjutnya digunakan sebagai parameter dalam pembuatan 3D *Building Multipatch*. Pembuatan DEM ini menggunakan kualitas *high* dan *filtering mode mild* dengan lama waktu pemrosesan 46 menit 46 detik menghasilkan ukuran data 11,54 GB. Selanjutnya yakni proses *orthophoto* untuk menghilangkan distorsi geometris yang disebabkan oleh *relief displacement*, posisi kamera, dan efek perspekstif. Data DEM yang diproses sebelumnya digunakan untuk menghilangkan efek distrosi sehingga menghasilkan ortofoto yang memiliki kesamaan geometris dengan peta. Proses *orthophoto* ini membutuhkan waktu 28 menit 15 detik dengan ukuran data sebesar 17,96 GB. Setelah *orthophoto* terbentuk, proses Mask R-CNN

diterapkan untuk *extract building footprint*. *Extract building footprint* adalah proses penting dalam analisis geospasial untuk mengidentifikasi bentuk dan batas bangunan dari data raster dalam hal ini ortofoto. Hasil *building footprint* akan dilakukan regularisasi (*regularize*) sehingga menghasilkan *footprint* bangunan yang tepat secara geometri. Yang terakhir yakni pembuatan model 3D bangunan dari *extrude multipatch* data DEM. Selisih data DTM dan DSM dilakukan untuk pembuatan *Normalize Digital Surface Model* (nDSM). nDSM ini yang digunakan untuk mendefinisikan tinggi bangunan dalam pembentukan 3D LOD1.



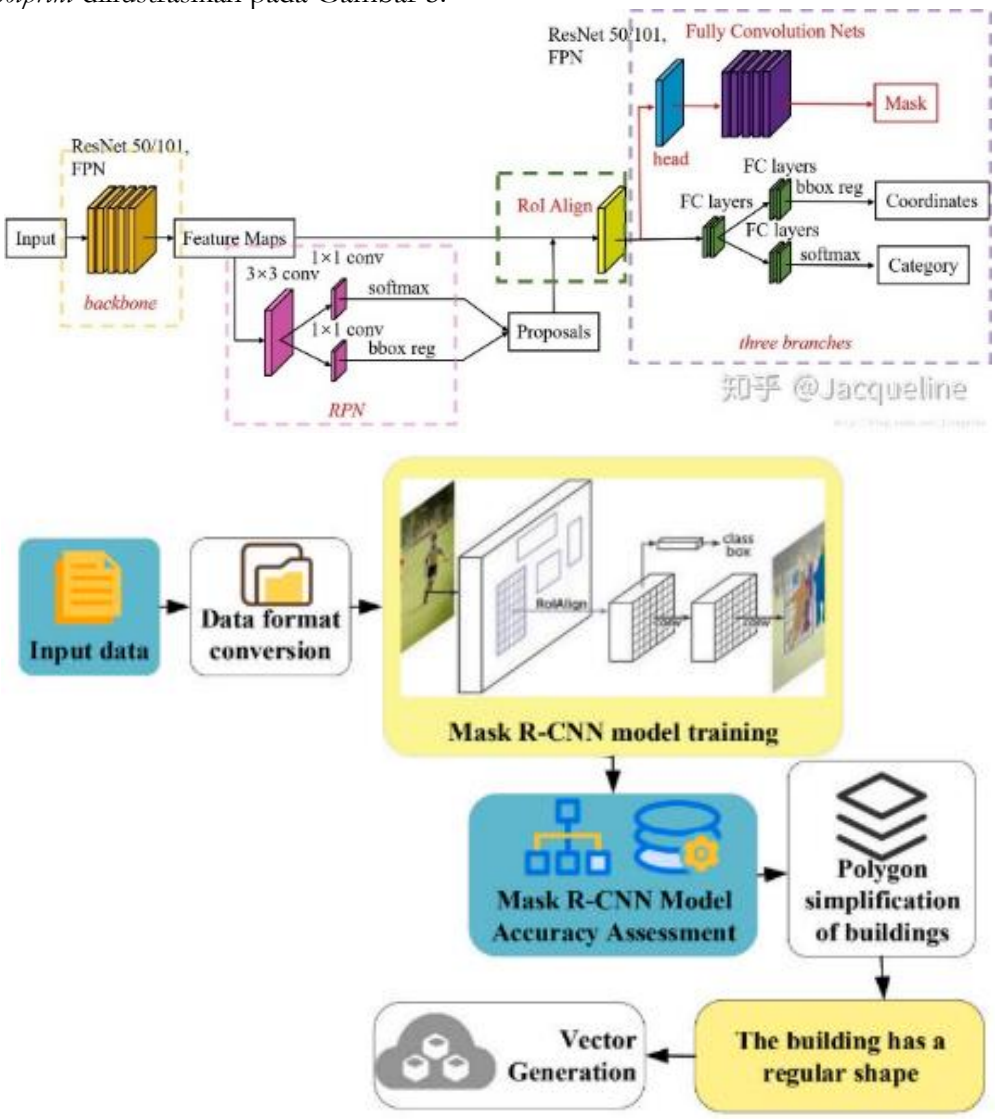
Gambar 4 Metodologi teknik SFM pada data foto udara

Salah satu tahap krusial dalam proses ini adalah *bundle adjustment*, yaitu teknik optimalisasi dalam fotogrametri yang bertujuan untuk menyempurnakan parameter kamera (posisi dan orientasi) serta titik-titik 3D hasil deteksi dari berbagai citra. Selama proses *bundle adjustment,* perlu dilakukan pengikatan sistem koordinat citra foto udara dengan titik GCP dan ICP yang biasa disebut dengan proses *premarking*. Hal ini berfungsi sebagai kendali eksternal untuk memverifikasi hasil *bundle adjustment*. Kemudian tahap selanjutnya yakni *optimize camera parameter* untuk mengontrol *reprojection error*. Hasil ini dikontrol dengan melihat nilai total *error* pada *check point*. Apabila nilai *check point* berada pada fraksi milimeter maka tahap *bundle adjustment* telah sukses dilakukan. Setelah tahap SfM selesai, dilakukan uji akurasi untuk mengevaluasi kualitas hasil pemotretan udara. Hasil pengolahan data berupa orthofoto harus memenuhi standar ketelitian peta sebagaimana diatur dalam Peraturan Kepala Badan Informasi Geospasial (BIG) Nomor 15 Tahun 2014 tentang ketelitian Peta Rupabumi Indonesia (RPI). Akurasi orthofoto dengan data ICP sebagai acuan diukur menggunakan nilai *Circular Error* 90% (CE90) untuk kesalahan horizontal dan *Linear Error* 90% (LE90) untuk kesalahan vertikal. Berdasarkan standar *United States National Map Accuracy Standards* (US NMAS), nilai CE90 yang diperbolehkan adalah kurang dari atau sama dengan 40 meter, sedangkan nilai LE90 adalah kurang dari atau sama dengan 2 meter. Rumus perhitungan kuarasi CE90 dan LE90 ditunjukkan pada persamaan berikut.

$$CE\ 90 = 1{,}5175 \times RMSEr \qquad (1)$$

$$LE\ 90 = 1{,}5499 \times RMSEz \qquad (2)$$

Orthofoto yang sudah memenuhi akurasi vertikal dan horizontal kemudian dilakukan proses *bulding footprint* dengan algoritma *deep learning* Mask-RCNN. Secara arsitektural, Mask R-CNN ini mengintegrasikan tiga komponen utama: *Region Proposal Network* (RPN) untuk mengidentifikasi *Region of Interest* (RoI), operasi konvolusi RoIAlign untuk ekstraksi fitur spasial, serta cabang paralel untuk klasifikasi objek dan generasi masker. Keunggulan utamanya terletak pada kemampuannya melakukan deteksi dan segmentasi secara simultan, presisi delineasi batas objek yang kompleks, serta adaptabilitasnya terhadap berbagai skala spasial. Algoritma penerapan Mask R-CNN dalam proses *building footprint* diilustrasikan pada Gambar 5.



Gambar 5 Penerapan Algoritma Mask R-CNN ke deteksi bangunan (Hou & Li, 2024)

Dalam penelitian ini, penulis menerapkan arsitektur Mask R-CNN dengan memanfaatkan *backbone Residual Network 101* (ResNet-101) yang diperkenalkan oleh He et al. (2017), sebagai alternatif dari *backbone* VGG-16. Modul residual pada ResNet dilengkapi dengan koneksi pintasan (*skip connection*), yang memungkinkan penggabungan aktivasi dari suatu lapisan ke lapisan lain yang terpisah sejauh 2-3 lompatan. Pada tahap segmentasi bangunan menggunakan Mask R-CNN dilakukan pada 3 simulasi percobaan yakni dengan persentase 50% training data, 70% training data, dan 80% training data. Pemilihan 3 skenario ini pada awalnya didasari oleh penelitian Makungo, dkk (2022) dalam mengevaluasi performa Mask-RCNN yang difokuskan untuk mendeteksi bentuk dasar bangunan (*footprint*) dari citra di Kota Cape Town, Afrika Selatan yang memiliki dua jenis kawasan dengan training data 80% dan validasi 20%. Kemudian untuk menguji peforma dan efektivitas

algoritma yang diterapkan, maka dilakukan augmentasi data dengan mengubah variasi persentasi training data 30% dan 70%. Dengan augmentasi data tersebut penulis ingin membuktikan apakah dengan training data < 80% lebih efektif untuk melakukan segmentasi bangunan. Chip gambar dan *mask* bangunan yang dihasilkan melalui tools "*Export Training Data for Deep Learning*" digunakan untuk melatih model Mask R-CNN dalam melakukan segmentasi objek (*instance segmentation*). Tiga model tersebut dilatih secara terpisah dengan data input berupa citra *orthophoto*. Proses pelatihan dilakukan menggunakan *ArcGIS API for Python di Jupyter Notebook* dengan memanfaatkan arsitektur ResNet101 yang telah dilatih sebelumnya (*pre-trained*) sebagai *backbone*.

Sampel validasi yakni data digitasi manual berfungsi untuk memantau konvergensi model dengan melihat penurunan nilai *loss* di akhir setiap *epoch*. Selama pelatihan, *batch size* 4 digunakan untuk semua model. *Learning rate* merupakan parameter kritis dalam pelatihan model *deep learning*. Jika terlalu tinggi, model dapat konvergen ke solusi yang tidak optimal, sedangkan jika terlalu rendah, proses pelatihan menjadi lambat (Hafidz, 2018). Model Mask R-CNN diimplementasikan menggunakan fungsi MaskRCNN dari *package deep learning* ArcGIS Pro. Pelatihan dilakukan pada GPU NVIDIA GeForce RTX 4080. Setiap model dilatih selama 30 *epoch* menggunakan data pelatihan dan validasi/testing.

Tabel 2 Spesifikasi pelatihan Mask R-CNN

Spesifikasi Deep Learning Mask R-CNN	**Nilai**
Data raster pelatihan	Orthofoto Kelurahan Jawa
Simulasi 1	Persentase training : 50% (775 fitur bangunan) Persentase validasi : 50% (775 fitur bangunan)
Simulasi 2	Persentase training : 70% (1085 fitur bangunan) Persentase validasi : 30% (465 fitur bangunan)
Simulasi 3	Persentase training : 80% (1241 fitur bangunan) Persentase validasi : 20% (310 fitur bangunan)
Epok	30
Batch	4
Backbone	ResNet-101
Kartu Grafis Komputer	NVIDIA GeForce RTX 4080

HASIL DAN PEMBAHASAN

Hasil Uji Akurasi Foto Udara UAV

Proses SfM diawali dengan alignment foto untuk membangun sparse point cloud, dilanjutkan dengan pembuatan dense point cloud, Digital Elevation Model (DEM) yang terdiri dari *Digital Surface Model* (DSM) dan *Digital Terrain Model* (DTM) dan akhirnya orthophoto dengan resolusi spasial atau GSD sebesar 0,02312 meter. Hal ini menunjukkan bahwa GSD foto udara telah memenuhi standar toleransi resolusi spasial yakni ≤ 0,12 m. Kualitas geometrik hasil olahan telah divalidasi menggunakan titik uji (ICP) sebanyak 12 titik yang tersebar merata di seluruh area studi. Dibawah ini merupakan tabel hasil perhitungan uji akurasi pada hasil pemetaan foto udara UAV (Tabel 3). Sistem koordinat yang digunakan pada pengukuran dan pengolahan ICP adalah Sistem Koordinat Proyeksi UTM 50S.

Tabel 3 Uji akurasi horizontal dan vertikal pada hasil pengukuran foto udara

Point	**Data ICP Lapangan**		
	Easting (m)	**Northing (m)**	**Elevation (m)**
ICP 01	515189,24	9945266,73	59,91
ICP 02	515391,17	9945534,46	61,72

Point	Data ICP Lapangan		
	Easting (m)	Northing (m)	Elevation (m)
ICP 03	515232,79	9945637,96	71,34
ICP 04	515043,28	9944678,35	57,3
ICP 05	515683,91	9944657,7	57,39
ICP 06	515257,15	9944607,79	56,73
ICP 07	515226,81	9944845,38	57,47
ICP 08	515514,59	9945045,82	67
ICP 09	515586,6	9945236,3	62,65
ICP 10	515441,37	9944696,55	57,3
ICP 11	515568,55	9945504,44	62,59
ICP 12	515301,84	9945288,59	60,63
Point	**Orthophoto – Foto Udara UAV**		**DEM Elevation (m)**
	Easting (m)	**Northing (m)**	
ICP 01	515189,19	9945266,69	59,74
ICP 02	515391,24	9945534,47	61,71
ICP 03	515232,89	9945637,96	71,35
ICP 04	515043,24	9944678,31	57,37
ICP 05	515683,95	9944657,78	57,5
ICP 06	515257,18	9944607,74	57,15
ICP 07	515226,81	9944845,26	57,56
ICP 08	515514,56	9945045,79	67,21
ICP 09	515586,77	9945236,22	62,8
ICP 10	515441,39	9944696,52	57,5
ICP 11	515568,52	9945504,36	62,47
ICP 12	515301,8	9945288,67	60,54
Akurasi Horizontal (CE90) = 0,139 m			
Akurasi Vertikal (LE90) = 0,282 m			

Nilai ketelitian horizontal yang dihitung berdasarkan CE90 mencapai 0,139 meter, mengkonfirmasi kecermatan posisi horizontal pada ortofoto yang sudah sesuai dengan kondisi aktual di lapangan. Sementara itu, ketelitian vertikal LE90 sebesar 0,282 meter menunjukkan kesesuaian tinggi antara elevasi pada DEM dengan pengukuran lapangan memiliki selisih yang tidak signifikan. Tingkat akurasi yang diperoleh membuktikan bahwa produk pemetaan foto udara ini memiliki reliabilitas tinggi baik dalam komponen horizontal maupun vertikal.

Perfoma Segmentasi Bangunan Menggunakan Mask R-CNN

Pengujian dilakukan di kelurahan Jawa, Kota Samarinda dengan luas wilayah 92 hektar. Pada kawasan pemukiman, model Mask-RCNN menunjukkan performa yang baik dengan mampu mendeteksi bangunan meskipun memiliki variasi bahan atap, bentuk, ukuran, dan tinggi yang beragam. Hasil ini menunjukkan keunggulan metode ini dalam memproses data skala besar secara efisien dibandingkan metode digitasi manual yang memakan waktu. Peforma Mask R-CNN dapat ditunjukkan secara statistik melalui penyajian grafik loss. Grafik loss yang menggambarkan proses pelatihan model Mask R-CNN untuk ekstraksi tapak bangunan terhadap dataset dapat dilihat pada Gambar 6. Selain model grafik, nilai evaluasi metrik juga ditunjukkan pada tabel 4, 5, dan 6 di bawah ini untuk mengetahui peformas Mask R-CNN secara statistik.

Tabel 4 Metrik evaluasi peforma Mask R-CNN IoU 0,5

Assessment IoU = 0,5	**Simulasi I : 30%**	**Simulasi II : 70%**	**Simulasi III : 80%**
Precision	0,206	0,346	0,385
Recall	0,071	0,076	0,099
F1 Score	0,106	0,124	0,157
AP	2,891	4,549	3,892
True Positive (TP)	150	160	208
False Positive (FP)	577	302	332
False Negative (FN)	1952	1942	1894

Tabel 5 Metrik evaluasi peforma Mask R-CNN IoU 0,75

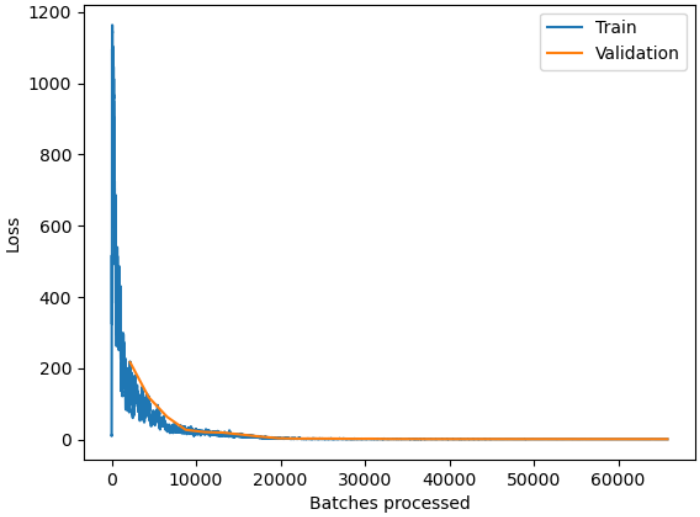
Assessment IoU = 0,75	**Simulasi I : 30%**	**Simulasi II : 70%**	**Simulasi III : 80%**
Precision	0,067	0,212	0,248
Recall	0,023	0,046	0,063
F1 Score	0,034	0,076	0,101
AP	2,891	4,549	3,892
True Positive (TP)	49	98	134
False Positive (FP)	678	364	406
False Negative (FN)	2053	2004	1968

Tabel 6 Metrik evaluasi peforma Mask R-CNN IoU 0,95

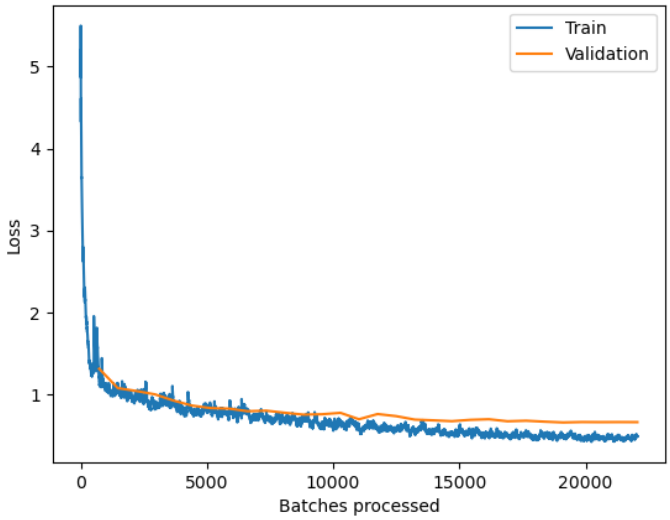
Assessment IoU = 0,95	**Simulasi I : 30%**	**Simulasi II : 70%**	**Simulasi III : 80%**
Precision	0	0	0,0037
Recall	0	0	0,001
F1 Score	0	0	0,0015
AP	0	0	3,892
True Positive (TP)	0	0	2
False Positive (FP)	727	462	538
False Negative (FN)	2102	2102	2100

Berdasarkan nilai statistik pada Tabel 4 dengan Intesection of Union (IoU) sebesar 0,5 atau sebesar 50% menunjukkan bahwa model training data pada ketiga simulai menunjukkan perbedaan nilai yang cukup signifikan. Terutama pada simulai III dengan 80% data tarining yang menunjukkan nilai presisi, recall, F1, TP, dan FP paling tinggi yakni 0,385; 0,099; 0,157; 208; dan 332. Nilai FN pada simulasi III paling rendah karena hal ini menunjukkan seberapa salah prediksi yang dimodelkan dengan nilai 1894. Selanjutnya pada Tabel 5 dengan IoU 0,75 juga masih menunjukkan nilai yang signifikan dimana pola asesmen sama dengan IoU 0,5 yakni simulasi 80% memiliki nilai presisi, recall, F1, TP, dan FP paling tinggi berturut-turut yakni 0,248; 0,063; 0,101; 134; 406. Kemudian diikuti nilai FN paling rendah yakni 1968. Nilai AP pada IoU 0,5 dan 0,75 memiliki hasil yang sama dimana simulasi II 70% paling tinggi yakni 4,549. Yang terakhir yakni pada IoU 0,95 dimana hasil precision, recall, F1 score, AP, TP pada simulai I dan II bernilai 0. Hal ini disebakan karena semakin tinggi persentase tumpang tindih antara prediksi dan *ground truth* (IoU) yakni bernilai 95% maka area tumpang tindih prediksi dan *ground truth* sangat akurat. Hasil ini menunjukkan bahwa hanya simulasi III dengan data training 80% yang mampu melakukan asesmen hingga 95% secara perlahan ditunjukkan dengan gafik warna biru yang masih tinggi di awal kemudian baru turun di akhir. Sedangkan *validation loss* yakni grafik warna kuning pada batch 10.000 masih mengalami *overfitting*. *Overfitting* yakni keadaan di mana model *deep learning* terlalu kompleks sehingga pre-trained pola data

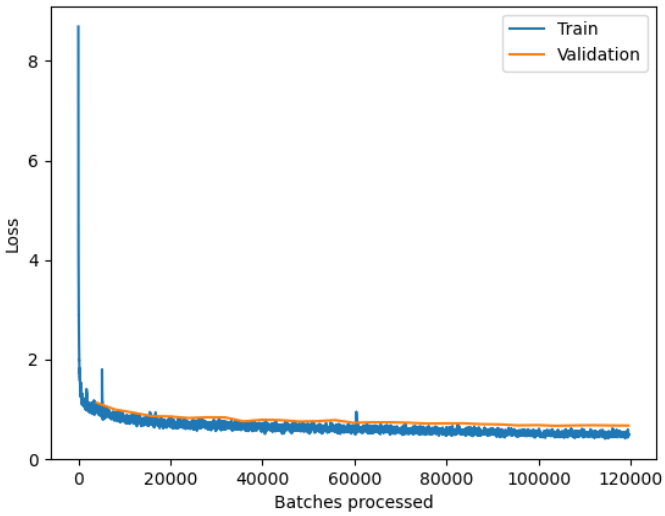
pelatihan secara berlebihan yang mengakibatkan model menunjukkan kinerja baik pada data pelatihan tetapi gagal melakukan generalisasi dengan baik pada data baru atau data validasi. Hal ini dapat disebabkan karena data pelatihan terlalu sedikit dibandingkan kompleksitas model yakni hanya 50% dari data keseluruhan. Terlalu sedikitnya data pelatihan ini juga menyebabkan pelatihan terlalu lama dan menimbulkan *overfitting*. Selanjutnya untuk simulasi II grafik *loss training* mulai turun dengan cepat di awal, lalu mendekati nilai rendah, sementara *validation loss* masih ada *overfitting* di awal. Simulasi III menunjukkan model grafik terbaik yakni *training loss* turun dengan cepat di awal dan *validation loss* stabil rendah yang mengindikasikan bahwa model dapat bekerja dengan baik pada data baru tanpa mengalami *overfitting*. Nilai presisi (*Average Precision Score*) untuk simulasi I, II, dan III secara berutut-turut adalah sebesar 0,589; 0,746; dan 0,794 menunjukkan bahwa model data training ≤50% belum mampu dilakukan untuk proses *deep learning*.



Simulasi I : 50% Training – 50% Validasi



Simulasi II : 70% Training – 30% Validasi



Simulasi III : 80% Training – 20% Validasi

Gambar 6 Grafik training loss vs validation loss pelatihan pada 3 model simulasi Mask R-CNN

Visualisasi Hasil Ekstraksi Building Footprint

Perbandingan hasil ekstraksi tapak bangunan secara visual terdapat sedikit perbedaan antara ketiga metode simulasi setelah dilakukan regularisasi. Ekstraksi tapak bangunan dari simulasi III cenderung tidak memiliki gap di tengah-tengah bangunan. Selain itu, overlap antar bangunan menjadi lebih minim pada ekstraksi bangunan dari simulasi III. Kedua hal tersebut dikarenakan adanya pemrosesan klasifikasi bangunan pada point cloud yang dihitung berdasarkan elevasi point cloud, serta dilakukan perhitungan parameter kesamaan ketinggian pada luas tertentu dan sudut tertentu. Sehingga gap maupun overlap dapat diminimalisir. Gambar 7 di bawah ini merupakan salah satu contoh yang menunjukkan adanya overlap dari bangunan yang dihasilkan oleh simulasi I dan II. Sementara pada simulasi III tidak terdapat adanya overlap di lokasi tersebut.

Simulasi I

Simulasi II

Simulasi III

Gambar 7 Cuplikan hasil ekstraksi bangunan dari data digitasi manual (kiri) dan segmentasi otomatis Mask-RCNN (kanan)

Secara kuantitatif, hasil ekstraksi tapak bangunan dari foto udara UAV segmentasi Mask RCNN diabndingkan dengan hasil dari digitasi manual. Ukuran yang menjadi parameter perbandingan adalah jumlah bangunan terdeteksi dan selisih luas antara hasil ekstraksi dengan hasil digitasi. Tabel 7 di bawah ini merupakan jumlah bangunan terdeteksi dan bangunan tidak terdeteksi dari hasil ekstraksi, beserta persentasenya terhadap data digitasi manual.

Tabel 7 Persentase bangunan terdeteksi terhadap hasil digitasi

Sumber Tapak Bangunan	Jumlah Bangunan Terdeteksi	Jumlah Bangunan Tidak Terdeteksi	Persentase Terhadap Digitasi Manual
Digitasi Manual	1551 (referensi)	-	-
Mask R-CNN Simulasi I	507	1044	33%
Mask R-CNN Simulasi II	653	898	42%
Mask R-CNN Simulasi III	855	696	55%

Hal ini dapat dilihat bahwa hasil ekstraksi bangunan terdeteksi pada foto simulasi III paling besar yakni 55% sedangkan presentase simulasi I hanya 33% disusul dengan simulasi II sebesar 42%. Banyaknya jumlah bangunan yang tidak terdeteksi ini dikarenakan banyak bangunan rapat dalam satu wilayah yang dideteksi menjadi satu objek oleh Mask-RCNN. Di sisi lain, hasil ekstraksi dari data simulasi III menunjukkan persentase >50% bangunan terdeteksi dibandingkan data simulasi I dan II. Dari segi optimalisasi dan efisiensi, perlu adanya pengecekan bangunan yang berhimpit pada citra agar deteksi objek sesuai dengan realisasinya. Perbandingan kuantitatif tidak hanya dilakukan dengan menghitung persentase bangunan terdeteksi, tetapi juga pada selisih luas yang dihasilkan (*mean difference*). Dengan menghitung *mean difference*, dapat diketahui seberapa besar perbedaan yang ada antara data yang tercatat dan data hasil digitasi foto udara. Tabel 8 menunjukkan perbandingan selisih luas antara hasil ekstraksi terhadap data digitasi manual.

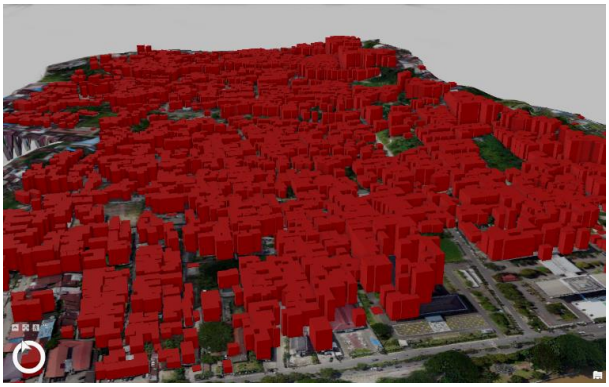
Tabel 8 Perbandingan selisih luas bangunan hasil ekstraksi dengan data digitasi manual

Luas Tapak Bangunan pada Digitasi Manual = 358145 m^2			
Sumber Tapak Bangunan	**Luas Total Bangunan (m^2)**	**Selisih Luas Total (m^2)**	**Rata-rata Selisih Luas Bangunan (m^2)**
Mask R-CNN Simulasi I	351074,607	7070,393	538,295
Mask R-CNN Simulasi II	352067,773	6077,227	436,633
Mask R-CNN Simulasi III	355262,391	2882,609	111,081

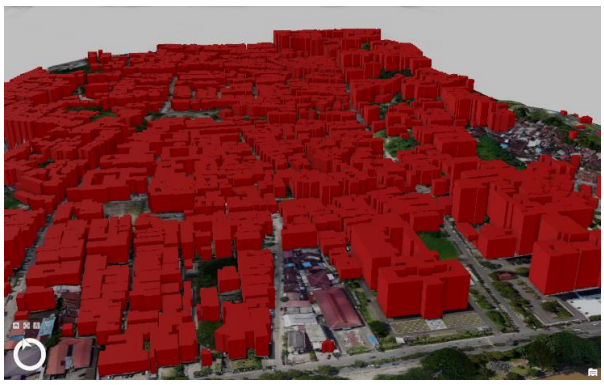
Luas total tapak bangunan berdasarkan digitasi manual adalah 358145 m². Simulasi I menghasilkan luas total bangunan sebesar 351074,607 m², dengan selisih luas total 7070,393 m² dan rata-rata selisih 538,295 m². Simulasi II menunjukkan hasil yang mirip, dengan luas total bangunan 352067,773 m², selisih total 6077,227 m², dan rata-rata selisih 436,633 m². Sementara itu, hasil ekstraksi bangunan dari simulasi III menghasilkan luas total yang paling mendekati dengan digitasi manual yaitu 355262,391 m², dengan selisih 2882,609 m² dan rata-rata selisih paling kecil sebesar 111,081 m². Analisis ini menunjukkan bahwa selisih yang cukup besar antara hasil ekstraksi dan digitasi manual mengindikasikan adanya perbedaan signifikan dalam akurasi pengukuran. Semakin banyak data training seperti pada simulasi III akan memberikan hasil yang lebih mendekati digitasi manual dibandingkan dengan Simulasi I dan II.

Hasil Model 3D LOD1

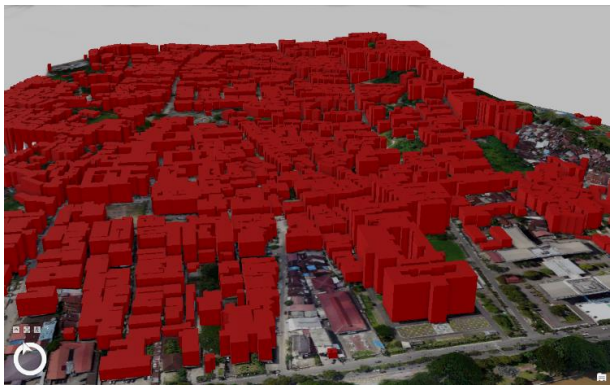
Pemodelan 3D yang dihasilkan dari elevasi ketinggian NDSM foto udara UAV dan tapak bangunan hasil Mask R-CNN memberikan representasi bangunan yang cukup akurat dalam bentuk *building feature*. Teknik ini memungkinkan penggambaran tapak bangunan dengan baik, meskipun ada beberapa keterbatasan dalam menangkap detail sudut dan sisi bangunan. Perspektif tegak di beberapa titik tidak mampu menangkap elemen yang tersembunyi atau terhalang vegetasi sehingga beberapa fitur bangunan tidak terdefinisi dengan sempurna. Namun, pendekatan ini tetap efektif untuk mendapatkan gambaran umum dari struktur bangunan, terutama ketika digunakan dalam kombinasi dengan data lain untuk meningkatkan akurasi. Hasil 3D LOD1 *building feature* dari elevasi NDSM foto udara UAV untuk ketiga simulasi dan perbandingannya dengan digitasi manual ditunjukkan pada Gambar 8-11.



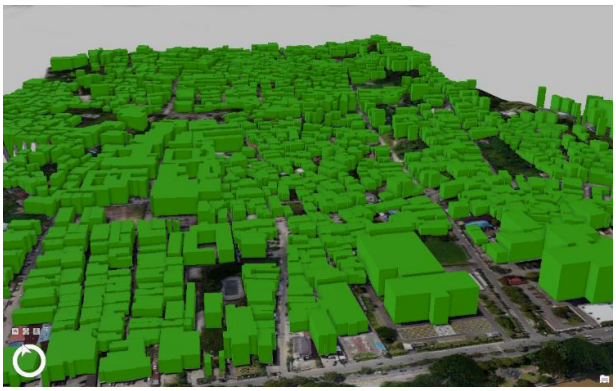
Gambar 8 Model 3D Segmentasi Semi Otomatis Mask-RCNN Simulasi I



Gambar 9 Model 3D Segmentasi Semi Otomatis Mask-RCNN Simulasi II



Gambar 10 Model 3D Segmentasi Semi Otomatis Mask-RCNN Simulasi III



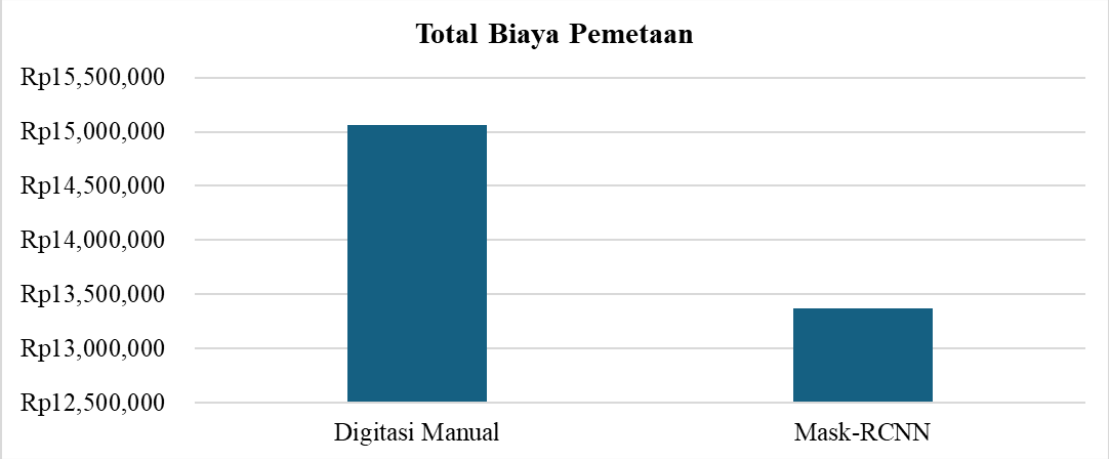
Gambar 11 Model 3D Digitasi Manual

Building footprint dengan hasil otomatisasi Mask R-CNN pada foto udara UAV memiliki visualisasi yang yang berbeda antara ketiga metode simulasi terhadap model 3D digitasi manual.

Footpritnt atau tapak bangunan pada metode Mask-RCNN lebih kecil dari bangunan sebenarnya. Hal ini disebabkan karena terdapat bangunan yang tidak terdeteksi atau tertutup vegetasi yang menyababkan noise. 3D building features jika mengikuti batas footprint otomatis akan menghasilkan model yang tidak beraturan dan kurang sesuai. Meskipun begitu, metode Mask R-CNN Simulasi III dengan 80% data training bisa digunakan untuk mass 3D model skala besar yang membutuhkan waktu cepat. Oleh karena itu penelitian ini menunjukkan bahwa Mask R-CNN dengan 80% data training cukup efektif untuk pembuatan model 3D.

Perbandingan Efisiensi Waktu dan Biaya

Pemodelan Tabel biaya pemrosesan ini menyajikan perbandingan menyeluruh mengenai estimasi anggaran yang diperlukan antara metode otomatis menggunakan Mask-RCNN dengan digitasi manual. Perhitungan biaya pemetaan dalam analisis ini merujuk pada Keputusan Kepala BIG Nomor 3 Tahun 2023 tentang Standar Biaya Penyelenggaraan Informasi Geospasial di Badan Informasi Geospasial Tahun 2023.



Gambar 12 Grafik perbandingan selisih biaya pemetaan antara ketiga metode

Berdasarkan perhitungan yang dilakukan, survei foto udara menggunakan UAV dengan metode Mask-RCNN memerlukan total biaya sebesar Rp13.367.504. Sementara itu, survei dengan metode konvensional melalui digitasi manual membutuhkan anggaran yang lebih tinggi, yaitu Rp15.064.172. Komponen biaya dalam kedua metode tersebut meliputi penyewaan UAV tipe VTOL, upah tenaga kerja (termasuk pilot dan surveyor), serta lisensi perangkat lunak untuk pengolahan data GNSS dan Sistem Informasi Geografis (SIG). Untuk rincian lebih lengkap mengenai perhitungan biaya tersebut, dapat dilihat pada Tabel 9 dan 10.

Tabel 9. Biaya pemetaan Digitasi Manual

No	Uraian	Nilai	Satuan		Waktu	Jumlah
1	Pilot	Rp.1.500.000	1	org	1	Rp. 1.500.000
2	Surveyor (Tenaga Terampil Jenjang 4)	Rp. 408.334	2	org	2	Rp. 1.633.336
3	Operator Pengolah Data (Tenaga Terampil Jenjang 5)	Rp. 473.334	1	org	4	Rp. 1.893.336
4	Sewa UAV (GNSS, PPK, Sensor Foto Udara)	Rp.2.000.000	1	Unit	1	Rp. 2.000.000

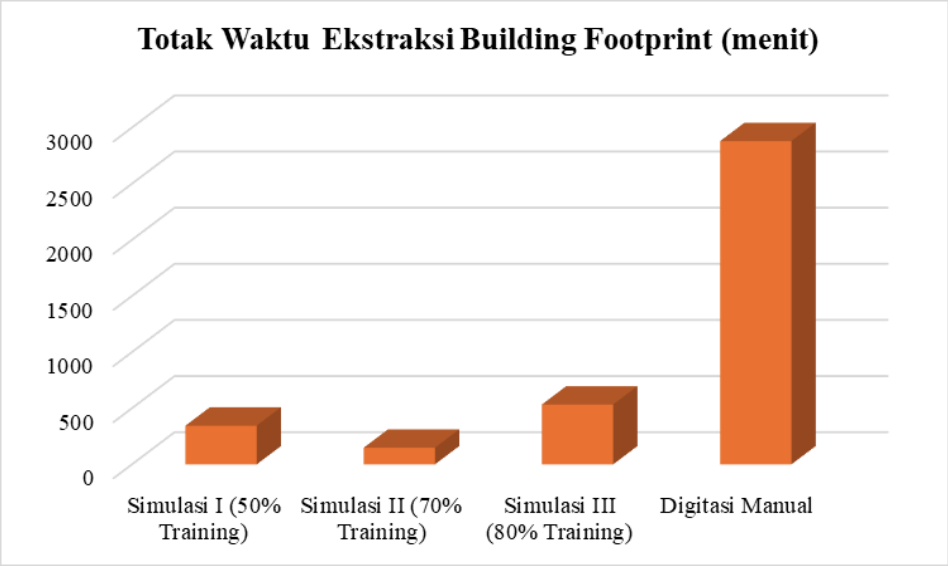
No	Uraian	Nilai	Satuan		Waktu	Jumlah
5	Sewa GNSS RTK	Rp.2.300.000	1	Unit	2	Rp. 4.600.000
6	Sewa Mobil dan BBM	Rp. 825.000	1	Unit	2	Rp, 1.650.000
7	Sewa Workstation	Rp. 150.000	1	Unit	4	Rp. 600.000
8	Sewa Perangkat Lunak GNSS Processing	Rp. 112.500	1	Unit	1	Rp. 112.500
9	Sewa Perangkat Lunak SIG	Rp. 225.000	1	Unit	3	Rp. 675.000
10	Sewa PC Workstation SFM Fotogrametri	Rp. 400.000	1	Unit	1	Rp. 400.000
	TOTAL BIAYA					**Rp. 15.064.172**

Tabel 10 Biaya pemetaan Mask-RCNN

No	Uraian	Nilai	Satuan		Waktu	Jumlah
1	Pilot	Rp.1.500.000	1	org	1	Rp. 1.500.000
2	Surveyor (Tenaga Terampil Jenjang 4)	Rp. 408.334	2	org	2	Rp. 1.633.336
3	Operator Pengolah Data (Tenaga Terampil Jenjang 5)	Rp. 473.334	1	org	2	Rp. 946.668
4	Sewa UAV (GNSS, PPK, Sensor Foto Udara)	Rp.2.000.000	1	Unit	1	Rp. 2.000.000
5	Sewa GNSS RTK	Rp.2.300.000	1	Unit	2	Rp. 4.600.000
6	Sewa Mobil dan BBM	Rp. 825.000	1	Unit	2	Rp, 1.650.000
7	Sewa Workstation	Rp. 150.000	1	Unit	2	Rp. 300.000
8	Sewa Perangkat Lunak GNSS Processing	Rp. 112.500	1	Unit	1	Rp. 112.500
9	Sewa Perangkat Lunak SIG	Rp. 225.000	1	Unit	1	Rp. 225.000
10	Sewa PC Workstation SFM Fotogrametri	Rp. 400.000	1	Unit	1	Rp. 400.000
	TOTAL BIAYA					**Rp. 13.367.504**

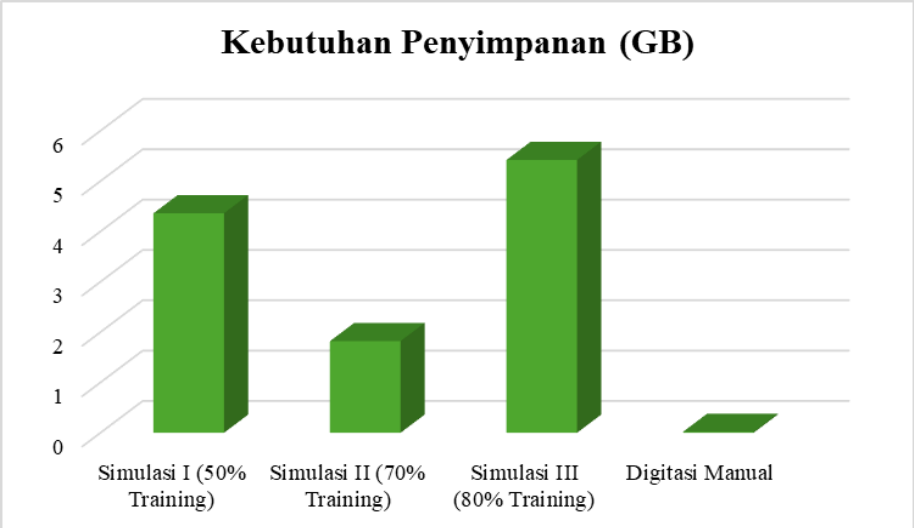
Berdasarkan analisis komparatif, metode digitasi manual menunjukkan peningkatan biaya sebesar 13% dibandingkan pendekatan otomatis berbasis Mask-RCNN. Temuan ini

menggarisbawahi pentingnya pertimbangan holistik dalam pemilihan metodologi, yang meliputi aspek anggaran, tingkat akurasi yang dibutuhkan, serta tujuan akhir pemetaan. Secara ekonomis, Mask-RCNN menawarkan efisiensi biaya namun dengan potensi keterbatasan dalam hal detail hasil. Di sisi lain, meskipun memerlukan investasi lebih besar, digitasi manual mampu memberikan presisi tinggi dan kelengkapan data yang ideal untuk kebutuhan analisis spasial mendalam. Pemilihan metodologi yang tepat pada akhirnya akan menentukan tingkat keberhasilan proyek melalui keseimbangan optimal antara efisiensi sumber daya dan kualitas output.



Gambar 13 Grafik perbandingan selisih total waktu pengolahan ekstraksi building footprint

Dari segi waktu training data sampai terbentuk *building footprint*, metode digitasi manual memerlukan waktu 2 hari atau jika dikonversi ke dalam menit yakni 2880 menit. Metode ini memerlukan waktu paling lama karena dikerjakan oleh tenaga manusia. Sedangkan metode Mask-RCNN mampu mereduksi total waktu training data dalam ekstraksi tapak bangunan dengan efektif dimulai dari metode Mask R-CNN simulasi III yang membutuhkan waktu 533 menit, simulasi I dengan waktu 344 menit, dan yang paling memerlukan sedikit waktu yakni yakni simulasi II hanya 147 menit.



Gambar 14 Grafik perbandingan selisih kebutuhan penyimpanan atau storage dalam ekstraksi tapak bangunan hingga pembentukan model 3D LOD1

Dari segi penyimpanan, digitasi manual hanya membutuhkan 18,7 MB yang mana metode ini hanya memerlukan sedikit penyimpanan untuk pembangunan model 3D LOD1. Metode Mask-

RCNN membutuhkan penyimpanan yang cukup besar karena terdapat masking data training per bangunan yang diperlukan untuk validasi hasil training. Kebutuhan penyimpanan paling besar dimulai dari simulasi III sebesar 5,42 GB, simulasi I sebesar 4,36 GB, dan yang paling kecil yakni simulasi II sebesar 1,82 GB. Tabel kebutuhan waktu dan penyimpanan setiap metode disajikan secara numerik di bawah ini.

Tabel 11 Tabel kebutuhan waktu dan penyimpanan setiap metode simulasi

Metode	Waktu Pemrosesan	Ukuran Penyimpanan
Simulasi I (50% Training)	5 jam 44 menit 41 detik	4,36 GB
Simulasi II (70% Training)	2 jam 27 menit 58 detik	1,82 GB
Simulasi III (80% Training)	8 jam 53 menit 11 detik	5,42 GB
Digitasi Manual	2 hari	18,7 MB

Dalam penelitian ini terjadi anomali pada simulasi III dimana seharusnya semakin banyak data traininng maka waktu pengolahan akan lebih singkat dan penyimpanan lebih sedikit karena data validasi yang dihasilkan akan lebih sedikit. Namun sebaliknya simulasi III yang menggunakan data training 80% lebih lama dari simulasi II dengan data training 70%. Hal ini dapat disebabkan karena saat data set bertambah 10% dari 70% ke 80%, model akan menghasilkan jumlah gambar metadata dan label yang semakin banyak. Selain itu data training yang banyak dapat terjadi karena bentuk bangunan yang heterogen sehingga model perlu mempelajari struktur bangunan lebih lama. Dengan training data yang lebih lama juga, maka epoch waktu pengolahan juga akan lebih lama yang menyebabkan penyimpanan membesar. Sebaliknya pada simulasi II dimana training 70% lebih cepat dari simulasi I training 50% karena saat training 70% komposisi bangunan yang digunakan lebih homogen daripada komposisi data training 50%. Saat training 70% dijalankan setelah training 50% maka kemungkinan data training 70% sudah tersimpan dalam chace (RAM atau disk) sedangkan saat tarining 50% belum tersimpan. Hal ini yang menyebabkan simulasi II training 70% lebih cepat dan membutuhkan penyimpanan lebih sedikit.

Dalam penelitian ini, Mask-RCNN dengan 70% data training dipilih sebagai metode utama untuk pemodelan 3D bangunan. Pemilihan ini didasarkan pada beberapa pertimbangan penting, terutama dalam konteks pembentukan model 3D LOD1, biaya, waktu pemrosesan, dan kebuthan penyimpanan. Digitasi manual memungkinkan kontrol penuh dan penyesuaian langsung oleh operator, memastikan bahwa setiap detail bangunan diidentifikasi dengan akurasi tinggi. Namun hal ini akan memakan waktu dan biaya yang cukup tinggi dalam pemetaan model 3D secara besar dan cepat. Namun kembali lagi ke tujuan pemetaan yang dilakukan apakah memerlukan tingkat akurasi yang tinggi atau untuk percepatan visualisasi. Meskipun metode digitasi manual ini lebih memakan waktu dan sumber daya, keandalan dan akurasinya menjadikannya pilihan yang tepat dalam pemetaan skala kecil yang memerlukan tingkat akurasi tinggi.

3D yang dihasilkan dari elevasi ketinggian NDSM foto udara UAV dan tapak bangunan hasil Mask R-CNN memberikan representasi bangunan yang cukup akurat dalam bentuk *building feature*. Teknik ini memungkinkan penggambaran tapak bangunan dengan baik, meskipun ada beberapa keterbatasan dalam menangkap detail sudut dan sisi bangunan. Perspektif tegak di beberapa titik tidak mampu menangkap elemen yang tersembunyi atau terhalang vegetasi sehingga beberapa fitur bangunan tidak terdefinisi dengan sempurna. Namun, pendekatan ini tetap efektif untuk mendapatkan gambaran umum dari struktur bangunan, terutama ketika digunakan dalam kombinasi dengan data lain untuk meningkatkan akurasi. Hasil 3D LOD1 *building feature* dari elevasi NDSM foto udara UAV untuk ketiga simulasi dan perbandingannya.

KESIMPULAN

Pemetaan foto udara menggunakan wahana UAV pada kelurahan Jawa, Kalimantan Timur menghasilkan resolusi spasial 0,023 meter dengan akurasi horisontal (CE90) sebesar 0,139 meter dan akurasi vertikal (LE90) 0,282 meter. Hasil ini telah memenuhi persyaratan ketelitian menurut SNI 8202:2019 untuk pemetaan skala 1:1000 Kelas 1, sekaligus memenuhi kriteria ketelitian vertikal untuk pembuatan garis kontur dengan interval 1 meter pada kategori Kelas 1.

Ekstraksi building footprint secara otomatis menggunakan deep learning Mask R-CNN dengan 3 simulasi data training pada foto udara UAV mampu diimplementasikan untuk pemetaan 3D skala besar dengan persentasi jumlah bangunan terdeteksi untuk simulasi I, II, dan III secara bertutut-turut yaitu 33%, 42%, dan 55% serta nilai presisi secara berturut-turut 0,589; 0,746; dan 0,794 . Model 3D LOD1 pada simulasi III dengan 80% data training menghasilkan visualisasi yang paling mendekati data digitasi manual diantara ketiga metode simulasi karena simulasi III cenderung tidak memiliki gap di tengah-tengah bangunan dan overlap antar bangunan menjadi lebih minim.

Dari segi biaya, metode Mask-RCNN sangat efektif dan efisien untuk diterapkan dalam modelan 3D skala besar yang cepat karena metode ini hanya membutuhkan biaya sebesar Rp 13.367.504 yang mempu mereduksi 13% biaya daripada menggunakan digitasi manual. Dari segi penyimpanan dan waktu, simulasi II dengan 70% data trainig menjadi rekomendasi dalam implentasi Mask R-CNN karena simulasi II memerlukan waktu pemrosesan paling cepat yakni 2 jam 27 menit dan 58 detik serta membutuhkan kapasitas penyimpanan paling kecil yakni sebesar 1,82 GB diantara ketiga metode simulasi.

Keterbatasan dan kendala dalam penelitian ini data orthophoto bangunan yang dihasilkan agar memperoleh hasil Mask R-CNN yang maksimal yakni bangunan yang memiliki pola seragam. Selain itu, pemilihan model 3D bangunan perlu memperhatikan jenis bangunan dan kondisi topografi yang berbeda seperti pada bangunan dengan bentuk yang tidak simetris atau daerah yang memiliki medan sulit. Penelitian lebih lanjut, eningkatan pemodelan 3D dapat dilakukan ke tahap selanjutnya yang lebih detail dengan menerapkan 3D LOD2, LOD3, atau bahkan LOD4 dengan menerapkan data LiDAR atau foto udara oblique agar menghindari potensi distorsi yang terjadi pada satu metode.

PERNYATAAN KONFLIK KEPENTINGAN

Penulis menyatakan tidak ada konflik kepentingan dalam artikel ini (*The authors declare no competing interest*).

REFERENSI

Abdul Rahman, R., Mohammad Yusof, Y., Kashefi, H., & Baharum, S. (2012). Developing Abdulla, W. (2017). Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow.

Anurogo, W., Lubis, M. Z., Khoirunnisa, H., Hanafi, D. S. P. A., Rizki, F., Surya, G., & Dewanti, N. A. (2017). A simple aerial photogrammetric mapping system overview and image acquisition using unmanned aerial vehicles (UAVs). Geospatial Information, 1(1), 11-18.

Danielczuk, M., Matl, M., Gupta, S., Li, A., Lee, A., Mahler, J., & Goldberg, K. (2019, May). Segmenting unknown 3d objects from real depth images using mask r-cnn trained on synthetic data. In 2019 International Conference on Robotics and Automation (ICRA) (pp. 7283-7290). IEEE.

Deliry, S. I., & Avdan, U. (2021). Accuracy of unmanned aerial systems photogrammetry and structure frommotion in surveying and mapping: a review. Journal of the Indian Society of Remote Sensing, 49(8), 1997-2017.

Eltner, A., & Sofia, G. (2020). Structure from motion photogrammetric technique. In Developments in Earth surface processes (Vol. 23, pp. 1-24). Elsevier.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2018). Mask R-CNN. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2), 386–397. https://doi.org/10.1109/TPAMI.2018.2844175

Huang, G., Liu, Z., Van Der Maaten, L., et al.: 'Densely connected convolutional networks'. IEEE Proc. Conf. Computer Vision and Pattern Recognition, Honolulu, HI, USA, 7 July 2017, vol. 1, p. 3

Javadnejad, F. (2017). Small Unmanned Aircraft Systems (UAS) for Engineering Inspections and Geospatial Mapping. Dissertation of Oregon State University, 1-132.

Jiang, S., Jiang, W., & Wang, L. (2021). Unmanned Aerial Vehicle-Based Photogrammetric 3D Mapping: A survey of techniques, applications, and challenges. IEEE Geoscience and Remote Sensing Magazine, 10(2), 135-171.

Khayyal, H. K., Zeidan, Z. M., & Beshr, A. A. (2022). Creation and spatial analysis of 3D city modeling based on GIS data. Civil Engineering Journal, 8(1), 105.

Kraff, N. J., Wurm, M., & Taubenbock, H. (2020). Uncertainties of Human Perception in Visual Image Interpretation in Complex Urban Environments. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 4229–4241. https://doi.org/10.1109/jstars.2020.3011543

Le, T. D., Huynh, D. T., & Pham, H. V. (2018, November). Efficient human-robot interaction using deep learning with mask R-CNN: Detection, recognition, tracking and segmentation. In 2018 15th International conference on control, automation, robotics and vision (ICARCV) (pp. 162-167). IEEE.

Nex, F., & Remondino, F. (2014). UAV for 3D mapping applications: a review. Applied geomatics, 6, 1-15.

Nys, G. A., Billen, R., & Poux, F. (2020). Automatic 3d buildings compact reconstruction from LiDAR point clouds. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 43, 473-478.

Peters, R., Dukai, B., Vitalis, S., van Liempt, J., & Stoter, J. (2022). Automated 3D reconstruction of LoD2 and LoD1 models for all 10 million buildings of the Netherlands. Photogrammetric Engineering & Remote Sensing, 88(3), 165-170.

Remondino, F. (2011). Heritage recording and 3D modeling with photogrammetry and 3D scanning. Remote sensing, 3(6), 1104-1138.

Templin, T. (2023). LOD 0 LOD1. The Routledge Handbook of Geospatial Technologies and Society, 348.

Ullman, S. (1979). The interpretation of structure from motion. Proceedings of the Royal Society of London. Series B. Biological Sciences, 203(1153), 405-426.

Xavier, A. I., Villavicencio, C., Macrohon, J. J., Jeng, J. H., & Hsieh, J. G. (2022). Object detection via gradient-based mask R-CNN using machine learning algorithms. Machines, 10(5), 340.

Zhang, C., Zhou, J., Wang, H., Tan, T., Cui, M., Huang, Z., ... & Zhang, L. (2022). Multi-species individual tree segmentation and identification based on improved mask R-CNN and UAV imagery in mixed forests. Remote Sensing, 14(4), 874.