ISSN 2809-1833
Prosiding FIT ISI Vol 1, 2021 (248-260)
@Ikatan Surveyor Indonesia – Teknik Geodesi UNDIP

PEMODELAN BANGUNAN 3D MENGGUNAKAN *FOOTPRINT* BANGUNAN HASIL EKSTRAKSI MASK R-CNN DAN *DENSE POINT CLOUD* DARI FOTO UDARA UAV

Tata Bahtera Negara[1], Harintaka[2]
[1]Departemen Teknik Geodesi-Fakultas Teknik Universitas Gadjah Mada
Jl. Grafika No 2, Sinduadi, Mlati, Sleman -55284 Telp./Faks: (+62274) 520226, e-mail: tbnegara@gmail.com
[2] Departemen Teknik Geodesi-Fakultas Teknik Universitas Gadjah Mada
Jl. Grafika No 2, Sinduadi, Mlati, Sleman -55284 Telp./Faks: (+62274) 520226, e-mail: harintaka@ugm.ac.id

ABSTRAK

Bangunan merupakan salah satu objek penting yang secara spasial dibutuhkan dalam berbagai pekerjaan khususnya untuk perencanaan dan pembangunan kota. Bangunan dalam representasi 3D telah terbukti mampu menunjang kegiatan perencanaan dengan baik mengingat dunia nyata berada dalam sistem 3D. Salah satu metode yang paling sederhana untuk membuat model bangunan 3D dalam cakupan wilayah yang luas adalah dengan melakukan ekstrusi *footprint* bangunan. Data yang umum digunakan dalam metode ini adalah *footprint* bangunan hasil digitasi manual pada *orthomosaic* dan komponen elevasi berupa *point cloud* dari *Light Detection and Ranging* (LiDAR). Pekerjaan digitasi manual umumnya memakan waktu yang relatif lama dan sumber daya manusia yang cenderung tinggi apabila data yang diproses semakin besar, selain itu hasil digitasi juga tidak konsisten relatif kepada keterampilan operator. Disisi lain, penggunaan *point cloud* LiDAR menyebabkan metode ini kurang terjangkau dari sisi biaya. Dalam penelitian ini, dilakukan pemodelan bangunan 3D menggunakan *footprint* bangunan yang dihasilkan secara otomatis dengan teknik *Mask Region-based Convolutional Neural Network* (Mask R-CNN) dan *dense point cloud* yang diperoleh dari pengolahan foto udara di kawasan kampus pusat Universitas Riau yang diakuisisi menggunakan *Unmanned Aerial Vehicle* (UAV). Metode yang diterapkan memberikan hasil yang cukup baik. Model Mask R-CNN yang dilatih dalam 25 *epoch* pembelajaran menghasilkan akurasi pembelajaran senilai 96,80% dan *footprint* bangunan yang dihasilkan memiliki nilai *recall* (kelengkapan) 88,83%, kepresisian 91,65%, dan nilai *Intersection over Union* (IoU) 91,90% ketika dibandingkan dengan data *ground truth*. Proses ekstrusi *footprint* bangunan hasil ekstraksi otomatis tersebut menghasilkan model bangunan 3D dalam *Level of Detail* (LOD) 2 dengan nilai *Root Mean Square Error* (RMSE) < 2 meter berdasarkan standar *City Geography Markup Language* (CityGML).

Kata kunci : Bangunan, Mask R-CNN, Model 3D, Orthomosaic, UAV.

ABSTRACT

Buildings are one of the notable objects which spatially needed for various jobs, especially for urban planning and development. Buildings in 3D representation have been proven to be able to support planning activities properly considering that the real world is in a 3D system. One of the simplest methods to create 3D building models over a wide area is to extrude the buildings footprint. The data commonly used in this method are the building footprint as result of manual digitizing on orthomosaic and the elevation component in the form of point cloud from Light Detection and Ranging (LiDAR). Manual digitizing generally takes a relatively long time and human resources tend to be high if the data processed is getting bigger, the results are also inconsistent relative to the skills of the operator. On the other hand, the use of LiDAR point cloud causes this method less affordable in terms of cost. In this study, 3D building modeling was carried out using building footprints generated automatically with the Mask Region-based Convolutional Neural Network (Mask R-CNN) technique and dense point cloud obtained from aerial imagery processing in the main campus area of Riau University acquired using Unmanned Aerial Vehicle (UAV). The method applied gives quite good results. The Mask R-CNN model that was trained in 25 learning epochs had a learning accuracy of 96.80% and the building footprint had a recall value (completeness) of 88.83%, 91.65% precision, and 91.90% Intersection over Union (IoU) when compared to ground truth data. The extrusion process of the auto-extracted building footprint produced 3D building models in Level of Detail (LOD) 2 with a Root Mean Square Error (RMSE) value of < 2 meters based on City Geography Markup Language (CityGML) standards.

Keywords : Building, Mask R-CNN, 3D Model, Orthomosaic, UAV.

1. PENDAHULUAN

Informasi tutupan lahan selalu berubah seiring dengan adanya aktivitas antropogenik dari manusia (Gómez *et al*., 2016). Dinamisnya bangunan sebagai salah satu informasi spasial dari tutupan lahan serta manfaatnya dalam berbagai fungsi menyebabkan informasi bangunan sangat dibutuhkan keberadaan dan kebaruannya sehingga dibutuhkan data yang cukup akurat, murah dan dapat diperoleh dengan cepat untuk memperoleh informasi bangunan terkini.

Cara dan metode dalam merepresentasikan bangunan selalu berubah seiring dengan perkembangan teknologi. Informasi bangunan berevolusi dari yang awalnya berbasis analog menjadi berbasis digital, dan dari planar dua dimensi (2D) menjadi representasi objek tiga dimensi (3D). Analisis yang dapat dilakukan dari perubahan representasi bangunan 2D menjadi 3D telah terbukti memberikan hasil analisis yang lebih representatif dan mendalam (Biljecki *et al*., 2017). Adanya model bangunan 3D telah menunjukkan manfaat yang tidak dapat diperoleh menggunakan informasi bangunan dalam representasi 2D dalam berbagai aplikasi seperti perencanaan dan manajemen wilayah perkotaan, simulasi bencana banjir, monitoring lahan, visualisasi tiga dimensi, asesmen potensi radiasi matahari dari bentuk atap bangunan, dan sebagainya.

Salah satu pendekatan yang paling sederhana untuk melakukan pemodelan bangunan 3D adalah dengan melakukan ekstrusi data *footprint* bangunan menggunakan data elevasi. Data yang paling sering dimanfaatkan untuk mengeksekusi pendekatan ini adalah data foto udara standar untuk memperoleh *footprint* bangunan dan data *Light Detection and Ranging* (LiDAR) berupa *point cloud* sebagai komponen elevasi (Kwak *et al*., 2011; Peeroo *et al*., 2017).

Namun, pembuatan model bangunan tiga dimensi menggunakan data *footprint* bangunan yang biasanya didigitasi dari foto udara standar dan data *point cloud* LiDAR membutuhkan biaya yang relatif tinggi karena membutuhkan dua buah sensor yang terdiri atas kamera udara metrik dan instrumen LiDAR untuk memperoleh data foto udara dan *point cloud*, sehingga tidak ekonomis dari sisi biaya terutama apabila wilayah yang akan dimodelkan tidak terlalu luas. 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 *et al*., 2020).

Untuk mengatasi kekurangan pada metode tersebut, dilakukan pembuatan model bangunan tiga dimensi melalui data *footprint* bangunan yang diperoleh secara otomatis dengan perangkat komputer menggunakan *deep learning* dan data *dense point cloud* yang diperoleh melalui proses fotogrametri dari foto udara yang diakuisisi menggunakan wahana *Unmanned Aerial Vehicle* (UAV).

UAV atau pesawat udara tanpa awak merupakan sistem pesawat kendali jarak jauh sehingga tidak memerlukan pilot yang langsung mengendalikan pada wahananya. Beberapa keunggulan teknologi foto udara UAV meliputi biaya yang murah, ukuran wahana yang lebih kecil sehingga memiliki aksesibilitas yang tinggi, mengurangi resiko kecelakaan pada saat akuisisi data, kemudahan pengoperasian, dan akuisisi data yang cepat serta kemudahan menyesuaikan wilayah akuisisi sesuai dengan kebutuhan (Ammour *et al*., 2017). Akuisisi data foto udara dengan wahana UAV mampu menghasilkan *point cloud* yang diperoleh melalui proses *Structure from Motion* (SFM)–fotogrametri (Mlambo *et al*., 2017; Fawcett *et al*., 2019).

Metode ekstraksi objek bangunan pada foto udara UAV saat ini dapat dilakukan secara otomatis dengan pendekatan *deep learning*. *Deep learning* mengacu pada jaringan saraf tiruan yang terdiri dari banyak lapisan neuron buatan yang meniru cara kerja otak manusia yang disusun secara hierarki untuk memproses informasi (Hemanth dan Estrela, 2017). *Deep learning* dapat bekerja secara efektif untuk proses pembelajaran pengenalan objek khususnya untuk deteksi dan ekstraksi objek bangunan pada foto udara secara cepat dan akurat (Uba, 2016). Salah satu model *deep learning* yang saat ini sedang berkembang dan banyak digunakan dalam proses ekstraksi bangunan adalah *Mask Region-based Convolutional Neural Network* (Mask R-CNN) (Zhao *et al.*, 2018; Chen dan Li, 2019). Model Mask R-CNN banyak digunakan karena kesederhanaan model, fleksibilitas, dan juga kemampuannya dalam mendeteksi objek yang rapat (He *et al*., 2018).

Teknologi SfM-fotogrametri berbasis UAV dan *deep learning* belum dimanfaatkan secara maksimal terutama pada pekerjaan pembuatan model bangunan 3D sementara memiliki potensi yang tinggi. Pemanfaatan teknologi ini dapat meningkatkan efisiensi dalam akuisisi data, menekan biaya, serta aspek otomatisasi dalam pembuatan model bangunan 3D.

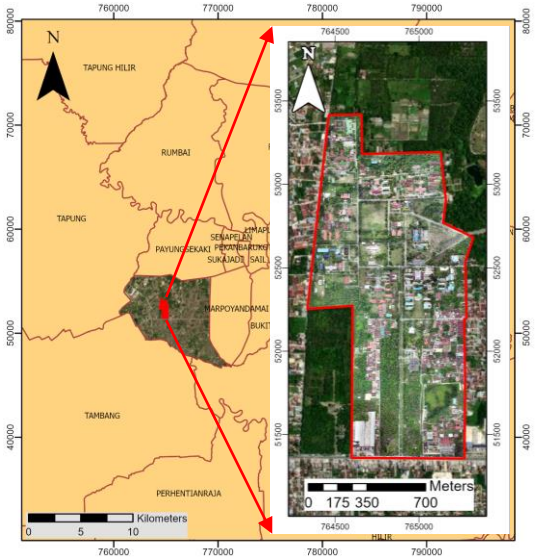
2. METODE PENELITIAN

2.1 Data dan Lokasi

Data yang digunakan dalam penelitian ini adalah data *orthomosaic* dengan format *Tag Image File Format* (TIFF) dan data *dense point cloud* dalam format *Laser* (LAS) dari hasil pemrosesan foto udara UAV dengan teknik SfM-fotogrametri. *Orthomosaic* dan *dense point cloud* yang digunakan melingkupi seluruh cakupan lokasi

penelitian dan memiliki sistem koordinat yang sudah terikat dengan sistem koordinat tanah karena telah diproses menggunakan *Ground Control Points* (GCP). *Orthomosaic* yang digunakan sudah terbebas dari efek bangunan rebah dan memiliki nilai *Ground Sampling Distance* (GSD) 5,7 cm, sementara *dense point cloud* yang digunakan terdiri atas 60,5 juta lebih titik dengan kerapatan ± 42 titik/m^2.

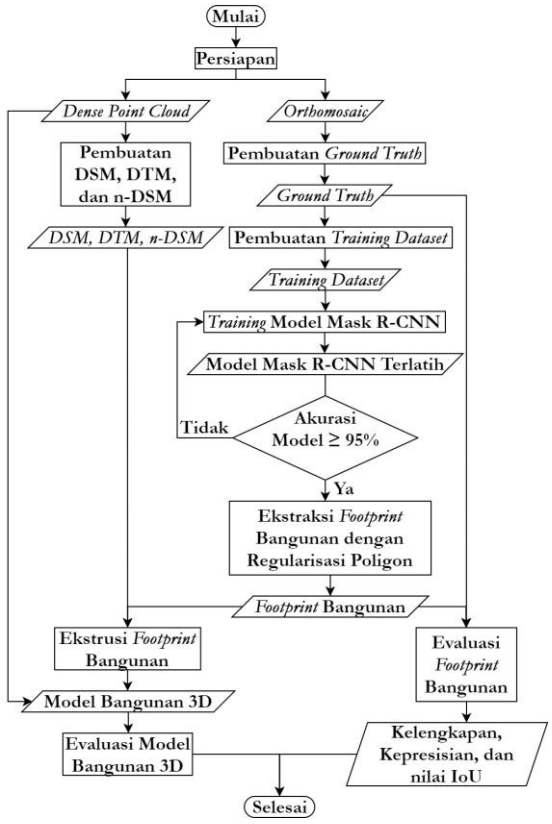
Lokasi penelitian ini berada pada wilayah kampus Universitas Riau (UNRI), Kota Pekanbaru, Provinsi Riau dengan luas ± 142,58 hektar. Batas lokasi yang menjadi area kajian pada penelitian ini disesuaikan dengan kebutuhan penelitian. Batas lokasi penelitian yang dimaksud ditunjukkan pada Gambar 1.



Gambar 1. Lokasi Penelitian (Sumber : Badan Informasi Geospasial dan *World Imagery Basemap* milik Esri)

2.2 Metodologi

Penelitian dilaksanakan dalam beberapa tahapan dimulai dari persiapan, pembuatan *ground truth*, pembuatan dataset untuk *training* model Mask R-CNN, dilanjutkan dengan proses *training* model Mask R-CNN, ekstraksi *footprint* bangunan secara otomatis menggunakan Mask R-CNN yang dilanjutkan dengan proses regularisasi, pemodelan bangunan 3D dengan melakukan ekstrusi *footprint* yang telah diekstraksi secara otomatis menggunakan Mask R-CNN, serta evaluasi bangunan 3D hasil pemodelan. Diagram alir tahapan pelaksanaan penelitian ditunjukkan pada Gambar 2.



Gambar 2. Diagram Alir Penelitian

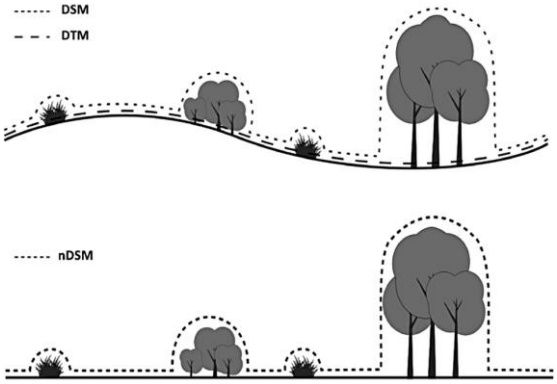
2.2.1 Persiapan

Selain bahan berupa data penelitian seperti yang sudah disebutkan sebelumnya, tentunya juga diperlukan alat agar penelitian dapat dilakukan. Alat yang dibutuhkan dalam penelitian ini adalah seperangkat komputer atau laptop yang dilengkapi dengan perangkat lunak ArcGIS Pro yang telah dilengkapi *framework deep learning* untuk keperluan *training* model Mask R-CNN, ekstraksi *footprint* bangunan, ekstrusi *footprint* untuk mendapatkan model bangunan 3D, hingga evaluasi hasil.

Laptop yang digunakan memiliki spesifikasi *processor* AMD Ryzen 7 4800H yang dilengkapi dengan *Graphical Processing Unit* (GPU) NVIDIA GeForce RTX 2060 dan RAM 16GB. Komputer atau laptop dengan spesifikasi GPU yang tinggi dibutuhkan untuk dapat memproses data lebih cepat khususnya pada saat melakukan *training* model Mask R-CNN dan ekstraksi *footprint* bangunan secara otomatis (Hemanth dan Estrela, 2017). Proses *training* model Mask R-CNN tidak akan dapat dijalankan apabila laptop yang digunakan tidak memiliki GPU yang memadai.

2.2.2 Pembuatan Model Elevasi Digital

Pada tahap ini, digunakan data *dense point cloud* yang telah diklasifikasi titik yang merupakan *ground* atau terrainnya untuk membuat model elevasi digital. Model elevasi digital adalah data digital yang merepresentasikan elevasi dari permukaan bumi ataupun objek-objek yang ada di atasnya. Model elevasi digital yang dihasilkan pada tahap ini terdiri atas *Digital Surface Model* (DSM), *Digital Terrain Model* (DTM), dan *Normalized Digital Surface Model* (nDSM). DSM merupakan model permukaan yang berisi informasi elevasi permukaan bumi beserta objek-objek yang ada diatasnya seperti bangunan, pepohonan, ataupun objek lainnya, dan DTM merupakan model permukaan yang berisi informasi elevasi permukaan bumi saja, sedangkan nDSM merupakan model permukaan yang berisi informasi elevasi objek-objek diatas permukaan bumi relatif terhadap permukaan bumi *(Miroslaw-Swiatek et al.*, 2016). Ilustrasi DSM, DTM, dan nDSM ditampilkan pada Gambar 3.



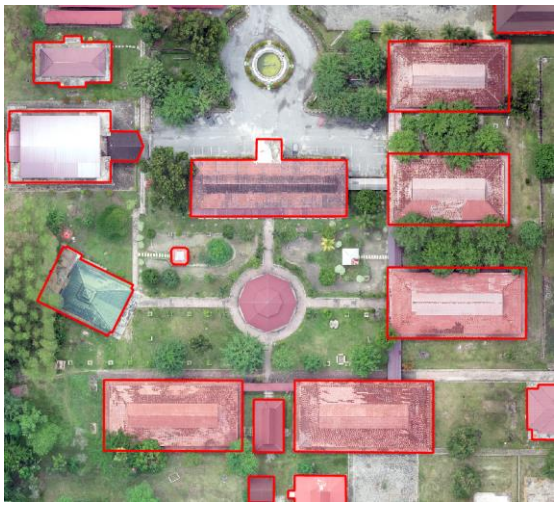
Gambar 3. DSM, DTM, dan nDSM (Miroslaw-Swiatek *et al*., 2016)

DSM diperoleh dengan menginterpolasi langsung seluruh titik yang ada pada *dense point cloud*, dan DTM diperoleh dengan menginterpolasi titik *ground* atau *terrain* yang telah diklasifikasi sebelumnya pada *dense point cloud*, sedangkan nDSM diperoleh dengan melakukan diferensiasi DSM dengan DTM mengikuti Persamaan (1).

$$nDSM = DSM - DTM \quad (1)$$

2.2.3 Pembuatan *Ground Truth* dan *Training Dataset*

Ground truth merupakan data acuan yang sesuai dengan kondisi sebenarnya di lapangan. *Ground truth* digunakan sebagai data pembanding untuk melakukan evaluasi akurasi dari hasil ekstraksi *footprint* bangunan menggunakan *deep learning* (Monteiro dan Campilho, 2005). Selain itu, data *ground truth* juga diperlukan sebagai input dalam proses pembuatan dataset untuk melakukan *training* model *deep learning*. *Ground truth* pada penelitian ini diperoleh melalui metode digitasi interaktif dengan interpretasi visual pada *orthomosaic* sesuai dengan klasifikasi penelitian yaitu kelas bangunan. Pembuatan *ground truth* menghasilkan data vektor bangunan berupa poligon yang pada seluruh area *orthomosaic*.



Gambar 4. Pembuatan *Ground Truth*

Training dataset merupakan kumpulan data yang akan digunakan sebagai input pada proses *training* model *deep learning*. *Training dataset* berisi data potongan *orthomosaic* beserta pasangan data *ground truth*-nya. Tahap pertama dalam melakukan pembuatan *training dataset* adalah membuat *boundary* yang menjadi *training area* dengan memastikan objek – objek bangunan dan non-bangunan yang ada di dalam area tersebut sudah mewakili keseluruhan data yang ada pada *orthomosaic*. Variasi objek bangunan yang ada pada *training area* didasarkan pada beberapa aspek seperti ukuran, bentuk, warna, maupun kerapatan antar bangunan. Dari data *ground truth* dan *boundary* tersebut dilakukan pembuatan *tile - tile* pasangan *orthomosaic* dan *ground truth* pada *training area*. Parameter yang digunakan dalam proses pembuatan *training dataset* ini adalah ukuran tile, dan lompatan/*stride*. Ukuran *tile* yang dibuat adalah 1024 × 1024 piksel. Ukuran ini dipertimbangkan melalui *trial by error* karena dianggap sebagai ukuran paling cocok dan sesuai dengan spesifikasi perangkat yang digunakan dan juga menyesuaikan ukuran bangunan dalam satuan piksel pada *orthomosaic* yang digunakan. Nilai lompatan yang digunakan adalah 512 piksel

sehingga terdapat pertampalan 50% antar *tile* yang bersebelahan, selain itu juga ditambahkan nilai rotasi 180^o. Hal ini dimaksudkan sebagai upaya augmentasi data sehingga model *deep learning* tidak mengalami kekurangan data pada saat proses pembelajaran untuk menghasilkan prediksi dan akurasi *training* yang baik. Proses pembuatan *training dataset* secara otomatis ini membutuhkan waktu 3 menit 6 detik dan menghasilkan sejumlah pasangan *tile orthomosaic* dan *ground truth*.

Gambar 5. Contoh pasangan *tile orthomosaic* dan *ground truth*-nya pada *training dataset*

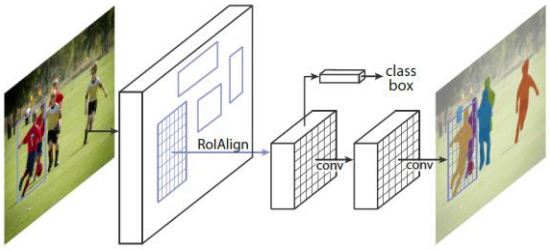
2.2.4 *Training* Model Mask R-CNN

Proses *training* merupakan proses bagi model *deep learning* untuk mempelajari hubungan data antara *ground truth* yang merupakan output dan *orthomosaic* yang merupakan input dalam *training dataset* sehingga memperoleh seperangkat bobot yang sesuai untuk melakukan estimasi dengan baik ketika diberikan data input yang berbeda *(Khan et al.*, 2018). *Training* dilakukan dengan *backbone* model ResNet50. ResNet50 memiliki 50 lapisan berisi *convolution* dan ReLU diikuti dengan *downsampling* menggunakan operasi *pooling* untuk menyandikan atau membuat kode dari input data kedalam representasi fitur pada berbagai tingkatan (He *et al*., 2016). Nilai *epoch* maksimal yang didefinisikan adalah 30 *epoch* dengan memberlakukan *early stopping* untuk menghindari model yang bersifat *overfitting*. *Training dataset* dibagi menjadi 80% sebagai data *training* untuk proses *training*, dan 20% sebagai data validasi untuk proses tes atau validasi untuk menghitung nilai akurasi pembelajaran model. Pembagian ini dilakukan secara acak oleh komputer. Akurasi *training* yang disyaratkan adalah minimal 95%, sehingga diharapkan akan mampu menghasilkan *footprint* bangunan dengan akurasi tinggi.

2.2.5 Ekstraksi *footprint* bangunan

Model Mask R-CNN yang sudah memiliki model pembobotan dapat digunakan untuk melakukan ekstraksi *footprint* bangunan. Proses ini dapat dilakukan melalui segmentasi *instance*. Segmentasi *instance* merupakan kombinasi antara pekerjaan deteksi objek dan segmentasi, sehingga menghasilkan keluaran fitur yang berbeda untuk setiap individu objek walaupun berada dalam kelas objek yang sama. Mask R-CNN mampu menghasilkan keluaran fitur yang berbeda setiap objek yang berdekatan. (He *et al*., 2018).

Dilihat dari arsitekturnya, Mask R-CNN terdiri atas 2 bagian arsitektur yaitu *backbone* dan *head. Backbone* berfungsi sebagai jaringan yang melakukan ekstraksi fitur setiap foto dan *head* untuk klasifikasi dan regresi beserta ekstraksi fitur pada tiap *Region of Interest* (RoI). Untuk meningkatkan kecepatan dan akurasi, Mask R-CNN menggunakan kombinasi ResNet-*Feature Pyramid Network* (FPN) sebagai *backbone* dan *head* (He *et al*., 2018).



Gambar 6. *Framework* model Mask R-CNN *(He et al.*, 2018)

Dalam penggunaan Mask R-CNN untuk ekstraksi *footprint* bangunan, perlu dilakukan regularisasi atau penyederhanaan karena *footprint* hasil ekstraksi memiliki bentuk yang tidak teratur dan banyak mengandung *noise*.

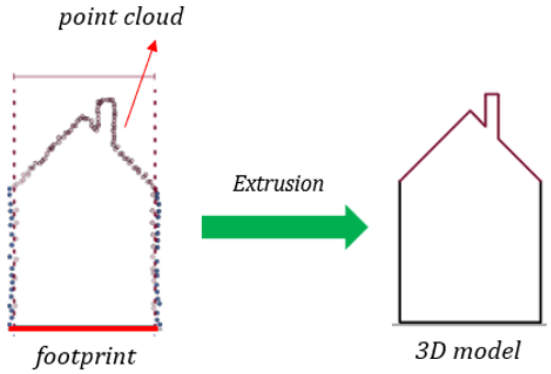
Gambar 7. Regularisasi *footprint* bangunan (Zhao *et al*., 2018)

Penyederhanaan ini dilakukan dalam 2 tahapan yang terdiri atas penyederhanaan bentuk dengan algoritma Douglas-Peucker, dan penyederhanaan sudut – sudut poligon. Hasil akhir dari proses regularisasi adalah *footprint* bangunan dengan bentuk yang teratur dan lebih sesuai dengan kondisi asli di lapangan. Meskipun begitu, nilai akurasi hasil ekstraksi sebelum dan sesudah

dilakukan regularisasi tidak akan berubah secara signifikan (Zhao *et al*., 2018).

2.2.5 Ekstrusi *Footprint* Bangunan

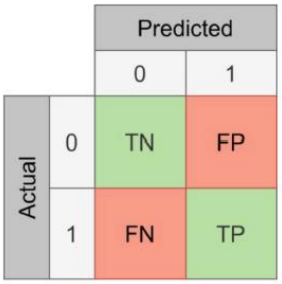
Pada tahap ini *footprint* bangunan yang telah diekstraksi secara otomatis menggunakan model Mask R-CNN diekstrusi menggunakan DSM, DTM, dan nDSM yang berasal dari *dense point cloud*. Seperti halnya proses ekstraksi *footprint*, proses ekstrusi *footprint* untuk menghasilkan model bangunan 3D ini juga dilakukan secara otomatis (Biljecki *et al*., 2017). Hasil dari tahap ini adalah model bangunan 3D yang tidak hanya dalam bentuk balok atau kubus, melainkan memiliki bentuk atap bangunan.



Gambar 8. Pemodelan bangunan 3D dengan cara melakukan ekstrusi *footprint* bangunan (Biljecki *et al*., 2017).

2.2.5 Evaluasi *Footprint* dan Model Bangunan 3D

Footprint bangunan yang telah diekstraksi dengan Mask R-CNN dapat dianalisis melalui nilai akurasinya. Analisis akurasi merupakan tahapan yang krusial dalam pekerjaan dengan *deep learning*. Salah satu alat yang dapat digunakan untuk menghitung nilai akurasi adalah menghitung nilai matrik konfusi dengan membandingkan hasil ekstraksi terhadap *ground truth*. Matriks konfusi berisi elemen hasil ekstraksi yang terdiri atas *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP), dan *False Negatives* (FN) (de Carvalho *et al*., 2021).



Gambar 9. Matriks konfusi (de Carvalho *et al*., 2021)

TP dalam hal ini merupakan jumlah elemen hasil ekstraksi yang benar sesuai dengan *ground truth*, sementara FP menunjukkan jumlah elemen hasil ekstraksi yang salah, dan FN merupakan jumlah bangunan yang gagal diekstraksi. Nilai utama dalam mengevaluasi hasil ekstraksi *footprint* bangunan adalah *precision* (kepresisian) dan *recall* (kelengkapan). Masing – masing nilai tersebut dihitung menggunakan Persamaan (2) dan (3).

$$Kepresisian = \frac{TP}{TP + FP} \quad (2)$$

$$Kelengkapan = \frac{TP}{TP + FN} \quad (3)$$

Selain itu, untuk menunjukkan kedekatan hasil ekstraksi dengan *ground truth*, dilakukan perhitungan indeks *intersection over union* (IoU) atau disebut juga dengan indeks *Jaccard*. Indeks IoU merupakan suatu metode intuitif untuk mengevaluasi hasil ekstraksi objek dengan *deep learning* (Zhang *et al*., 2020). Indeks IoU dihitung sebagai rasio luasan yang saling tumpang tindih (*intersection*) antara *ground truth* dengan hasil ekstraksi menggunakan *deep learning* dibandingkan dengan total luasannya (*union*). Perhitungan akurasi menggunakan indeks IoU dilakukan dengan menggunakan Persamaan (4).

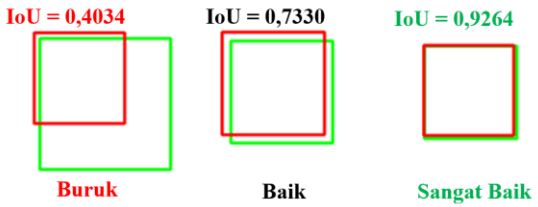
$$IoU\ (A, B) = \frac{Area\ of\ Intersection\ (A \cap B)}{Area\ of\ Union\ (A \cup B)} \quad (4)$$

Dalam hal ini :

A : *footprint* hasil ekstraksi dengan Mask R-CNN
B : *ground truth*

Indeks IoU yang tinggi menunjukkan tingkat kesamaan hasil ekstraksi otomatis dengan data *ground truth* yang tinggi. Rentang indeks IoU yang mungkin didapatkan adalah 0 – 1 atau 0% - 100%. Indeks IoU 1 atau 100% menunjukkan kesamaan 100% antara data hasil ekstraksi menggunakan *deep learning* dengan data

ground truth, sebaliknya apabila indeks IoU memiliki nilai IoU 0,2 atau 20% berarti bahwa tingkat kesamaan antara data data hasil ekstraksi menggunakan *deep learning* dengan data *ground truth* hanya 20% dan tergolong rendah, menandakan bahwa hasil ekstraksi yang didapatkan tidak akurat.



Gambar 10. Ilustrasi indeks IoU (Cowton *et al*., 2019)

Evaluasi model bangunan 3D dilakukan mengacu kepada standar CityGML. CityGML merupakan standar data 3D yang diterbitkan oleh *Open Geospasial Consortium* (OGC). Berdasarkan CityGML, objek yang sama dapat direpresentasikan dalam *Level of Detail* (LOD) yang berbeda secara bersamaan, khususnya untuk objek bangunan yang terdiri atas 5 LOD. Untuk LOD 0 yang paling kasar pada dasarnya adalah model terain digital dua setengah dimensi. LOD 1 adalah model blok terdiri dari bangunan prismatik atap yang flat atau datar (belum memiliki bentuk atap). LOD 2 menunjukkan bangunan yang sudah memiliki struktur atap. LOD 3 menunjukkan model arsitektur dengan struktur dinding dan atap yang mendetail. LOD 4 melengkapi LOD 3 dengan menambahkan struktur interior. Misalnya, bangunan terdiri dari kamar, pintu interior, tangga, dan furniture (Biljecki *et al*., 2016).

Gambar 11. LOD berdasarkan CityGML (Biljecki *et al*., 2016)

Selain itu, analisis secara geometri juga dilakukan. Analisis uji akurasi geometri bangunan dibutuhkan untuk menguji tingkat akurasi ukuran dimensi bangunan meliputi panjang, lebar dan tinggi bangunan dari hasil yang dimodelkan dengan ukuran aslinya di lapangan. Ukuran asli yang dimaksud disini adalah pengukuran yang diperoleh dari *measurement* langsung pada *dense point cloud*. Nilai akurasi bangunan berdasarkan CityGML menggunakan acuan nilai *Root Mean Square Error* (RMSE). Nilai RMSE yang dihasilkan pada model bangunan LOD 2 sesuai dengan standar CityGML yaitu < 2 m (Gröger *et al*., 2012). Nilai RMSE dihitung dengan menggunakan Persamaan (5).

$$RMSE = \sqrt{\frac{\sum(R - R_1)^2}{n}} \quad (5)$$

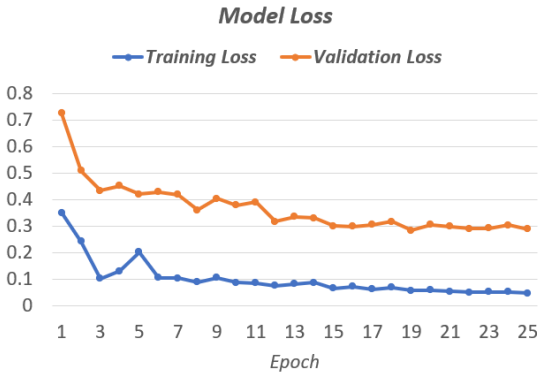
Dalam hal ini :

R : panjang, lebar, atau tinggi model bangunan
R_1: panjang, lebar, atau tinggi bangunan asli di lapangan
n : jumlah sampel

3. HASIL DAN PEMBAHASAN

3.1 Akurasi *Training* Model Mask R-CNN

Training model Mask R-CNN dilakukan dalam 25 *epoch* dengan menggunakan 1256 pasang *tile orthomosaic* dan *ground truth* yang kemudian dibagi menjadi 1005 pasang digunakan untuk *training*, dan 251 pasang digunakan untuk tes atau validasi untuk menghitung nilai akurasi *training*. Proses *training* yang dilakukan dengan mengaktifkan GPU memerlukan waktu 12 jam 3 menit 10 detik dan menghasilkan model Mask R-CNN terlatih yang siap digunakan untuk ekstraksi *footprint* bangunan.



Gambar 12. Grafik *loss* dari model Mask R-CNN selama proses *training*

Dalam proses *training* model *deep learning* dikenal dengan adanya nilai *loss*. Nilai *loss* juga berarti nilai kesalahan atau *error* dalam proses *training*. *Loss* pada model terdiri atas *training loss* dan *validation loss*. *Training loss* merupakan nilai kesalahan pada model yang telah melewati proses *training* dengan memproses data *training*, sedangkan *validation loss* merupakan nilai kesalahan model dengan memproses data validasi (Sornapudi *et al*., 2018).

Pada Gambar 12 dapat diamati bahwa nilai *validation loss* selalu lebih besar dari pada nilai *training loss*. Hal ini lumrah terjadi dikarenakan nilai *validation loss* merupakan nilai kesalahan model dalam

memproses data validasi yang tidak dikenal oleh model karena tidak dimasukkan dalam proses *training*.

Pada *epoch* 1 model memiliki *training loss* dan *validation loss* senilai 0,3520 dan 0,7265. Seiring dengan berjalannya proses *training*, model menjadi semakin membaik dikarenakan penurunan nilai *training loss* dan *validation loss*, hingga pada *epoch* 25 proses *training* dihentikan. Hal ini dikarenakan perbaikan model semakin kecil dan tidak signifikan ditandai dengan grafik yang semakin landai. Hasil akhir dari proses *training* ini diperoleh model Mask R-CNN yang memiliki nilai *training loss* dan *validation loss* masing – masing 0,0465 dan 0,2892 dengan akurasi *training* model senilai 96,80%.

3.2 Evaluasi *Footprint* Bangunan Hasil Ekstraksi

Proses ekstraksi *footprint* bangunan dilakukan secara otomatis menggunakan model Mask R-CNN yang telah melewati proses *training*. Proses ekstraksi ini membutuhkan waktu 22 menit 13 detik dan menghasilkan *footprint* bangunan dalam model data vektor yang melingkupi seluruh area penelitian.

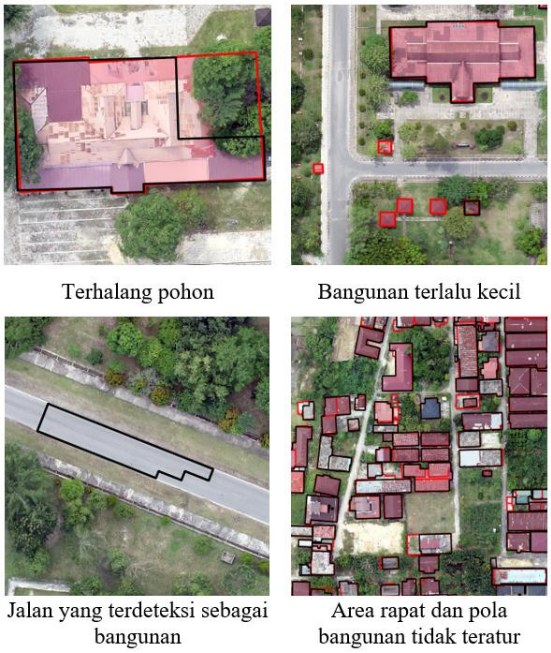
Proses ekstraksi *footprint* bangunan memberikan hasil yang cukup baik. Hal ini dapat terjadi dikarenakan model Mask R-CNN yang digunakan memiliki nilai akurasi *training* yang tinggi.

Gambar 13. Hasil ekstraksi di area renggang bangunan

Gambar 14. Hasil ekstraksi di area rapat bangunan

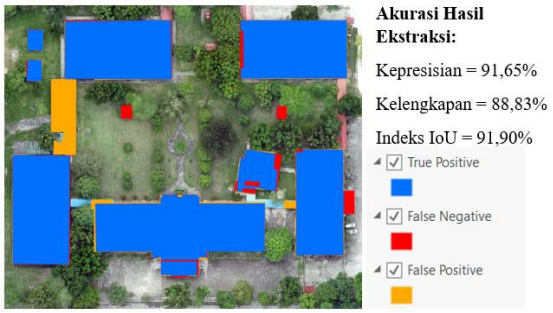
Dapat diamati pada Gambar 13 dan 14 bahwa model Mask R-CNN mampu menghasilkan *footprint* bangunan dengan baik pada area yang memiliki kerapatan rendah, ataupun area yang memiliki tingkat kerapatan bangunan yang relatif tinggi dengan pola bangunan yang relatif teratur. Selain itu, *footprint* yang diperoleh juga memiliki tingkat kedetailan yang tinggi karena mampu mengikuti lekukan bangunan yang bervariasi.

Namun, ekstraksi secara otomatis ini masih terdapat beberapa kekurangan. Kekurangan ekstraksi tersebut disebabkan oleh beberapa faktor seperti, bentuk *footprint* yang tidak sempurna karena bangunan yang terhalang pohon, gagalnya model dalam mendeteksi bangunan karena ukuran bangunan yang kecil dibandingkan ukuran bangunan rerata secara keseluruhan, objek non-bangunan seperti jalan yang terdeteksi sebagai bangunan karena tekstur yang menyerupai bangunan dengan bentuk atap datar, hingga kecacatan hasil ekstraksi karena area yang terlalu rapat dengan pola bangunan yang tidak teratur. Ketidakakuratan hasil ekstraksi ini disebabkan oleh nilai *loss* atau *error* pada model yang telah melalui proses *training* sehingga model tidak mampu melakukan ekstraksi secara sempurna.



Gambar 15. Kesalahan hasil ekstraksi

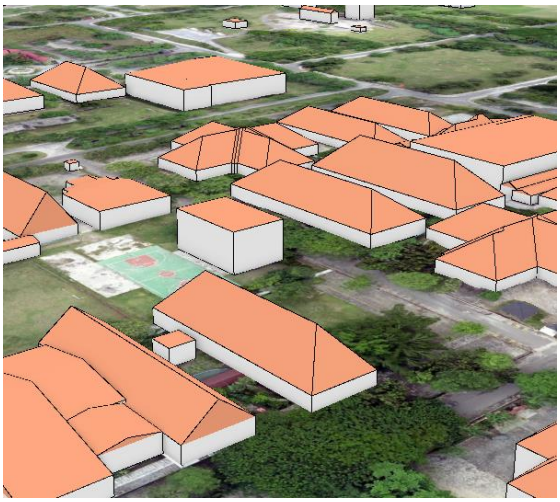
Walaupun terdapat beberapa kesalahan, secara keseluruhan performa model Mask R-CNN dalam melakukan ekstraksi *footprint* bangunan sudah cukup baik. Hal ini dibuktikan dari nilai akurasi yang secara kuantitatif diketahui melalui nilai kepresisian, kelengkapan, dan indeks IoU yang relatif tinggi. Hasil ekstraksi memiliki tingkat kepresisian senilai 91,65% dengan jumlah bangunan yang berhasil diekstraksi memiliki kelengkapan 88,83%, dan tingkat kesamaan antara hasil ekstraksi dengan *ground truth* diketahui melalui nilai indeks IoU yakni senilai 91,90%.



Gambar 16. Akurasi hasil ekstraksi

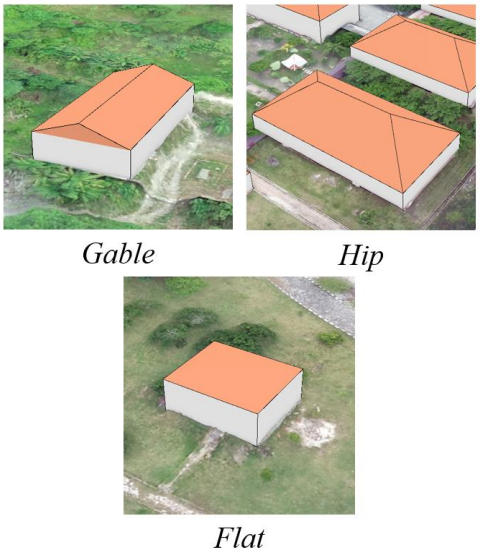
3.3 Evaluasi Model Bangunan 3D

Proses ekstrusi *footprint* bangunan hasil ekstraksi otomatis menggunakan Mask R-CNN dengan data elevasi berupa DSM, DTM, dan nDSM yang berasal dari *dense point cloud* menghasilkan model 3D bangunan sejumlah 620 bangunan dalam waktu 4 menit 38 detik.



Gambar 17. Bangunan hasil pemodelan 3D

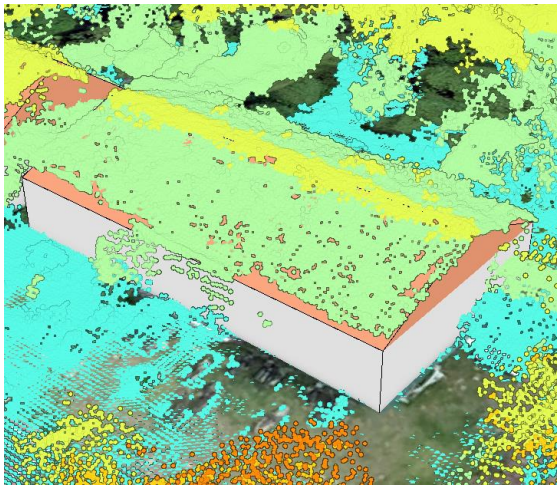
Model bangunan 3D yang dihasilkan sudah memiliki bentuk atap, sehingga berdasarkan tingkat kedetailannya, model bangunan 3D yang diperoleh termasuk dalam LOD 2. Terdapat beberapa kelompok bangunan berdasarkan jenis atapnya, yaitu bangunan dengan jenis atap *hip*, *gable*, *flat*. Bangunan dengan jenis atap *gable* memiliki 2 sisi miring, *hip* memiliki 4 sisi miring, dan *flat* tidak memiliki sisi miring sama sekali, atau juga bisa disebut sebagai atap datar.



Gambar 18. Model bangunan 3D berdasarkan jenis atapnya

Model bangunan 3D yang dihasilkan memiliki beberapa atribut sebagai parameter yang menentukan kesesuaian bentuknya dengan kondisi asli di lapangan. Parameter tersebut terdiri atas jenis atap dan orientasi

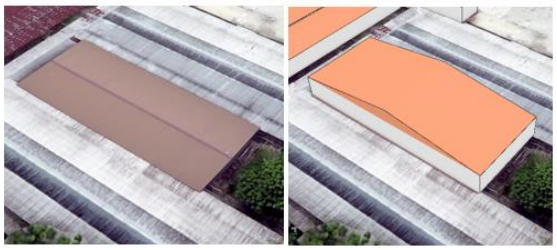
atap. Jika dibandingkan dengan data *dense point cloud* sebagai acuan kondisi sebenarnya di lapangan, terdapat model bangunan 3D yang sesuai, dan juga model bangunan 3D yang tidak sesuai dengan kondisi aslinya. Diamati bahwa terdapat 4 jenis kesalahan yang terjadi dalam pemodelan bangunan 3D kali ini, yaitu kesalahan model 3D karena kesalahan ekstraksi *footprint* bangunan sebelumnya, kesalahan jenis atap, kesalahan orientasi atap, dan kesalahan geometri (RMSE).



Gambar 19. Model 3D dengan atap *gable* yang sesuai dengan kondisi asli di lapangan

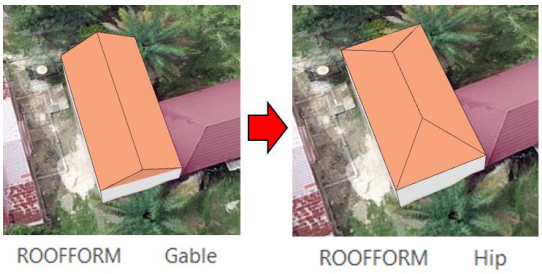
Gambar 20. Kesalahan karena *footprint* bangunan yang tidak tepat

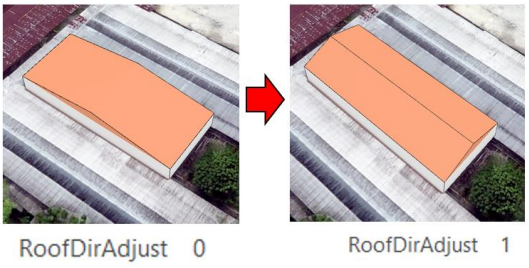
Gambar 21. Kesalahan jenis atap bangunan



Gambar 22. Kesalahan orientasi atap bangunan

Meskipun demikian, kesalahan model bangunan 3D berupa kesalahan jenis atap ataupun orientasi atap masih dapat diperbaiki dengan mengubah nilai masing-masing parameter tersebut secara manual. Sebagai contoh, untuk memperbaiki jenis atap bangunan yang salah, dilakukan perubahan nilai atribut jenis atap menjadi jenis atap yang sesuai, begitu juga untuk kesalahan orientasi atap.





Gambar 22. Perbaikan jenis dan orientasi atap bangunan

Untuk kesalahan geometri dari model bangunan 3D, dapat diketahui melalui nilai akurasi geometri dengan menghitung RMSE. Digunakan sampel ukuran bangunan sebanyak 30 sampel ukuran yang dibandingkan dengan *measurement* langsung pada *dense point cloud*. Ukuran tersebut merupakan sisi-sisi bangunan baik berupa panjang, lebar, ataupun tinggi. Perhitungan akurasi geometri model bangunan 3D ditampilkan pada Tabel 1.

Tabel 1. Uji akurasi geometri model bangunan 3D

No	Ukuran model 3D (m)	Ukuran asli (m)	Selisih	Selisih^2
1	3,67	3,77	-0,1	0,01
2	10,61	10,49	0,12	0,0144
3	8,84	8,41	0,43	0,1849
4	9,33	9,25	0,08	0,0064
5	6,39	5,49	0,9	0,81
6	21,57	22,9	-1,33	1,7689
7	3,13	3,61	-0,48	0,2304
8	21,64	21,25	0,39	0,1521
9	7,84	8,1	-0,26	0,0676
10	21,04	21,57	-0,53	0,2809
11	13,79	14,71	-0,92	0,8464
12	22,26	22,85	-0,59	0,3481
13	47,73	49,38	-1,65	2,7225
14	13,95	14,67	-0,72	0,5184
15	8,31	7,84	0,47	0,2209
16	22,28	23,8	-1,52	2,3104
17	11,55	12,73	-1,18	1,3924
18	37,83	35,26	2,57	6,6049
19	22,63	19,26	3,37	11,3569
20	12,12	13,64	-1,52	2,3104
21	25,46	25,53	-0,07	0,0049
22	12,13	13,16	-1,03	1,0609
23	12,08	13,53	-1,45	2,1025
24	29,52	29,03	0,49	0,2401
25	9,53	10,06	-0,53	0,2809
26	30,94	30,44	0,5	0,25
27	9,94	10,46	-0,52	0,2704
28	48,09	48,1	-0,01	0,0001
29	15,56	14,88	0,68	0,4624
30	9,29	9,78	-0,49	0,2401
			RMSE	**1.111593**

Model bangunan 3D yang dihasilkan memiliki nilai akurasi RMSE 1,11 m. Dengan demikian, model bangunan 3D yang dihasilkan memenuhi standar geometri dari bangunan LOD 2 mengacu kepada standar CityGML yang mensyaratkan model bangunan LOD 2 memiliki RMSE < 2 m.

4. KESIMPULAN DAN SARAN

Dengan menggunakan model Mask R-CNN yang memiliki akurasi *training* senilai 96,80%, diperoleh *footprint* bangunan secara otomatis dengan tingkat kepresisian 91,65%, kelengkapan 88,83%, dan tingkat kesamaan berdasarkan indeks IoU senilai 91,90%. Model bangunan 3D yang dihasilkan melalui teknik ekstrusi *footprint* hasil ekstraksi secara otomatis menggunakan data *dense point cloud* dari foto udara UAV mampu mencapai kriteria LOD 2 karena sudah memiliki bentuk atap dan memiliki nilai akurasi geometri (RMSE) < 2 m.

Untuk meningkatkan kualitas model bangunan 3D yang dihasilkan menggunakan teknik ini, dapat ditambahkan data berupa nDSM dalam proses *training* Mask R-CNN untuk mengurangi terjadinya kesalahan ekstraksi seperti jalan yang diekstraksi sebagai bangunan. Selain itu, dibutuhkan perangkat keras berupa laptop dengan spesifikasi yang sangat tinggi untuk meningkatkan performa model dan menyingkat waktu pemrosesan. Hasil model 3D juga dapat ditingkatkan lagi dengan proses pemberian tekstur bangunan atau bahkan peningkatan LOD bangunan menjadi LOD 3 dengan bantuan data fotogrametri jarak dekat.

UCAPAN TERIMA KASIH

Terimakasih diucapkan kepada Departemen Teknik Geodesi Universitas Gadjah Mada dan Rektorat Universitas Riau atas izin pelaksanaan dan lokasi yang diberikan kepada penulis untuk melakukan penelitian. Penulis juga mengucapkan terimakasih kepada PT Hasanah Surveyor Raya yang membantu dalam penyediaan alat sehingga data yang dibutuhkan dalam penelitian ini dapat direalisasikan.

DAFTAR PUSTAKA

Ammour, N., Alhichri, H., Bazi, Y., Benjdira, B., Alajlan, N., & Zuair, M. (2017). Deep learning approach for car detection in UAV imagery. *Remote Sensing*, *9*(4). https://doi.org/10.3390/rs9040312

Biljecki, F., Ledoux, H., & Stoter, J. (2016). An improved LOD specification for 3D building models. In *Computers, Environment and Urban Systems* (Vol. 59). https://doi.org/10.1016/j.compenvurbsys.2016.04.005

Biljecki, F., Ledoux, H., & Stoter, J. (2017). Generating 3D city models without elevation

data. *Computers, Environment and Urban Systems*, *64*, 1–18. https://doi.org/10.1016/j.compenvurbsys.2017.01.001

Chen, M., & Li, J. (2019). Deep convolutional neural network application on rooftop detection for aerial image. *arXiv*, 0–3.

Cowton, J., Kyriazakis, I., & Bacardit, J. (2019). Automated Individual Pig Localisation, Tracking and Behaviour Metric Extraction Using Deep Learning. *IEEE Access*, *7*, 108049–108060. https://doi.org/10.1109/ACCESS.2019.2933060

de Carvalho, O. L. F., de Carvalho, O. A., Albuquerque, A. O. de, Bem, P. P. de, Silva, C. R., Ferreira, P. H. G., de Moura, R. D. S., Gomes, R. A. T., Guimarães, R. F., & Borges, D. L. (2021). Instance segmentation for large, multi-channel remote sensing imagery using mask-RCNN and a mosaicking approach. *Remote Sensing*, *13*(1), 1–24. https://doi.org/10.3390/rs13010039

Fawcett, D., Azlan, B., Hill, T. C., Kho, L. K., Bennie, J., & Anderson, K. (2019). Unmanned aerial vehicle (UAV) derived structure-from-motion photogrammetry point clouds for oil palm (Elaeis guineensis) canopy segmentation and height estimation. *International Journal of Remote Sensing*, *40*(19), 7538–7560. https://doi.org/10.1080/01431161.2019.1591651

Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *116*, 55–72. https://doi.org/10.1016/j.isprsjprs.2016.03.008

Gröger, G., Kolbe, T. H., Nagel, C., & Häfele, K.-H. (2012). OpenGIS City Geography Markup Language (CityGML) Encoding Standard, Version 2.0.0. *OGC Document No. 12-019*, 344. https://portal.opengeospatial.org/files/?artifact_id=47842

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

Hemanth, D. J., & Estrela, V. V. (2017). *DEEP LEARNING FOR IMAGE PROCESSING APPLICATIONS*. IOS Press.

Khan, S., Rahmani, H., Shah, S. A. A., & Bennamoun, M. (2018). A Guide to Convolutional Neural Networks for Computer Vision. In *Synthesis Lectures on Computer Vision* (Vol. 8, Nomor 1). https://doi.org/10.2200/s00822ed1v01y201712cov015

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

Kwak, E., Habib, A., & Al-Durgham, M. (2011). Model-based automatic 3d building model generation by integrating lidar and aerial images. *Archiwum Fotogrametrii, Kartografii i Teledetekcji*, *22*(January 2015), 187–200.

Miroslaw-Swiatek, D., Szporak-Wasilewska, S., Michalowski, R., Kardel, I., & Grygoruk, M. (2016). Developing an algorithm for enhancement of a digital terrain model for a densely vegetated floodplain wetland. *Journal of Applied Remote Sensing*, *10*(3), 036013. https://doi.org/10.1117/1.jrs.10.036013

Mlambo, R., Woodhouse, I. H., Gerard, F., & Anderson, K. (2017). Structure from motion (SfM) photogrammetry with drone data: A low cost method for monitoring greenhouse gas emissions from forests in developing countries. *Forests*, *8*(3). https://doi.org/10.3390/f8030068

Monteiro, F. C., & Campilho, A. C. (2005). Performance evaluation of image segmentation algorithms. *Handbook of Pattern Recognition and Computer Vision, 3rd Edition*, 525–542. https://doi.org/10.1142/9789812775320_0028

Peeroo, U., Idrees, M. O., & Saeidi, V. (2017). Building extraction for 3D city modelling using airborne laser scanning data and high-resolution aerial photo. *South African Journal of Geomatics*, *6*(3), 363. https://doi.org/10.4314/sajg.v6i3.7

Sornapudi, S., Long, L. R., Almubarak, H., & Antani, S. K. (2018). Deep Learning Nuclei Detection in Digitized Histology Images by Superpixels. *Journal of Pathology Informatics*, *9*(1). https://doi.org/10.4103/jpi.jpi

Uba, N. (2016). *Land Use and Land Cover Classification Using Deep Learning*

Techniques. Arizona State University.

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2020). *Dive Into Deep Learning*. https://doi.org/10.1016/j.jacr.2020.02.005

Zhao, K., Kang, J., Jung, J., & Sohn, G. (2018). Building extraction from satellite images using mask R-CNN with building boundary regularization. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2018-June*, 242–246. https://doi.org/10.1109/CVPRW.2018.00045