Ekstraksi Tapak Bangunan dari Orthophoto Menggunakan Model Mask R-CNN (Studi Kasus: Kelurahan Darmo, Kota Surabaya)

Alfian Bimanjaya, Hepi Hapsari Handayani, dan Mohammad Rohmaneo Darminto
Departemen Teknik Geomatika, Institut Teknologi Sepuluh Nopember (ITS)
e-mail: hapsari@geodesy.its.ac.id

***Abstrak*—Peta dasar skala besar sangat dibutuhkan oleh kota besar/metropolitan seperti Kota Surabaya untuk perencanaan kota dan menunjang pembangunan kota cerdas. Salah satu informasi utama yang paling dibutuhkan dari peta skala besar adalah fitur bangunan. Ekstraksi tapak bangunan sendiri adalah pekerjaan yang sangat menantang karena banyak alasan, termasuk sifat heterogen dari geometri dan spektral bangunan, kompleksitas bangunan yang sulit diprediksi, dan data sensor yang kurang baik (yaitu bayangan, kontras yang buruk, dan perspektif citra yang buruk). Intepretasi yang dilakukan oleh operator secara visual masih merupakan pendekatan yang umum digunakan untuk ekstraksi informasi dari *orthophoto*. Akurasi intepretasi yang dihasilkan tergantung pada keterampilan dan pengalaman dari operator. Sehingga, dapat terjadi inkonsistensi pada data yang dihasilkan oleh operator yang berbeda. Beberapa tahun terakhir ini, ekstraksi otomatis tapak bangunan dari citra satelit resolusi tinggi maupun *orthophoto* menjadi isu penelitian penting dan menantang yang mendapat perhatian lebih besar. Banyak penelitian terbaru telah mengeksplorasi metode deteksi objek berbasis *deep learning* untuk meningkatkan kualitas ekstraksi bangunan. Dalam penelitian ini, penulis menerapkan metode deteksi objek berbasis *Mask Region-based Convolutional Neural Network* (Mask R-CNN) untuk ekstraksi tapak bangunan memanfaatkan *orthophoto* di daerah urban, yaitu Kelurahan Darmo, Kecamatan Wonokromo, Kota Surabaya. Model Mask R-CNN secara kuantitatif sudah cukup baik dalam mendeteksi objek namun kualitas geometri delineasi batas bangunan masih perlu diperbaiki. Beberapa strategi dirancang dan digabungkan dengan model deteksi objek berbasis Mask R-CNN, termasuk segmentasi *orthophoto*, *post-processing* menggunakan alat otomatis yang terdiri dari regularisasi poligon untuk membuat bangunan lebih teratur, remove overlap untuk menghilangkan tumpang tindih antar bangunan, *fill gap* untuk mengisi celah antar bangunan dan integrasi hasil ekstraksi untuk keseluruhan area studi. Metode otomatis yang penulis terapkan menghasilkan kinerja yang baik dengan presisi 91,43%; kelengkapan (*recall*) 82,97%; dan skor-F1 86,99%.**

***Kata Kunci*—*Convolutional Neural Network*, *Deep Learning*, Ekstraksi Bangunan, *Orthophoto*, Peta Skala Besar.**

I. PENDAHULUAN

PETA Rupabumi Indonesia (RBI) adalah peta dasar yang mengintegrasikan wilayah darat dan wilayah laut termasuk wilayah pantai. Pada Undang-Undang nomor 11 tahun 2020 tentang Cipta Kerja Substansi Informasi Geospasial [1]. Dijelaskan bahwa informasi geospasial terdiri dari peta dasar dan peta tematik dengan skala 1:1.000.000 hingga 1:1000. Pembuatan peta dasar skala 1:5.000 hingga 1:1.000.000 mencakup seluruh wilayah NKRI. Sedangkan, peta dasar skala 1:1.000 diperuntukkan wilayah tertentu sesuai kebutuhan, seperti kota besar/metropolitan, wilayah dengan pertumbuhan ekonomi tinggi, dan wilayah rawan bencana. Kota Surabaya sebagai kota terbesar kedua di Indonesia setelah Jakarta termasuk sebagai kota metropolitan yang membutuhkan peta skala besar untuk menunjang perencanaan kota.

Salah satu informasi utama yang paling krusial dan dibutuhkan dari peta skala besar adalah fitur bangunan. Ekstraksi tapak bangunan sendiri adalah pekerjaan yang sangat menantang karena banyak alasan, termasuk sifat heterogen dari geometri dan spektral bangunan, kompleksitas bangunan yang sulit diprediksi, dan data sesor yang kurang baik (yaitu bayangan, kontras yang buruk, dan perspektif citra yang buruk) [2]. Selain itu, dengan desain arsitektur bangunan yang semakin beragam akan semakin menyulitkan proses segmentasi tapak bangunan secara otomatis. Penelitian untuk mendeteksi bangunan telah dipelajari selama beberapa tahun menggunakan berbagai teknologi penginderaan, seperti citra satelit [3], *orthophoto* [4] dan pemindaian LiDAR [5].

Intepretasi secara interaktif yang dilakukan oleh operator manusia secara visual masih merupakan pendekatan utama untuk klasifikasi digital informasi dari *orthophoto*. Akurasi intepretasi yang dihasilkan tergantung pada keterampilan dan pengalaman dari operator. Sehingga, dapat terjadi inkonsistensi pada data yang dihasilkan oleh operator yang berbeda. Intepretasi secara interaktif ini membutuhkan waktu dan sumber daya manusia yang semakin banyak apabila data yang diolah semakin besar [6]. Perkembangan baru-baru ini dari Artificial Intelligence (AI), telah menunjang babak baru studi menuju analisis dan pengenalan citra secara otomatis [7]. Metode ini menghasilkan kinerja yang melampaui ekstraksi informasi fitur secara manual/tradisional di berbagai aplikasi citra digital [8].

Ekstraksi tapak bangunan sebagian besar didasarkan pada *orthophoto* karena resolusi spasial yang baik dalam beberapa tahun terakhir, kemajuan besar telah dicapai dalam *machine learning*, terutama dalam *deep learning* [9]. Segmentasi dan pemetaan tapak bangunan secara otomatis yang akurat dari *orthophoto* bisa dimanfaatkan untuk berbagai aplikasi, termasuk untuk perencanaan kota, manajemen bencana, pemodelan kota, pemetaan nasional dan manajemen kependudukan. Oleh karena itu, mengembangkan metode baru yang lebih cepat dan akurat untuk mengekstrak tapak bangunan dari citra penginderaan jauh resolusi tinggi atau *orthophoto* akan memiliki banyak manfaat [10].

Dalam beberapa tahun terakhir, pendekatan yang didasarkan pada *deep learning*, seperti *Convolutional Neural Networks* (CNNs) dan variannya, menjadi banyak digunakan untuk pendeteksian objek. Wei dkk. [11]

melakukan ekstraksi bangunan menggunakan *Mask Region -based Convolutional Neural Network* (Mask R-CNN) memanfaatkan data *orthophoto* dilanjutkan dengan proses regularisasi. Zhao, Persello, dan Stein [7] menggunakan Mask R-CNN memanfaatkan data *orthophoto*. Pendekatan *deep learning* bisa diterapkan pada data *orthophoto* untuk mengekstraksi tapak bangunan secara otomatis untuk efisiensi waktu, tenaga dan biaya pemrosesan data.

Informasi geospasial objek bangunan dapat digunakan untuk mengetahui struktur dari bangunan serta memperkirakan berbagai atribut yang berkaitan, seperti bentuk, luas, dan tinggi bangunan. Secara tradisional, sebenarnya informasi tapak bangunan bisa didapatkan dari survey langsung ke lapangan menggunakan metode terestris, namun seiring dengan perkembangan teknologi akuisisi dan pengolahan data geospasial, informasi unsur rupabumi dan informasi atribut terkait lainnya bisa didapat dan diproses secara otomatis [10]. Untuk mengembangkan metode ekstraksi unsur rupabumi pada penelitian ini diperlukan data-data yang mendukung, seperti data pelatihan tapak bangunan untuk melatih model *deep learning*. Data tersebut dapat diperoleh dari *ground truth* dan hasil digitasi manual *orthophoto* yang diakusisi pada tahun 2016.

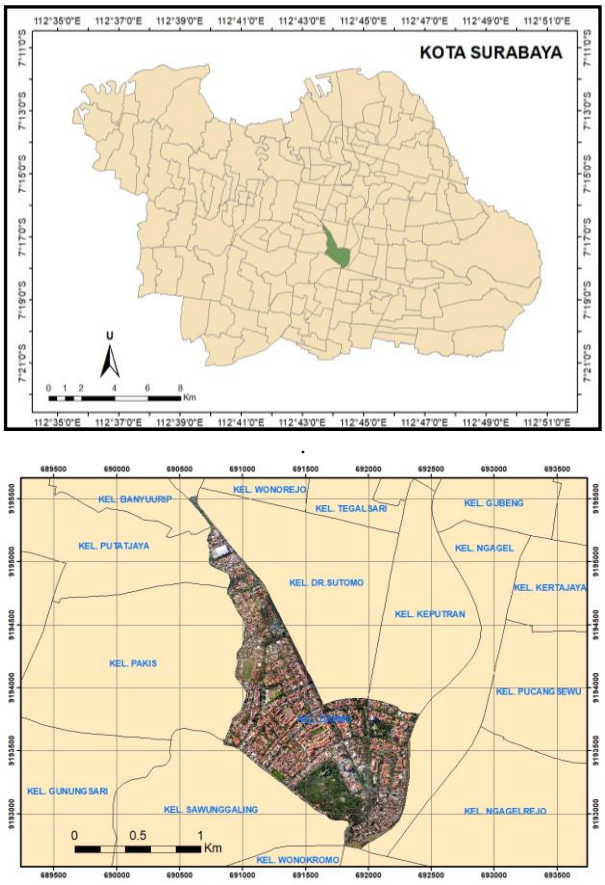
Dalam penelitian ini dilakukan ekstraksi tapak bangunan di Kelurahan Darmo, Kecamatan Wonokromo, Kota Surabaya memanfaatkan data *orthophoto*. Kelurahan Darmo dipilih karena jumlah bangunan yang banyak serta memiliki karakteristik data geospasial yang multi obyek dan multi struktur. Metode yang digunakan adalah penerapan *deep learning* menggunakan model Mask R-CNN dengan *backbone residual network* 50 (*ResNet*-50) untuk mengekstraksi tapak bangunan secara otomatis. Selanjutnya, dilakukan *post-processing* untuk menghaluskan bentuk bangunan supaya lebih teratur. Dilakukan evaluasi kuantitatif dengan presisi, kelengkapan, dan matriks F hasil ekstraksi tapak bangunan untuk menilai performa dari model yang diterapkan pada area yang memiliki heterogenitas tinggi, seperti di Kelurahan Darmo. Dari penelitian ini diharapkan bisa menjadi solusi untuk percepatan pemetaan skala besar sebagai penunjang dan rekomendasi pengambilan kebijakan terkait dengan perencanaan Kota Surabaya.

II. MATERIAL DAN METODE

A. Studi Area

Lokasi penelitian ini dilakukan di Kelurahan Darmo, Kecamatan Wonokromo, Kota Surabaya (7°16'28"–7°18'0" LS dan 112°43'34"–112°44'34" BT). Lokasi ini dipilih karena jumlah bangunan dan infrastruktur yang banyak serta memiliki karakteristik data geospasial yang multi obyek dan multi struktur. Area penelitian ditunjukkan pada Gambar 1.

Data *orthophoto* ini memiliki *ground sample distance* (GSD) 8 cm/pixel yang diakuisisi pada tahun 2016 dengan kamera udara metrik yang terpasang pada pesawat yang terbang sekitar 800 meter di atas permukaan area studi. Data ini didapatkan dari Dinas Perumahan Rakyat dan Kawasan Permukiman, Cipta Karya dan Tata Ruang, Kota Surabaya. Untuk melatih model Mask R-CNN penulis menggunakan data poligon tapak bangunan dari dua nomor lembar peta (NLP) di area studi.



Gambar 1. Peta lokasi penelitian di Kelurahan Darmo, Kecamatan Wonokromo, Kota Surabaya.

B. Persiapan Data

Dalam penelitian ini dilakukan ekstraksi tapak bangunan di Kelurahan Darmo, Kecamatan Wonokromo, Kota Surabaya memanfaatkan data *orthophoto* yang didapatkan dari Dinas Perumahan Rakyat dan Kawasan Permukiman, Cipta Karya dan Tata Ruang, Kota Surabaya. Data *orthophoto* pada area studi dibagi menjadi 5 zona untuk meringankan pemrosesan. Selanjutnya data *orthophoto* dilakukan segmentasi untuk menyederhanakan variasi spektral tanpa mengurangi resolusi spasial sehingga dapat memudahkan model untuk mendeteksi objek. Proses segmentasi menggunakan beberapa ukuran homogenitas melalui pengelompokan piksel yang berdekatan dengan nilai atribut yang sama. Hal tersebut tidak hanya diperoleh dari informasi spektral, tetapi juga geometrik dan tekstur [12].

Algoritma segmentasi mengelompokkan individu piksel ke dalam segmen sesuai dengan kriteria berikut [13]: homogenitas dalam segmen, kemampuan untuk dipisahkan dari elemen yang berdekatan, dan bentuk homogenitas. Tiga kriteria tersebut kadang-kadang saling bertentangan dan tidak dapat dipenuhi pada saat yang sama. Oleh karena itu, algoritma segmentasi menekankan satu atau dua saja dari tiga kriteria yang ada [14].

C. Model Mask R-CNN untuk Ekstraksi Bangunan

Dalam beberapa tahun terakhir, algoritma *deep learning* (DL) telah mengungguli kemampuan komputasi *Machine Learning* ke tingkat presisi dan performa yang lebih tinggi dengan meningkatkan jumlah lapisan atau kedalaman [15]. Algoritma *deep learning* seperti *Convolutional Neural Networks* (CNNs) telah dikenal luas sebagai pendekatan yang menonjol untuk banyak aplikasi computer vision

(pengenalan, ekstraksi, dan klasifikasi gambar / video) dan telah menunjukkan hasil yang luar biasa di banyak aplikasi [16]. *Deep Learning* memungkinkan ekstraksi fitur yang cepat dan otomatis dari dataset yang cukup besar secara berulang menggunakan model yang kompleks untuk mengurangi kesalahan klasifikasi menggunakan metode regresi [17].

Deep learning telah menjadi metode inti dalam banyak penelitian mengenai ekstraksi tapak bangunan dan jaringan jalan. Wei dkk. [11] melakukan ekstraksi bangunan menggunakan Mask R-CNN memanfaatkan *orthophoto* dilanjutkan dengan proses regularisasi. Zhao, Persello, dan Stein [7] menggunakan Mask R-CNN memanfaatkan *orthophoto*. Pendekatan *deep learning* bisa diterapkan pada *orthophoto* untuk mendeteksi unsur rupabumi, seperti bangunan dan jalan. Ketidakteraturan bentuk, sulitnya memodelkan objek bangunan menjadi celah untuk meneliti dan mencari pendekatan strategis untuk otomatiasi ekstraksi tapak bangunan dan jaringan jalan memanfaatkan *orthophoto*, dengan metode pemrosesan *deep learning*.

Faster R-CNN [18] dan Mask R-CNN [19] umumnya dianggap sebagai arsitektur mutakhir untuk ekstraksi objek dan segmentasi instan. Kedua arsitektur ini adalah bagian dari keluarga CNN berbasis region. Faster R-CNN secara efektif adalah dua jaringan; jaringan *Region Proposal Network* (RPN) dan jaringan pendeteksi Fast R-CNN. RPN menggantikan metode pencarian selektif di versi sebelumnya (R-CNN dan Fast R-CNN) yang pemrosesannya lebih memakan waktu. Dalam Faster R-CNN, RPN menyelesaikan ekstraksi dengan menggunakan peta fitur yang diturunkan dari lapisan konvolusi terakhir di *backbone* CNN (yaitu VGG16/ResNet) sebagai proposal untuk jaringan detektor. Ini memungkinkan arsitektur CNN digunakan untuk proposal dan klasifikasi region. Jaringan kemudian melakukan *pooling Region of Interest* (RoI), lapisan yang terhubung penuh dan klasifikasi seperti di R-CNN [2]. RPN menggunakan jangkar untuk menghasilkan kotak pembatas potensial bersama dengan skor yang menentukan seberapa besar kemungkinan setiap kotak ekstraksi objek. Lokasi proposal awal ditentukan oleh penyertaan jangkar. Jangkar adalah kotak proposal tertentu dengan pusat x, y, yang ditentukan oleh posisi jendela di atas peta fitur pada setiap langkah yang diberikan. Fungsi *loss* halus L1 digunakan untuk pelatihan dan didefinisikan sebagai:

$$L(p_1,t_i) = \frac{1}{N_{cls}}\sum L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}}\sum p_i^* L_{reg}(t_i,t_i^*) \quad (1)$$

Di mana *i* adalah indeks jangkar dalam *batch* kecil, p_i adalah kemungkinan proposal menjadi objek, t_i adalah koordinat kotak pembatas yang diprediksi, L_{cls} adalah kehilangan log, p_i^* adalah label objek kebenaran dari ground truth, L_{reg} adalah *loss* halus L_1, dan t_i^* adalah koordinat kotak yang sebenarnya. Fungsi *loss* dari *regressor* kemudian dapat didefinisikan sebagai:

$$SmoothL_1(x) = \begin{cases} 0{,}5x^2 if|x| < 1 \\ |x| - 0{,}5 otherwise \end{cases} \quad (2)$$

Mask Region-based Convolutional Neural Network (Mask R-CNN) diperkenalkan sebagai perpanjangan dari Faster R-CNN untuk memungkinkan segmentasi berbasis piksel yang akurat [19]. Mask R-CNN terdiri dari dua tahap utama, yaitu *Feature Pyramid Network* (FPN) dan *Region Proposal Network* (RPN). Dalam FPN, sejumlah proposal berbeda telah dibuat tentang prediksi objek berdasarkan masukan citra. Mask R-CNN dibuat langsung di Faster R-CNN dengan menambahkan cabang keluaran ketiga untuk proposal *mask* setiap objek kandidat. Oleh karena itu, model ini dapat dilatih secara paralel dengan jaringan Faster R-CNN, yang menghasilkan *mask* yang berpasangan untuk setiap RoI. Fungsi *loss* diperbarui untuk kemudian termasuk prediksi *mask* sehingga; $L = L_{cls} + L_{box} + L_{mask}$, di mana L adalah total *loss*.

D. Evaluasi Kuantitatif

Ekstraksi tapak bangunan dievaluasi dengan matriks presisi (precision), kelengkapan (recall), dan skor-F1 yang dihitung dari perpotongan/interseksi antara objek bangunan terdeteksi dengan bangunan ground truth. Interseksi didefinisikan sebagai intersection over union (IoU) untuk mengevaluasi keakuratan poligon bangunan yang terdeteksi. IoU sama dengan luas persimpangan poligon bangunan yang terdeteksi (dilambangkan dengan A) dan poligon bangunan *ground truth* (dilambangkan dengan oleh B) dibagi dengan luas gabungan A dan B. Jika poligon bangunan yang terdeteksi berpotongan dengan lebih dari satu poligon bangunan ground truth, maka bangunan *ground truth* dengan nilai IoU tertinggi akan dipilih:

$$IoU = \frac{Area(A \cap B)}{Area(A \cup B)} \quad (3)$$

Setiap satuan objek bangunan terdeteksi yang memiliki interseksi dengan satuan objek bangunan *ground truth* akan dianggap sebagai ekstraksi yang berhasil (*true positive*), jika satuan objek bangunan terdeteksi tidak berinterseksi dengan satuan objek bangunan *ground truth* maka dihitung sebagai kesalahan ekstraksi (*false positive*). Dan sebaliknya, jika satuan objek bangunan *ground truth* tidak berinterseksi dengan satuan objek bangunan terdeteksi maka dihitung sebagai kesalahan ekstraksi (*false negative*). Maka dilakukan evaluasi dengan persamaan berikut:

$$precision = \frac{TP}{TP + FP} \quad (4)$$

$$recall = \frac{TP}{TP + FN} \quad (5)$$

$$F_1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (6)$$

Di mana TP adalah *True Positive*, FP adalah *False Positive*, dan FN adalah *False Negative*.

III. HASIL DAN ANALISA

A. Pelatihan Model Mask R-CNN untuk Ekstraksi Bangunan

Pada penelitian ini, penulis mengimplementasikan Mask-RCNN menggunakan *backbone residual network* 50 (ResNet-50) yang dikenalkan oleh He dkk. [20] dibandingkan menggunakan *backbone* VGG-16. Modul residual dari ResNet dengan pintasan yang memungkinkan untuk mengambil aktivasi dari satu lapisan dan memasukkannya ke lapisan lain yang berjarak sekitar 2-3 lompatan sehingga sangat efektif dalam mengurangi masalah degradasi yang ditimbulkan oleh jaringan yang lebih dalam.

Model dilatih menggunakan dataset pelatihan poligon tapak bangunan dari dua Nomor Lembar Peta (NLP) di Kelurahan Darmo, Kecamatan Wonokromo, Kota Surabaya sebanyak 73 *epoch*, dengan ukuran *batch* 2. Waktu yang dibutuhkan untuk melatih model ini pada GPU NVIDIA GeForce GTX 1060 6GB adalah selama 93 jam 25 menit 28 detik. Grafik *loss* yang menggambarkan proses pelatihan model Mask R-CNN untuk ekstraksi tapak bangunan terhadap dataset dapat dilihat pada Gambar 2. Dataset pada area studi secara acak dibagi menjadi 70% sampel pelatihan dan 30% sampel validasi untuk jaringan model ekstraksi bangunan. Jumlah data sampel pelatihan dan validasi dapat dilihat pada Tabel 1.

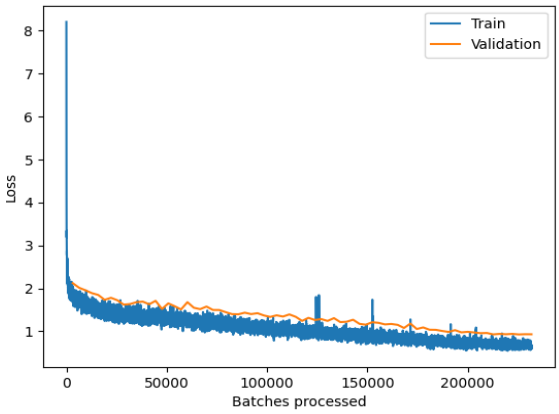
B. Ekstraksi Tapak Bangunan

Model yang telah terlatih digunakan untuk mendeteksi tapak bangunan untuk keseluruhan Kelurahan Darmo. Hasil ekstraksi bangunan dapat dilihat pada Gambar 3. Terlihat bahwa model sudah cukup baik dalam mendeteksi bangunan namun delineasi batas bangunan masih perlu diperbaiki. Meskipun Mask R-CNN dapat melakukan segmentasi tingkat objek, hasil ekstraksi tapak bangunan masih perlu dilakukan *post-processing* untuk membuat bangunan lebih teratur. Ini karena Mask R-CNN kehilangan detail ketika secara langsung melakukan *up-sampling* peta fitur yang sangat kecil ke ukuran penuh dari citra *input* [11].

C. Integrasi dan Post-Processing

Model *deep learning* Mask R-CNN yang telah terlatih sebelumnya akan digunakan untuk mendeteksi bangunan pada 5 zona di area studi. Selanjutnya, 5 peta tapak bangunan hasil ekstraksi menggunakan model Mask R-CNN tersebut diintegrasikan menjadi satu peta, ditunjukkan oleh Gambar 4. Diperlukan *post-processing* untuk memperbaiki bentuk geometri dari tapak bangunan yang sudah terdeteksi. Hasil ekstraksi tapak bangunan menggunakan model Mask R-CNN masih kurang akurat sehingga perlu dilakukan regularisasi bangunan untuk membuat bentuk geometri bangunan lebih teratur [11]. Algoritma regularisasi poligon efektif untuk penyesuaian batas antar bangunan yang tersegmentasi menjadi tapak bangunan terstruktur [21]. Mulai dari poligon awal hasil ekstraksi, algoritma regularisasi batas pada penelitian ini terdiri dari dua langkah: penyesuaian kasar yang menghilangkan kesalahan, nyata dari ekstraksi dan penyempurnaan yang menyesuaikan arah garis dan posisi sudut.

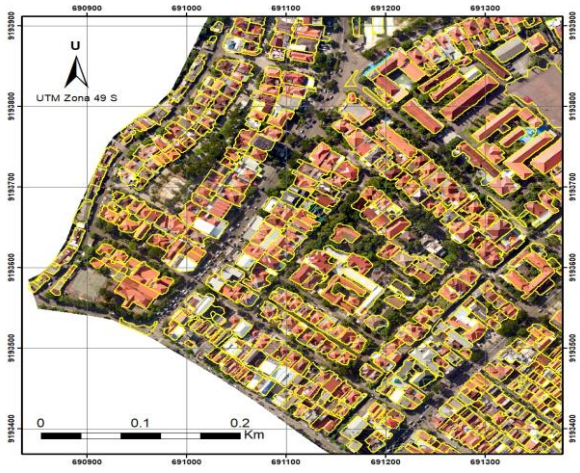
Dilanjutkan dengan proses *remove overlap* untuk menghilangkan tumpang tindih antar bangunan, dan proses



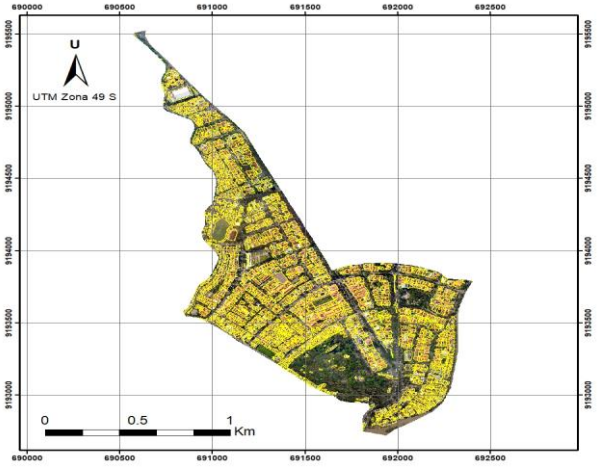
Gambar 2. Grafik *training loss* vs *validation loss* pelatihan model Mask R-CNN untuk deteksi tapak bangunan.

Tabel 1.
Perbandingan data pelatihan dan data validasi

Jenis Dataset	Presentase (%)	Jumlah data
Data pelatihan	70%	630
Data validasi	30%	270

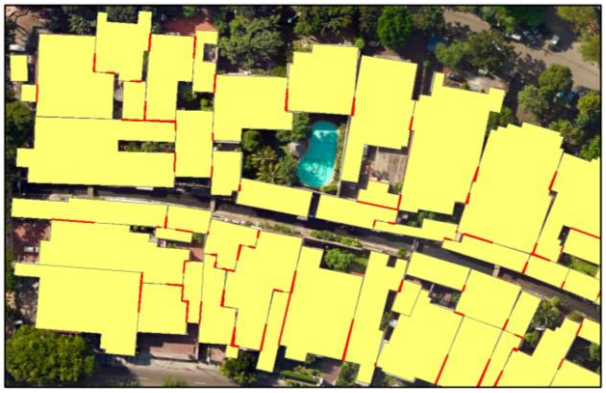


Gambar 3. Hasil deteksi bangunan sebelum dilakukan *post-processing.*



Gambar 4. Hasil deteksi bangunan sebelum dilakukan *post-processing.*

fill gap untuk mengisi celah antar bangunan seperti yang ditunjukkan pada Gambar 5. Maksud dari proses menghilangkan tumpang tindih dan celah di antara poligon bangunan adalah untuk memperbaiki topologi dari objek bangunan yang terdeteksi. Aturan topologi sendiri



Gambar 5. Hasil deteksi bangunan setelah dilakukan regularisasi namun masih terdapat tumpang tindih dan celah antar bangunan ditunjukkan area berwarna merah.

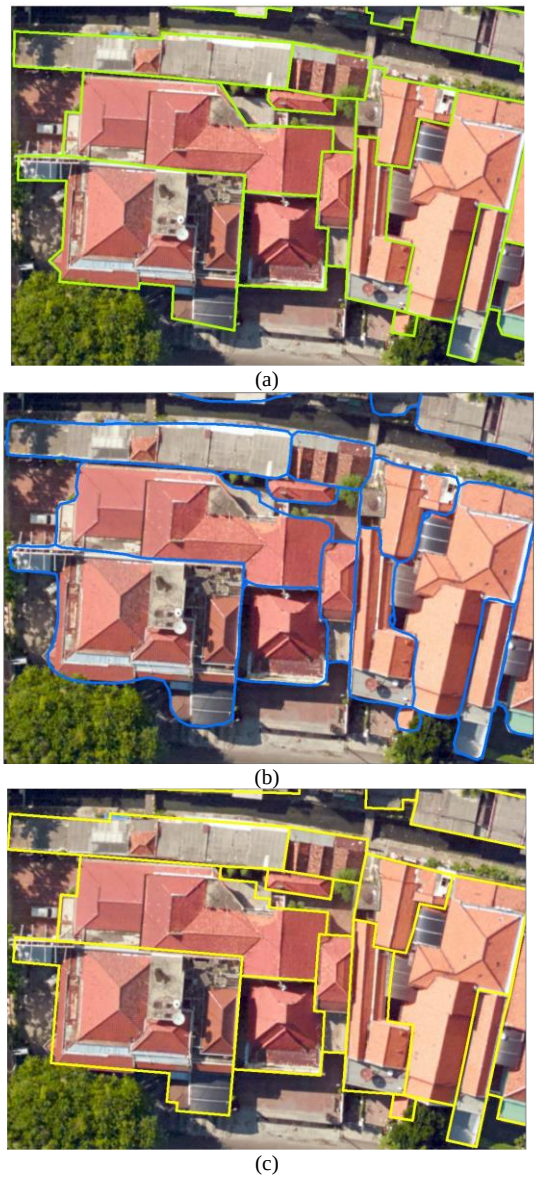
digunakan untuk memodelkan hubungan spasial antara kelas fitur dalam suatu dataset dan memastikan fitur tersebut konsisten dalam perilaku dan hubungan spasialnya [22]. Semua alur *post-processing* tapak bangunan ini dilakukan sebanyak 3 iterasi. Perbandingan hasil ekstraksi tapak bangunan sebelum dan sesudah *post-processing* ditunjukkan pada Gambar 6.

D. Evaluasi Model Secara Kuantitatif

Penulis menganalisis dan mengevaluasi secara kuantitatif model Mask R-CNN yang digunakan untuk mengekstraksi tapak bangunan di Kelurahan Darmo, Kecamatan Wonokromo, Kota Surabaya. Hasil ekstraksi tapak bangunan akan diuji dengan dataset validasi berdasarkan interseksi setiap bangunan antara keduanya. Intersection over union (IoU) digunakan untuk mengevaluasi keakuratan model Mask R-CNN dalam mendeteksi bangunan. Dalam penelitian ini penulis menggunakan IoU 50% yang berarti objek dianggap positif jika terdapat interseksi minimal 50% antara poligon bangunan hasil ekstraksi dengan bangunan validasi.

Presisi (precision), kelengkapan (recall), dan skor-F1 dihitung berdasarkan Persamaan 3-6, dimana true positive (TP) menunjukkan jumlah poligon bangunan yang terdeteksi dengan benar, *false* positive (FP) menunjukkan jumlah objek lain yang terdeteksi sebagai poligon bangunan karena kesalahan, dan *false* negative (FN) menunjukkan jumlah poligon bangunan yang tidak terdeteksi. Poligon bangunan akan dinilai sebagai terdeteksi dengan benar jika IoU antara poligon bangunan yang terdeteksi dan poligon bangunan *ground truth* lebih besar dari 50%. Evaluasi kuantitatif hasil ekstraksi tapak bangunan menggunakan model Mask R-CNN dapat dilihat pada Tabel 2.

Berdasarkan evaluasi model secara kuantitatif tersebut telihat bahwa bangunan tidak terdeteksi yang termasuk *false* negative (FN) terdapat 46 bangunan. Hal ini dikarenakan stuktur pola bangunan di Kelurahan Darmo, khususnya di permukiman yang saling berhimpit satu sama lain sehingga menyulitkan model untuk melakukan segmentasi batas antar bangunan. Selain itu, variasi warna dan pola atap bangunan yang sangat beragam juga menjadi salah satu faktor penyebab kesalahan ekstraksi. Gambar 7 menunjukkan beberapa contoh hasil ekstraksi tapak bangunan dari model yang penulis gunakan, di mana poligon hijau menunjukkan true positive (TP), poligon merah menunjukkan false positive (FP), dan kuning menunjukkan false negative (FN).



(a) (b) (c)

Gambar 6. Perbandingan data bangunan validasi dengan bangunan hasil deteksi menggunakan model Mask R-CNN. (a) data bangunan validasi, (b) hasil deteksi bangunan sebelum *post-processing*, dan (c) hasil deteksi bangunan setelah *post-processing.*

Tabel 2.
Evaluasi kuantitatif hasil deteksi tapak bangunan menggunakan model Mask R-CNN

TP	FP	FN	*Precision*	*Recall*	F_1-*score*
224	21	46	91,43%	82,97%	86,99%

Gambar 7. Poligon hijau menunjukkan *true positive* (TP), poligon merah menunjukkan *false positive* (FP), dan kuning menunjukkan *false negative* (FN).

IV. KESIMPULAN

Dalam penelitian ini, penulis menerapkan metode deteksi objek berbasis *Mask Region-based Convolutional Neural Network* (Mask R-CNN) untuk ekstraksi tapak bangunan memanfaatkan *orthophoto* di daerah urban, yaitu Kelurahan Darmo, Kecamatan Wonokromo, Kota Surabaya. Penulis melatih model Mask R-CNN untuk ekstraksi tapak bangunan menggunakan *orthophoto* dan data poligon tapak bangunan dari dua nomor lembar peta (NLP) di area studi. Beberapa strategi dirancang dan digabungkan dengan model deteksi objek berbasis Mask R-CNN, termasuk segmentasi *orthophoto*, yang terdiri dari regularisasi poligon untuk membuat bangunan lebih teratur, remove overlap untuk menghilangkan tumpang tindih antar bangunan, *fill gap* untuk mengisi celah antar bangunan dan integrasi hasil ekstraksi untuk keseluruhan area studi untuk meningkatkan kualitas hasil ekstraksi.

Hasil penelitian ini menunjukkan bahwa metode yang diterapkan berkinerja cukup baik dengan nilai presisi 91,43%; kelengkapan (recall) 82,97%; dan skor-F1 86,99%. Beberapa bias dan kesalahan yang tersisa serta ketidakakuratan poligon bangunan yang dihasilkan umumnya disebabkan oleh pepohonan yang menutupi sebagian bangunan, atap bangunan yang warnanya sangat bervariasi, pekerjaan konstruksi yang sedang berlangsung dan struktur bangunan yang kompleks. Dalam penelitian masa depan, penulis menyarankan untuk menggabungkan *orthophoto* dengan data elevasi permukaan untuk meningkatkan ketelitian dan memperkaya informasi hasil ekstraksi karena ada informasi ketinggian. Penulis juga menyarankan pembuatan *toolbox* untuk *post-processing* sehingga beberapa proses perbaikan hasil ekstraksi objek bisa dilakukan sekaligus dengan satu tool dalam sekali jalan. Metode yang penulis terapkan belum bisa sepenuhnya mendapatkan hasil yang baik, perlu adanya quality control dan perbaikan manual oleh operator untuk mendapatkan informasi geospasial tapak bangunan yang akurat. Meskipun begitu, metode ini masih mengungguli kinerja operator dalam hal efisiensi waktu, tenaga dan biaya pemrosesan data. Oleh karena itu, penulis percaya bahwa metode yang penulis terapkan pada penelitian ini memiliki potensi besar untuk bisa menunjang percepatan penyediaan peta dasar skala besar di Indonesia.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah membantu pelaksanaan penelitian ini, terutama kepada Dinas Perumahan Rakyat dan Kawasan Permukiman, Cipta Karya dan Tata Ruang, Kota Surabaya yang telah menyediakan data spasial utama penelitian ini berupa *orthophoto* untuk area penelitian di Kelurahan Darmo, Kecamatan Wonokromo, Kota Surabaya.

DAFTAR PUSTAKA

[1] Pemerintah Republik Indonesia, *Undang-Undang Republik Indonesia Nomor 11 Tahun 2020 tentang Cipta Kerja*. Jakarta: Pemerintah Republik Indonesia, 2020.

[2] D. Griffiths and J. Boehm, "Improving public data for building segmentation from convolutional neural networks (CNNs) for fused airborne lidar and image data using active contours," *ISPRS J. Photogramm. Remote Sens.*, vol. 154, pp. 70–83, 2019.

[3] W. Li, C. He, J. Fang, J. Zheng, H. Fu, and L. Yu, "Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source GIS data," *Remote Sens.*, vol. 11, no. 4, p. 403, 2019.

[4] D. Yu, S. Ji, J. Liu, and S. Wei, "Automatic 3D building reconstruction from multi-view aerial images with deep learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 171, pp. 155–170, 2021.

[5] S. Du, Y. Zhang, Z. Zou, S. Xu, X. He, and S. Chen, "Automatic building extraction from LiDAR data fusion of point and grid-based features," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 294–307, 2017.

[6] Y. Shi, Q. Li, and X. X. Zhu, "Building segmentation through a gated graph convolutional neural network with deep structured feature embedding," *ISPRS J. Photogramm. Remote Sens.*, vol. 159, pp. 184–197, 2020.

[7] W. Zhao, C. Persello, and A. Stein, "Building outline delineation: From aerial images to polygons with an improved end-to-end learning framework," *ISPRS J. Photogramm. Remote Sens.*, vol. 175, pp. 119–131, 2021.

[8] Xiao Xiang Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, and F. Xu, "Deep learning in remote sensing: a comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36.

[9] G. Pasquali, G. C. Iannelli, and F. Dell'Acqua, "Building footprint extraction from multispectral, spaceborne earth observation datasets using a structurally optimized u-net convolutional neural network," *Remote Sens.*, vol. 11, no. 23, p. 2803, 2019.

[10] Y. Xu, Z. Xie, Y. Feng, and Z. Chen, "Road extraction from high-resolution remote sensing imagery using deep learning," *Remote Sens.*, vol. 10, no. 9, p. 1461, 2018.

[11] S. Wei, S. Ji, and M. Lu, "Toward automatic building footprint delineation from aerial images using CNN and regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 2178–2189, 2020.

[12] Q. Yu, P. Gong, N. Clinton, G. S. Biging, M. Kelly, and D. Schirokauer, "Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery," *Photogramm Eng Remote Sens.*, vol. 72, no. 7, pp. 799–811, 2006.

[13] T. Veljanovski, U. Kanjir, and K. Oštir, "Object-based image analysis of remote sensing data," *Geod. vestn.*, vol. 55, no. 04, pp. 641–664, 2011.

[14] Nussbaum, Sven, Menz, and Gunter, *Object-Based Image Analysis and Treaty Verification: New Approaches in Remote Sensing - Applied to Nuclear Facilities in Iran*. New York: Springer, 2008.

[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[16] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, 2018.

[17] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 779–788, 2016.

[19] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Int. Conf. Comput. Vis.*, pp. 2980–2988, 2017.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv:1512.03385*, 2015. .

[21] K. Zhao, J. Kang, J. Jung, and G. Sohn, "Building extraction from satellite images using mask R-CNN with building boundary regularization," *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 242–246, 2018.

[22] D. B. Susetyo, D. Nuraeni, and A. P. Perdana, "Aturan topologi untuk unsur perairan dalam skema basis data spasial rupabumi Indonesia," in *Seminar Nasional II Pengelolaan Pesisir dan Daerah Aliran Sungai*, 2016, p. 11.