

Penggunaan *Deep Learning* dan *Post-Processing* Algoritma Douglas-Peucker untuk Ekstraksi Jaringan Jalan pada Area Urban dari *Orthophoto*

The Use of Deep Learning and Douglas-Peucker Post-Processing Algorithm for Road Network Extraction in Urban Area from Orthophoto

Alfian Bimanjaya¹, Hepsari Handayani^{1*}, Reza Fuad Rachmadi²

¹Departemen Teknik Geomatika, FTSPK-ITS, Kampus ITS Sukolilo, Surabaya, 60111, Indonesia

²Departemen Teknik Komputer, FTEIC-ITS, Kampus ITS Sukolilo, Surabaya, 60111, Indonesia

*Korespondensi penulis: hapsari@geodesy.its.ac.id

Diterima: 01022022; Diperbaiki: 22042024; Disetujui: 18072024; Dipublikasi: 18072024

Abstrak: Peta dasar skala besar sangat dibutuhkan oleh kota besar/metropolitan seperti Kota Surabaya untuk perencanaan kota dan menunjang pembangunan kota cerdas. Beberapa informasi utama yang paling dibutuhkan dari peta skala besar adalah fitur bangunan dan jaringan jalan. Ekstraksi jaringan jalan merupakan pekerjaan yang menantang karena banyak alasan, termasuk sifat heterogen dari geometri dan spektral, kompleksitas objek yang sulit dimodelkan, dan data sensor yang kurang baik. Interpretasi yang dilakukan oleh operator secara visual masih merupakan pendekatan yang umum digunakan untuk ekstraksi informasi dari *orthophoto*. Akurasi interpretasi 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 jalan dari *orthophoto* maupun CSRT menjadi isu penelitian penting dan menantang yang mendapat perhatian lebih besar. Dalam penelitian ini, penulis menerapkan metode deteksi objek berbasis *Mask Region-based Convolutional Neural Network* (Mask R-CNN) untuk ekstraksi jaringan jalan memanfaatkan *orthophoto* dan DSM LiDAR di daerah urban Kota Surabaya. Beberapa strategi dirancang dan digabungkan dengan model deteksi objek berbasis Mask R-CNN, termasuk *post-processing* yang terdiri dari regularisasi poligon algoritma Douglass-Peucker, *remove overlap*, *fill gap*, dan penghalusan poligon. Metode yang penulis terapkan menghasilkan kinerja yang cukup baik untuk ekstraksi jalan menghasilkan nilai presisi 90,28%; kelengkapan (*recall*) 85,85%; skor-F1 88,01%; dan IoU 78,59%; serta *overall accuracy* 95,25 % dan nilai kappa 90,5%.

Copyright © 2024 Geoid. All rights reserved.

Abstract: A Large-scale base map is needed by metropolitan cities such as Surabaya City for urban planning and smart city development. Some of the most needed information from a large-scale base map are building and road geospatial information. Road network extraction is a challenging task for many reasons, including the heterogeneous attribute of the geometry and spectral, the complexity of objects that are difficult to model, and poor sensor data. Visual interpretation by operators is still a commonly used approach for extracting information from *orthophoto*. Interpretation accuracy depends on the skill and experience of the operator. So, there can be inconsistencies in the data generated by different operators. In recent years, the automatic extraction of road from *orthophoto* or VHR image has become an important and challenging research issue. Many recent studies have explored deep learning to improve the quality of building and road extraction. In this study, we applied a *Mask Region-based Convolutional Neural Network* (Mask R-CNN) model for the building footprint and the road network extraction using *orthophoto* and DSM in urban area in Surabaya City. The quality of the extracted geometry needs to be improved. Several *post-processing* strategies, including polygon regularization using Douglas-Peucker algorithm and polygon smoothing are designed to achieve optimal extraction results. The method produces a good performance for road extraction, the precision is 90.28%; recall 85.85%; F1-score 88.01%; and IoU 78.59%; and the overall accuracy is 95.25% and the kappa value is 90.5%.

Kata kunci: *Deep Learning*, Ekstraksi Objek, *Orthophoto*, Peta Skala Besar, Algoritma Douglas-Peucker

Cara untuk sitasi: Bimanjaya, A., Hapsari, H.H., & Rachmadi R.F. (2024). Penggunaan Deep Learning dan Post-Processing Algoritma Douglas-Peucker untuk Ekstraksi Jaringan Jalan pada Area Urban dari *Orthophoto*. *Geoid*, 19(2), 371 - 385.

Pendahuluan

Peta Rupabumi Indonesia (RBI) merupakan peta dasar yang mengintegrasikan wilayah darat dan wilayah laut termasuk wilayah pantai. Kebutuhan peta skala besar merupakan salah satu komponen informasi geospasial dasar yang penting, mengingat berdasarkan Undang-Undang nomor 11 tahun 2020 tentang Cipta Kerja Substansi Informasi Geospasial yang menyebutkan bahwa segala aspek kebijakan pembangunan yang terkait dengan aspek keruangan harus didasari oleh informasi geospasial yang dapat dipertanggungjawabkan. Peta skala besar sangat berguna sebagai sistem pendukung pengambilan kebijakan dalam rangka mengoptimalkan pembangunan, khususnya 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 perencanaan kota yang menunjang pembangunan kota cerdas.

Salah satu informasi utama yang paling krusial dan dibutuhkan dari peta skala besar adalah jaringan jalan. Ekstraksi jaringan jalan sendiri merupakan pekerjaan yang sangat menantang karena fitur spektral yang lebih beragam dan fitur geometris yang lebih detail pada citra beresolusi tinggi. Secara tradisional, informasi jaringan jalan 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 (Xu dkk., 2018). Metode ekstraksi jalan yang dikerjakan secara tradisional juga memiliki kelemahan apabila objek mempunyai karakteristik jalan yang kompleks. Radian dan kelengkungan jalan yang rumit mempersulit metode tradisional untuk mengekstraksi persimpangan jalan yang utuh dan berkesinambungan (Zhu dkk., 2021). Selain itu, objek jalan dalam citra yang tertutupi oleh pepohonan atau objek tinggi lain, bayangannya akan semakin menyulitkan proses ekstraksi (Gao dkk., 2018).

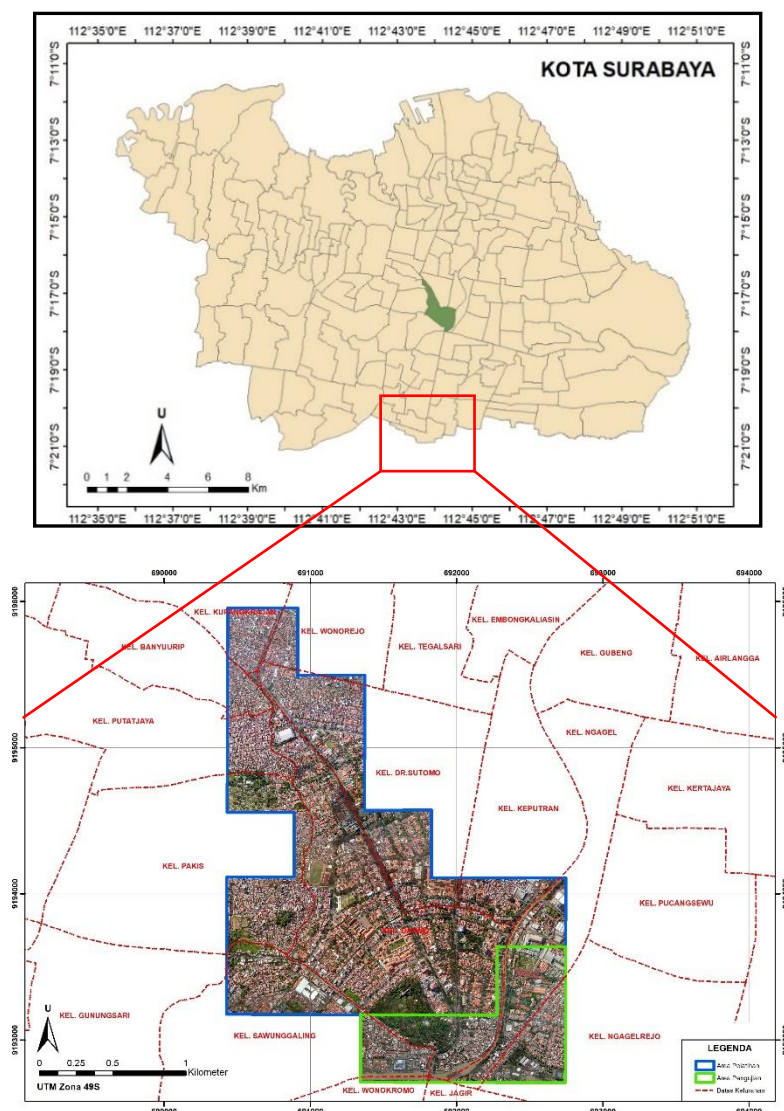
Ekstraksi objek sebagian besar didasarkan pada citra *orthophoto* karena resolusi spasial yang baik dalam beberapa tahun terakhir, kemajuan besar telah dicapai dalam *machine learning*, terutama dalam *deep learning* (Pasquali dkk., 2019). Pendekatan *deep learning* bisa diterapkan pada data *orthophoto* untuk mendeteksi jaringan jalan secara otomatis untuk menghemat waktu dan biaya pemrosesan data (Li, Liu, dkk., 2021). *Orthophoto* merupakan foto yang telah dikoreksi untuk distorsi akibat kemiringan kamera selama survei foto udara, distorsi dari lensa kamera, dan distorsi relief (Rossi, 2004). Pengembangan metode ekstraksi unsur rupabumi diperlukan data-data yang mendukung seperti data *training* jaringan jalan untuk melatih model *deep learning*. Data tersebut dapat diperoleh dari *ground truth* dan foto udara yang telah terkoreksi secara geometris. Interpretasi secara interaktif yang dilakukan oleh operator manusia secara visual masih merupakan pendekatan utama untuk klasifikasi digital informasi dari *orthophoto*. Akurasi interpretasi yang dihasilkan tergantung pada keterampilan dan pengalaman dari operator. Sehingga, dapat terjadi inkonsistensi pada data yang dihasilkan oleh operator yang berbeda. Segmentasi dan pemetaan unsur rupabumi 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 unsur rupabumi dari citra penginderaan jauh resolusi tinggi atau citra *orthophoto* akan memiliki banyak manfaat (Xu dkk., 2018).

Ekstraksi jaringan jalan sebagian besar didasarkan pada *orthophoto* atau CSRT karena resolusi spasial yang baik. Dalam beberapa tahun terakhir, pengembangan *deep learning* telah mendorong kemajuan ekstraksi jalan (Mnih dan Hinton, 2010; Cheng dkk., 2017; Yang dkk., 2019; Yuan dkk., 2020). Ekstraksi jalan dianggap sebagai alat penting untuk perencanaan kota dan pengambilan keputusan (Huang dkk., 2018), layanan mengemudi otomatis (Claussmann dkk., 2019), bantuan kemanusiaan (Bonafilia dkk., 2019), dan pembaharuan data informasi geospasial (Levin dan Duke, 2012). Dalam beberapa tahun terakhir, pendekatan yang didasarkan pada *deep learning*, seperti *Convolutional Neural Networks* (CNN) dan variannya, menjadi banyak digunakan untuk pendeteksian objek. Beberapa model yang umum digunakan telah mencapai kinerja mutakhir tidak hanya dalam *computer vision*, seperti U-Net (Ronneberger dkk., 2015), DeepLab (Chen dkk., 2017), Pyramid Scene Parsing Network (PSPNet) (Zhao dkk., 2017), yang mengintegrasikan fitur di beberapa lapisan CNN untuk mengeksplorasi informasi multi-skala pada tingkat semantik yang berbeda. Mencoba untuk memperkaya jaringan *encoder-decoder*, beberapa penelitian menggunakan strategi penyempurnaan untuk memasukkan informasi kontekstual multi-skala. *Mask Region-based Convolutional Neural Network*

(Mask R-CNN) yang merupakan model segmentasi semantik juga bisa dimanfaatkan untuk pekerjaan ini (He dkk., 2018).

Dalam penelitian ini dilakukan ekstraksi jaringan jalan di area urban memanfaatkan data *orthophoto* dan DSM LiDAR Kota Surabaya yang diakuisisi pada tahun 2016. Metode yang digunakan adalah penerapan *deep learning* untuk mendeteksi jaringan jalan secara otomatis. Selanjutnya, dilakukan *post-processing* untuk jalan, termasuk regularisasi poligon menggunakan algoritma Douglas-Peucker (1973) dan penghalusan poligon untuk memastikan bahwa poligon yang dihasilkan tidak hanya sesuai dengan *ground truth* tetapi juga memiliki tampilan geometri yang realistis. Dilakukan evaluasi dengan menghitung presisi, *recall*, dan F_1 -score hasil ekstraksi jaringan jalan untuk menilai performa dari model yang diterapkan pada area urban. Dari penelitian ini diharapkan bisa menjadi solusi untuk percepatan pemetaan skala besar sebagai penunjang pembangunan kota cerdas dan rekomendasi pengambilan kebijakan terkait dengan perencanaan Kota Surabaya.

Data dan Metode



Gambar 1. Area Penelitian Kelurahan Darmo, Surabaya

Data *orthophoto* Kota Surabaya yang digunakan pada penelitian ini diakuisisi pada tahun 2016 dalam format *.tif. Data ini didapatkan dari Dinas Perumahan Rakyat dan Kawasan Permukiman, Cipta Karya dan Tata

Ruang, Pemerintah Kota Surabaya. Resolusi spasial *orthophoto* yang digunakan adalah 8 cm dengan ukuran masing-masing lembar peta adalah 480 m × 480 m. Lokasi penelitian dilakukan di Kelurahan Darmo, Kecamatan Wonokromo, Kota Surabaya (7°16'28"-7°18'0" LS dan 112°43'34"-112°44'34" BT).

Lokasi Kelurahan Darmo dipilih karena memiliki karakteristik pola jaringan jalan yang sangat bervariasi dan heterogen. Sedangkan, keadaan topografi di Kelurahan Darmo memiliki ketinggian permukaan tanah di sebagian besar wilayahnya relatif datar, berkisar 0,25 m hingga 24,28 m. Batas wilayah kelurahan darmo sebelah utara yaitu kelurahan Dr. Soetomo dan Kelurahan Keputran, sebelah timur Kelurahan Ngagel, sebelah selatan Kelurahan Sawunggaling, Kelurahan Wonokromo, Kelurahan Jagir, dan batas sebelah barat yaitu Kelurahan Banyu Urip, Kelurahan Putat Jaya, Kelurahan Pakis.

Deep Learning merupakan salah satu metode yang telah banyak digunakan dalam penelitian mengenai ekstraksi jaringan jalan. 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 (LeCun dkk., 2015). Algoritma *Deep Learning* seperti *Convolutional Neural Networks* (CNNs) telah dikenal luas sebagai pendekatan yang menonjol untuk banyak aplikasi *computer vision* (pengenalan, deteksi, dan klasifikasi gambar / video) dan telah menunjukkan hasil yang luar biasa di banyak aplikasi (Cheng dkk., 2018). *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 (Pan dan Yang, 2010). Cira dkk. (2020) melakukan ekstraksi jaringan jalan skala besar memanfaatkan data *orthophoto*. Pendekatan *deep learning* bisa diterapkan pada *orthophoto* untuk mendeteksi unsur rupabumi, seperti bangunan dan jalan. Ketidakteraturan bentuk, sulitnya memodelkan objek menjadi celah untuk meneliti dan mencari pendekatan strategis untuk otomatisasi pendeteksian jaringan jalan memanfaatkan data *orthophoto*, dengan metode pemrosesan *deep learning*.

Faster R-CNN (Ren dkk., 2016) dan Mask R-CNN (He dkk., 2017) umumnya dianggap sebagai arsitektur mutakhir untuk ekstraksi objek dan segmentasi yang instan. Kedua arsitektur ini merupakan bagian dari keluarga CNN berbasis region. Faster R-CNN secara efektif terdiri dari dua jaringan; jaringan *Region Proposal Network* (RPN) dan jaringan pendeteksi Fast R-CNN. RPN dapat menggantikan metode pencarian selektif di versi sebelumnya (R-CNN dan Fast R-CNN) yang pemrosesannya lebih memakan waktu. Dalam Faster R-CNN, RPN dapat melakukan ekstraksi dengan menggunakan peta fitur yang diturunkan dari lapisan konvolusi terakhir di *backbone* CNN (yaitu VGG16/ResNet) sebagai proposal untuk jaringan detektor. Hal tersebut, memungkinkan arsitektur CNN digunakan untuk proposal dan klasifikasi region. Jaringan tersebut kemudian melakukan *pooling Region of Interest* (RoI), lapisan yang terhubung penuh dan klasifikasi seperti di R-CNN (Griffiths dan Boehm, 2019). 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 L_l digunakan untuk pelatihan dan didefinisikan sebagai:

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

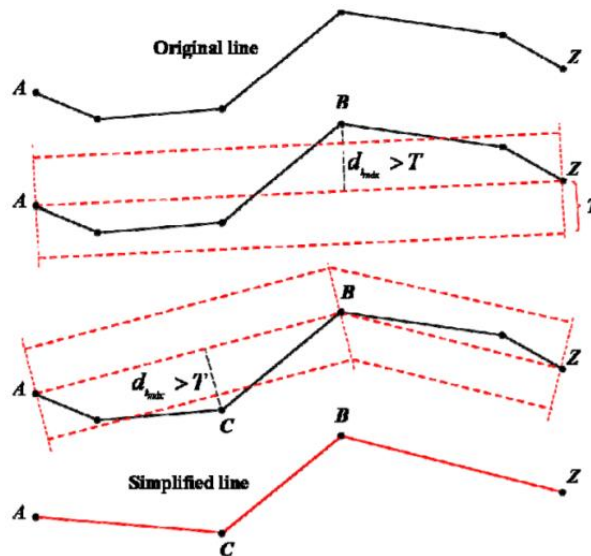
di mana i adalah indeks jangkar dalam *batch* kecil, p_i adalah kemungkinan proposal menjadi objek, t_i adalah koordinat kotak pembatas yang diprediksi, L_{cls} adalah kehilangan log, p_i^* adalah label objek kebenaran dari *ground truth*, L_{reg} adalah *loss* halus L_l , dan t_i^* adalah koordinat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

$$\text{Smooth}L_1(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases} \quad (2)$$

Mask Region-based Convolutional Neural Network (Mask R-CNN) diperkenalkan sebagai perpanjangan dari *Faster R-CNN* untuk melakukan segmentasi berbasis piksel yang akurat (He dkk., 2017). 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 *masker* 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; $L = L_{cls} + L_{box} + L_{mask}$, dimana L adalah total *loss*.

Algoritma Douglas-Peucker yang diimplementasikan dalam aplikasi ini merupakan metode penyederhanaan baris kedua yang disajikan oleh Douglas dan Peucker (1973). Dasar dari algoritma terdiri dari pemilihan titik-titik spesifik garis asli yang disebut titik kritis atau jangkar yang akan membangun garis umum. Untuk memilihnya, faktor toleransi de toleransi $T > 0$ ditetapkan sebelumnya. Ini disebut toleransi sederhana dan dinyatakan dalam satuan panjang. Jika garis tertutup, titik pertama dan terakhir tidak menentukan garis karena A dan Z bertepatan karena memiliki koordinat yang sama, jarak maksimum ke segmen awal diganti dengan jarak terbesar ke titik-titik ini. Namun demikian, dalam hal ini adalah kontroversial untuk memilih titik kritis sebagai titik kurva hanya karena telah didigitalkan di tempat pertama atau terakhir.



Gambar 2. Penerapan Algoritma Douglas-Peucker (Crespo dkk., 2014)

Toleransi T menentukan perpindahan terbesar di antara titik yang ditentukan. Peningkatan nilai T akan menyebabkan penyesuaian terhadap garis asli dan kesalahan topologi meningkat, namun peningkatan nilai T akan menurunkan titik-titik garis yang dihasilkan, menurunkan waktu pemrosesan, dan menurunkan penghalusan. Oleh karena itu, perlu dilakukan evaluasi terhadap hasil yang diperoleh untuk mencapai toleransi terbaik dalam setiap proses yang akan memberikan hasil yang seimbang dalam aspek-aspek tersebut. Ekstraksi jaringan jalan dievaluasi dengan perhitungan matriks kesalahan. Menurut Lunetta dan Lyon (2005) sebuah matrik kesalahan adalah sebuah himpunan persegi dari nilai-nilai yang disusun dalam baris dan kolom yang menyatakan jumlah dari sampel misalnya piksel, kelompok piksel atau poligon yang diklasifikasi untuk kategori tertentu relatif terhadap kategori sebenarnya. Secara khusus kolom merepresentasikan data sedangkan baris mengindikasikan peta yang dihasilkan dari data citra. Data referensi dianggap benar yang dapat dikumpulkan dari berbagai sumber, termasuk interpretasi citra setelit, *orthophoto*, hasil pengamatan dan pengukuran di lapangan. Matrik kesalahan setelah dibuat dengan benar, maka dapat digunakan sebagai tahap awal untuk sebuah rangkaian deskriptif dan teknik analisis statistik. Evaluasi kinerja model memerlukan derivasi matriks kesalahan yang mengidentifikasi kasus (a) *true positive*, (b) *false positive*, (c) *false negative*, and (d) *true negative* yang diprediksi oleh masing-masing model.

Tabel 1. Matriks Kesalahan

Model	Validasi	
	Positive (+)	Negative (-)
Positive (+)	True Positive	False Negative
Negative (-)	False Positive	True Negative

Statistik deskriptif paling umum dan paling sederhana adalah *overall accuracy*, yang menjumlahkan hasil dengan membagi total benar dari sampel, yaitu jumlah dari diagonal utama dengan jumlah unit sampel dalam matrik kesalahan. Persamaan *overall accuracy* adalah sebagai berikut:

$$\text{Overall accuracy} = \frac{\Sigma \text{ sampel benar (diagonal)}}{\Sigma \text{ sampel}} \times 100 \quad (3)$$

Selanjutnya, untuk menghitung nilai *kappa* untuk pengujian reliabilitas antar-penilai atau intra-penilai. Mirip dengan koefisien korelasi, dapat berkisar dari -1 hingga +1, di mana 0 mewakili jumlah kesepakatan yang dapat diharapkan dari peluang acak, dan 1 mewakili kesepakatan sempurna antara penilai. Nilai *kappa* dihitung dengan persamaan sebagai berikut:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (4)$$

dimana p_o adalah kesepakatan relatif yang diamati di antara penilai, dan p_e adalah probabilitas hipotetis dari kesepakatan kemungkinan, menggunakan data yang diamati untuk menghitung probabilitas setiap pengamat secara acak melihat setiap kategori. Nilai *kappa* dapat diinterpretasikan sebagaimana dijelaskan pada Tabel 2 (McHugh, 2012).

Tabel 2. Interpretasi Nilai *Kappa*

Nilai <i>Kappa</i>	Tingkat Kesepakatan
0 – 0,20	Tidak ada (<i>none</i>)
0,21 – 0,40	Minimal
0,41 – 0,60	Lemah (<i>weak</i>)
0,61 – 0,80	Sedang (<i>moderate</i>)
0,81 – 0,90	Kuat
Di atas 0,90	Sangat kuat

Confusion matrix dapat digunakan untuk mengevaluasi kinerja model dalam melakukan ekstraksi jaringan jalan. Label dibagi menjadi sampel positif dan sampel negatif, sedangkan hasil prediksi dibagi menjadi hasil positif (benar) dan hasil negatif (salah). Penggunaan TP (*true positive*) dan TN (*true negative*) untuk mewakili prediksi yang benar, sementara FP (*false positive*) dan FN (*false negative*) digunakan untuk mewakili prediksi yang salah. Berdasarkan indikator tersebut hasil ekstraksi ekstraksi jalan akan dievaluasi berdasarkan luasnya (Xin dkk., 2019; Shao dkk., 2021). Presisi menunjukkan berapa banyak sampel positif yang diprediksi positif. Ini mewakili proporsi objek berlabel yang diambil dari model:

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (5)$$

Recall menunjukkan jumlah contoh positif yang diprediksi benar dalam sampel. Ini mewakili proporsi objek yang diberi label dengan benar oleh model:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

F1-score, rata-rata harmonik presisi dan *recall*, adalah indeks yang banyak untuk mengukur akurasi model klasifikasi biner:

$$F_1 \text{ score} = 2 \times \frac{\text{Presisi} \times \text{recall}}{\text{Presisi} + \text{recall}} \quad (7)$$

Intersect over Union (IoU) adalah metrik evaluasi umum untuk segmentasi semantik dan deteksi objek, yang mengukur akurasi objek dasar yang terdeteksi dari dataset dan mengukur tingkat kesesuaian antara hasil ekstraksi dan label sebenarnya. Semakin besar nilai IOU, semakin banyak area hasil yang tumpang tindih.

$$IoU = \frac{TP}{TP + FP + FN} \quad (8)$$

Hasil dan Pembahasan

Data *orthophoto* yang digunakan pada penelitian ini berformat *Georeferenced Tagged Image File Format* (GeoTIFF) dengan ekstensi *.tif. Data ini berbekal kanal *red*, *green*, dan *blue* (RGB) yang sudah diolah hingga memiliki geometrik yang sesuai. Memiliki resolusi spasial 0,08 m, data ini tergolong resolusi sangat tinggi. Data *orthophoto* yang merupakan gabungan dari 20 lembar peta selanjutnya dibagi 80% untuk pelatihan (10% dari pelatihan untuk validasi) dan 20% untuk pengujian. Pada Gambar 1 ditunjukkan data *orthophoto* yang digunakan untuk pelatihan dan pengujian, serta ditampilkan pada batas administrasi Kelurahan Darmo. Referensi batas kelurahan yang digunakan diperoleh dari Peta Rupabumi Indonesia (RBI). Wilayah penelitian ini berada di dalam Kecamatan Wonokromo yang berbatasan dengan Kecamatan Sawahan dan Kecamatan Tegalsari.

Data label pelatihan jalan yang digunakan pada penelitian ini berformat *shapefile* (*.shp). Data ini berlokasi di Kelurahan Darmo, Kecamatan Wonokromo, Kota Surabaya yang menggabungkan 16 lembar peta (80% dari dataset). Jaringan jalan digabungkan menjadi satu *record* untuk meringankan proses pelatihan model. Poligon jalan diusahakan seakurat mungkin jika ditampilkan dengan *orthophoto*. Semakin banyak dan beragam *dataset* yang digunakan untuk melatih model maka akan semakin baik karena model bisa belajar dari banyaknya variasi objek yang digunakan dalam proses pelatihan. Hal utama yang penting dalam mempersiapkan data label pelatihan jalan adalah poligon harus mengikuti struktur geometri jalan yang kompleks dan harus meliputi bagian jalan yang tertutup oleh pepohonan atau objek lain.

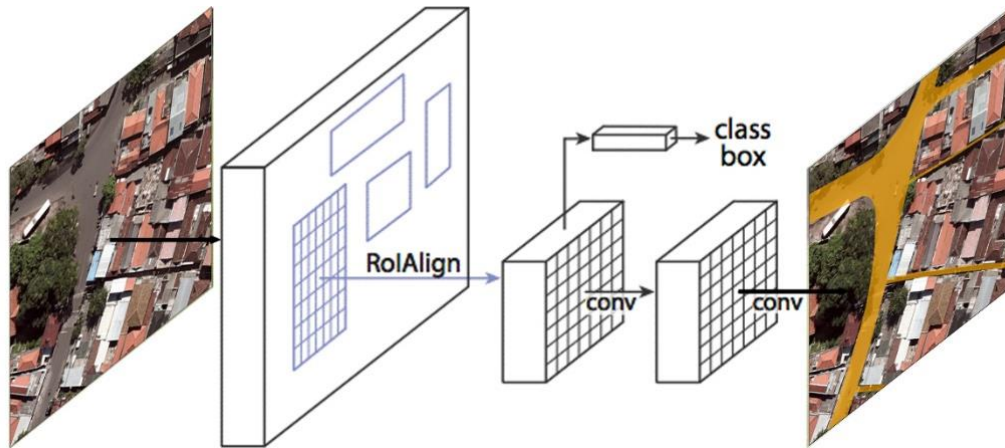
Data *orthophoto* dan set data pelatihan jalan tersebut selanjutnya dilakukan proses pemecahan (*split*) menjadi bagian-bagian kecil (*chip*) dan konversi ke format *deep learning*. Dalam melakukan proses *split* ini ukuran dari *orthophoto* tersebut dijadikan 512 piksel x 512 piksel. Sedangkan, untuk *stride* berukuran 256 piksel x 256 piksel diterapkan dalam proses *split* ini yang menghasilkan *overlap* antar *chip* sebesar 50%. Hal ini berarti setiap jarak 256 piksel secara vertikal dan horizontal akan menghasilkan satu *chip orthophoto* berukuran 512 piksel x 512 piksel. Dilakukan proses augmentasi berupa rotasi dengan sudut 180°. Sehingga, *chip* berukuran 512 piksel x 512 piksel yang dihasilkan dari proses konversi ini berjumlah 12.338 citra dan 15.952 fitur data (lihat Tabel 3). Set data ini yang akan digunakan untuk proses pelatihan dalam jaringan saraf tiruan untuk mendapatkan model yang dapat mendeteksi jaringan jalan di lokasi penelitian.

Tabel 3. Hasil *Split* Data Pelatihan Jalan

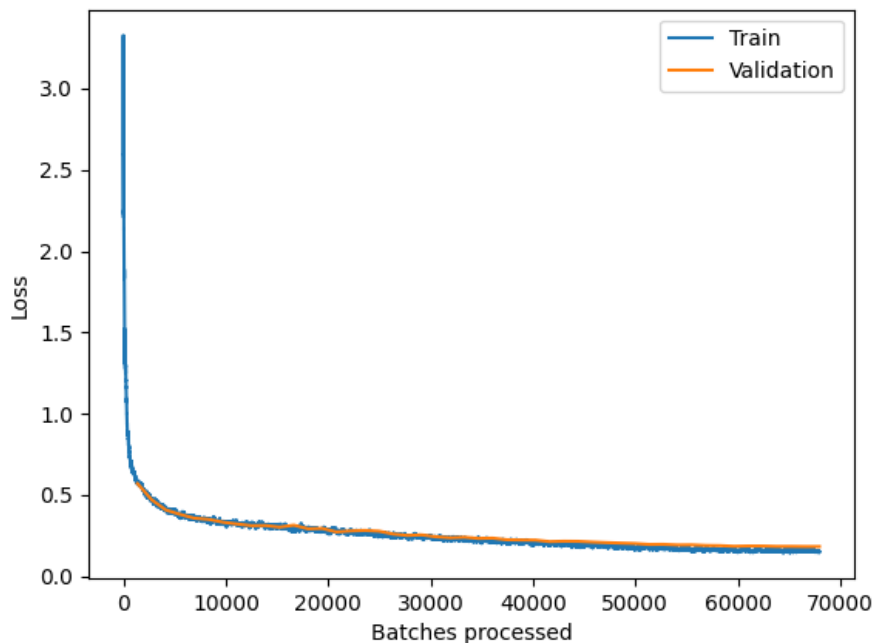
Nilai kelas	Nama kelas	Jumlah citra	Jumlah fitur	Ukuran rata-rata
1	Jalan	12.338	15.952	241,50

Model ekstraksi jalan pada penelitian ini dilatih menggunakan *dataset* pelatihan poligon jaringan jalan dari 16 lembar peta di Kelurahan Darmo. *Dataset* tersebut sebelumnya sudah dilakukan dilakukan proses pemecahan (*split*) menjadi bagian-bagian kecil (*chip*) dan konversi ke format *deep learning*. Pada penelitian ini, pelatihan model deteksi bangunan dilakukan menggunakan *backbone* ResNet-50 dan *framework* PyTorch pada Windows 10. Model dilatih dengan Adam *optimizer* sebanyak 50 *epoch*, dengan ukuran *batch* 8, dan *learning rate* 0,0001. Banyaknya *epoch* pelatihan ini didasarkan pada *trial and error* untuk mendapatkan hasil pelatihan yang paling optimal. Waktu yang dibutuhkan untuk melatih model ini pada GPU NVIDIA GeForce GTX 2080 8GB adalah selama 40 jam 13 menit 22 detik. Proses pelatihan model Mask R-CNN untuk deteksi objek akan menghasilkan data definisi model berekstensi *Esri Model Definition* (*.emd). Dalam data definisi model

tersebut ada informasi terkait dengan deteksi objek, seperti ukuran file citra pada saat melakukan pelatihan, tipe dari model, parameter model, *learning rate*, lokasi file dari model, jumlah *band* dari citra yang digunakan, jumlah kelas yang akan dideteksi dan rata-rata presisi yang dihasilkan dari model. Arsitektur pelatihan model Mask R-CNN untuk ekstraksi jalan memanfaatkan *orthophoto* bisa dilihat pada Gambar 3. Grafik *loss* yang menggambarkan proses pelatihan model Mask R-CNN untuk ekstraksi jaringan jalan dapat dilihat pada Gambar 4.

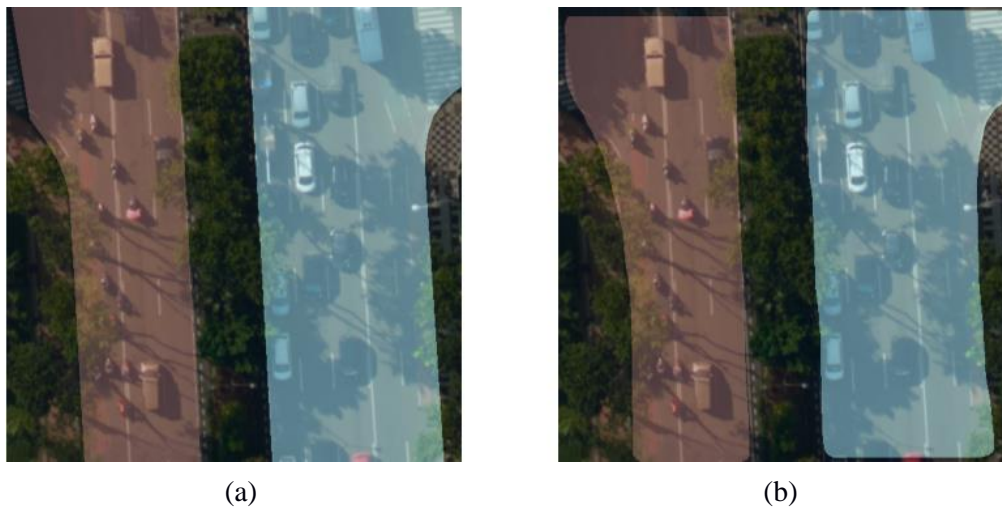


Gambar 3. Arsitektur Pelatihan Model Mask R-CNN Ekstraksi Jalan Memanfaatkan *Orthophoto*



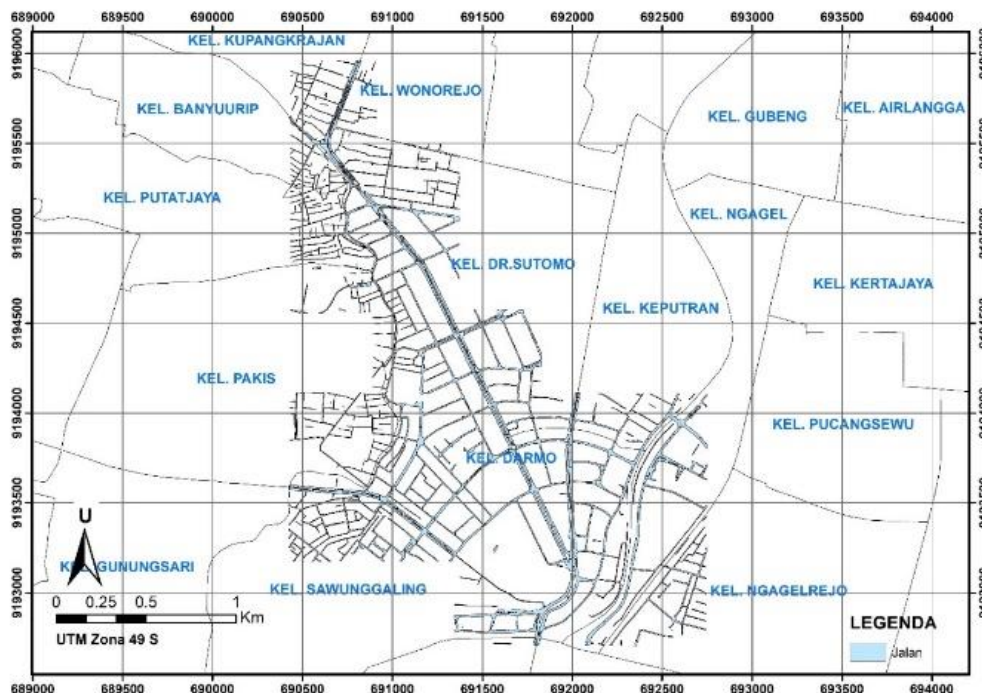
Gambar 4. Grafik *Training Loss* vs *Validation Loss* Model Mask R-CNN Ekstraksi Jaringan Jalan

Dalam melakukan proses pelatihan model, tidak semua data dilibatkan tetapi akan ada sebagian data yang digunakan untuk pengujian model guna mengevaluasi akurasi model dalam mendeteksi jaringan jalan. Gambar 5. memperlihatkan hasil pelatihan model, dimana akan terlihat perbedaan antara *ground truth* dengan prediksi. Model yang sudah dilatih akan digunakan untuk mendeteksi jaringan jalan untuk mendapat informasi unsur rupabumi di area urban. Deteksi objek berbasis wilayah (*region*) ini menghasilkan segmen-segmen vektor namun geometrinya belum sempurna, khususnya pada tikungan dan persimpangan jalan sehingga perlu dilakukan *post-processing*.



Gambar 5. Perbandingan (a) *Ground Truth* dan (b) Prediksi Model Ekstraksi Jaringan Jalan

Model Mask R-CNN untuk ekstraksi jalan dari *orthophoto* yang telah terlatih berekstensi *Esri Model Definition (*.emd)*, selanjutnya digunakan untuk mengekstrak jaringan jalan pada lokasi penelitian. Model ekstraksi jalan dari *orthophoto* akan mengekstrak jalan berdasarkan spektral tiga kanal, yaitu *Red*, *Green*, dan *Blue*. Secara spektral warna dari badan jalan tidak terlalu bervariasi. Kanopi yang menutupi badan jalan, seperti tutupan bangunan, vegetasi tinggi, dan bayangan menjadi tantangan bagi model untuk bisa melakukan deliniasi batas antara badan jalan dengan objek non-jalan secara tepat. Selain digunakan pada data pengujian, model juga digunakan untuk mengekstrak jaringan jalan pada area pelatihan namun yang akan digunakan untuk evaluasi akurasi adalah hasil pada area pengujian. Terlihat bahwa model sudah cukup baik dalam mendeteksi objek namun deliniasi batas jalan. Namun, hasil dari ekstraksi jaringan jalan masih berupa segmen-segmen yang masih overlap dan geometrinya masih perlu diperbaiki, terutama pada tikungan dan persimpangan jalan. Ekstraksi jalan ini menghasilkan jalan dengan luas total 498.379,687 m². Meskipun Mask R-CNN dapat melakukan segmentasi tingkat objek, hasil ekstraksi jaringan jalan perlu dilakukan *post-processing* untuk memperbaiki geometri jalan.

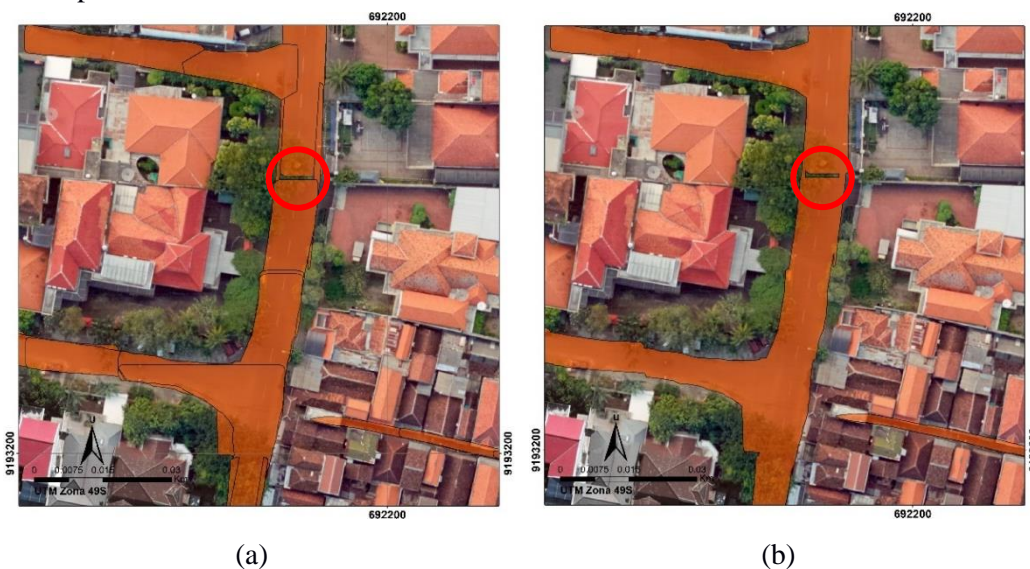


Gambar 6. Hasil Ekstraksi Jaringan Jalan Sebelum *Post-Processing*



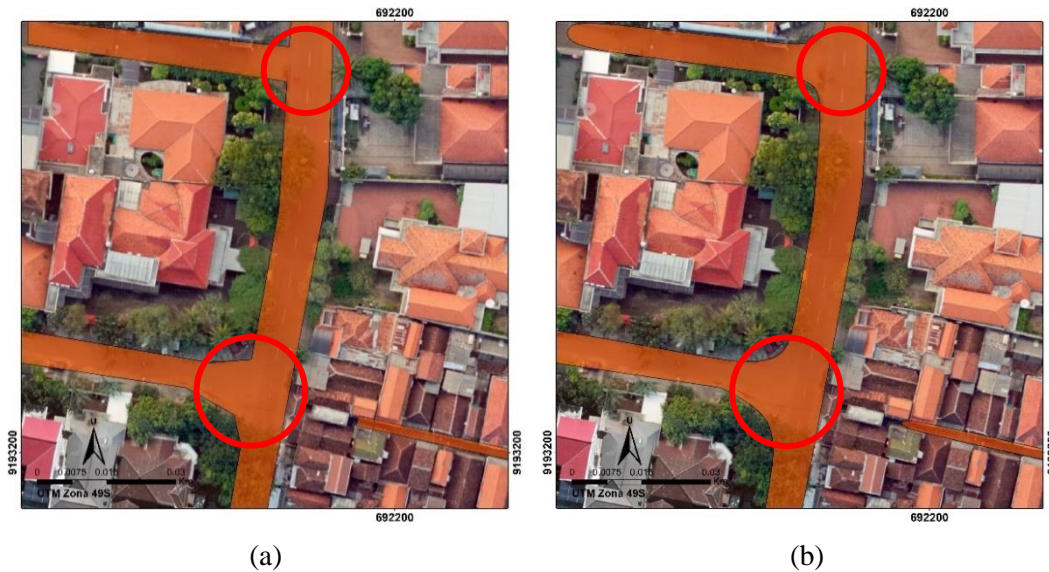
Gambar 7. Hasil Ekstraksi Jalan Masih Berupa Segmen-Segmen dan Masih *Overlap* Antar Segmen

Ekstraksi jaringan jalan menggunakan model Mask R-CNN yang telah dilatih menghasilkan poligon jalan yang sudah cukup baik namun masih berupa segmen-segmen dan secara geometri masih perlu dilakukan perbaikan. Diperlukan *post-processing* untuk menggabungkan segmen-segmen jalan menjadi satu kesatuan jaringan jalan serta memperbaiki bentuk geometri dari jaringan jalan yang sudah terdeteksi. Hasil ekstraksi jalan dilakukan penghilangan *noise* dilanjutkan dengan proses *dissolve* poligon untuk menyatukan segmen-segmen jalan yang masih saling tumpang tindih menjadi satu *record* jaringan jalan. Hasil dari penggabungan segmen jaringan jalan bisa dilihat pada Gambar 8.



Gambar 8. Hasil Ekstraksi Jaringan Jalan (a) Sebelum dan (b) Setelah *Dissolve*

Poligon ekstraksi jalan yang telah dilakukan penggabungan segmen-segmen ini masih menyisakan jaringan jalan yang terpotong ditunjukkan lingkaran berwarna merah pada Gambar 8 (b). Selain itu, jalan setelah dilakukan *dissolve* masih menyisakan bentuk yang kasar, khususnya pada jalan yang lurus sehingga perlu dilakukan penyederhanaan poligon jalan untuk membuat bentuk geometri jalan lebih teratur.



Gambar 9. Hasil Ekstraksi Jaringan Jalan (a) Setelah regularisasi dan (b) Setelah penghalusan

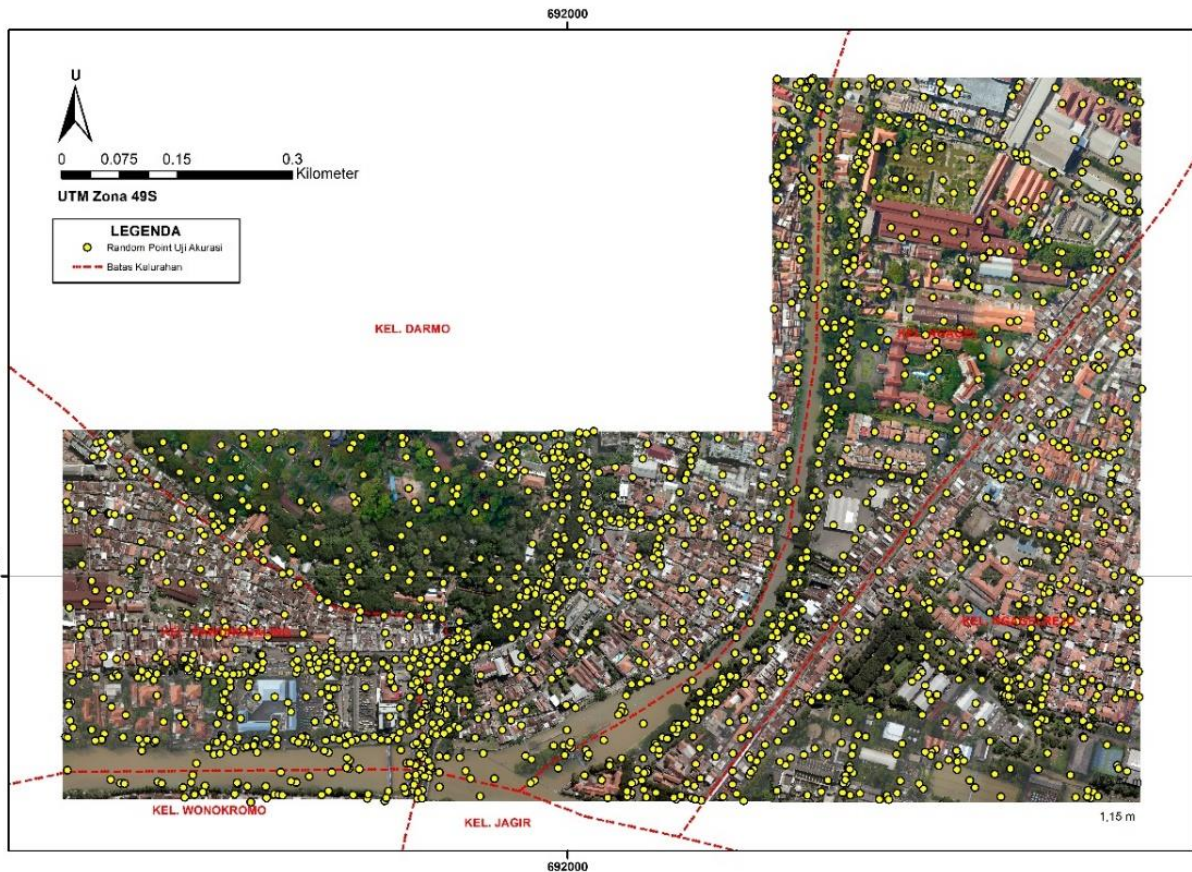
Yang utama dalam proses perbaikan geometri ini adalah algoritma Douglas-Peucker. Mulai dari poligon awal hasil ekstraksi, algoritma penyederhanaan pada penelitian ini terdiri dari dua langkah: penyesuaian kasar yang menghilangkan kesalahan nyata dari ekstraksi dan penyempurnaan yang menyesuaikan arah garis dan posisi sudut. Penerapan toleransi pada algoritma Douglas-Peucker menghasilkan beberapa konsekuensi. Penyempurnaan geometri jaringan jalan dilakukan dengan proses penghalusan poligon. Proses ini dilakukan untuk memperbaiki bentuk geometri yang kurang sempurna dari hasil *post-processing* sebelumnya yang masih menyisakan bentuk kasar, khususnya di tikungan dan persimpangan jalan yang ditunjukkan lingkaran berwarna merah pada Gambar 9. Proses penghalusan poligon ini merupakan akhir dari semua proses ekstraksi jaringan jalan. Proses penyederhaan dan penghalusan pada *post-processing* poligon jaringan jalan dilakukan sebanyak 3 iterasi dengan parameter nilai toleransi paling optimal sebesar 0,25 m yang didapatkan berdasarkan *trial and error*.

Presisi (*precision*), kelengkapan (*recall*), skor-F1, dan IoU dihitung berdasarkan Persamaan 5 – 8, 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 deteksi, dan *false negative* (FN) menunjukkan jumlah poligon bangunan yang tidak terdeteksi berdasarkan data pengujian. Hasil ekstraksi jalan akan dievaluasi berdasarkan luas interseksinya. Evaluasi hasil ekstraksi jaringan jalan berdasarkan luasnya pada area pengujian dapat dilihat pada Tabel 3.

Tabel 3. Uji Akurasi Deteksi Bangunan Model Mask R-CNN

TP	FP	FN	
81383,241 m ²	8760,380 m ²	13409,842 m ²	
<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	IoU
0,9028	0,8585	0,8801	0,7859

Berdasarkan evaluasi model secara kuantitatif dapat diketahui bahwa metode yang peneliti terapkan menghasilkan kinerja yang cukup baik untuk ekstraksi jaringan jalan dengan presisi 90,28%; kelengkapan (*recall*) 85,85%; skor-F1 88,01%; dan IoU 78,59%. Selanjutnya, dilakukan uji akurasi dengan total 2000 *random point* (1000 kelas jalan dan 1000 kelas non-jalan) di area pengujian. Sebaran *random point* tersebut bisa dilihat di Gambar 10.



Gambar 10. Sebaran *Random Point* untuk Uji Akurasi Hasil Ekstraksi Jalan Dengan Model Mask R-CNN

Nilai *producer accuracy* berfungsi sebagai penilaian secara tematik, yaitu menunjukkan tingkat kebenaran hasil ekstraksi terhadap *dataset* pengujian. *User accuracy* menjelaskan mengenai ketelitian hasil ekstraksi terhadap seluruh obyek yang dapat diidentifikasi. *Overall accuracy* menggambarkan nilai akurasi total kenampakan objek yang benar hasil ekstraksi dengan *dataset* pengujian. Nilai *kappa* mempertimbangkan faktor kesalahan proses klasifikasi, sehingga nilai *kappa* lebih rendah dari nilai *overall accuracy* dimana hanya mempertimbangkan data yang benar antara hasil ekstraksi dengan *dataset* pengujian. Uji akurasi hasil ekstraksi jaringan jalan menggunakan model Mask R-CNN dapat dilihat pada Tabel 4.

Berdasarkan evaluasi model secara kuantitatif tersebut terlihat bahwa model cukup baik untuk melakukan ekstraksi jalan. Nilai *producer accuracy* dan *user accuracy* semuanya di atas 90%. Dengan *Overall accuracy* 95,25 % dan nilai *kappa* 90,5%. Pada penelitian ini nilai *kappa* 90,5% diinterpretasikan memiliki kesepakatan sangat kuat berdasarkan Tabel 2. Ketidaksesuaian dan kesalahan ekstraksi umumnya bersumber dari kanopi yang menutupi badan jalan, seperti tutupan bangunan, vegetasi tinggi, dan bayangan sehingga model akan kesulitan melakukan delineasi batas antara badan jalan dengan objek non-jalan. Selain itu, model juga kesulitan dalam melakukan ekstraksi jalan yang kecil seperti di jalan lokal permukiman padat karena jalan yang sempit dan umumnya tertutup oleh bayangan dari bangunan. Gambar 11 menunjukkan contoh perbandingan hasil ekstraksi jaringan jalan dengan *dataset* pengujian.

Tabel 4. Uji Akurasi Deteksi Jalan Model Mask R-CNN

Kelas	Jalan	Non-jalan	Total	UA
Jalan	983	17	1000	0,983
Non-jalan	78	922	1000	0,922
Total	1061	939	2000	
PA	0,926	0,982		



Gambar 11. Hasil Ekstraksi Jalan (*Polyline Kuning*) dengan *Dataset Validasi (Polyline Merah)*

Kesimpulan

Dalam proses pelatihan model Mask R-CNN dengan backbone ResNet-50 *backbone* ResNet-50 dan *framework* PyTorch. Model dilatih dengan Adam *optimizer* sebanyak 50 *epoch*, dengan ukuran *batch* 4, dan *learning rate* 0,0001 menggunakan GPU NVIDIA GeForce GTX 2080 8GB. Proses pelatihan model Mask R-CNN ekstraksi jaringan jalan menggunakan *orthophoto* memakan waktu selama 40 jam 13 menit 22 detik. Dari proses pelatihan tersebut menghasilkan model Mask R-CNN ekstraksi jaringan jalan yang sudah cukup baik untuk bisa melakukan delineasi batas antara objek jalan dan non-jalan. Ekstraksi jalan menggunakan model Mask R-CNN yang sudah dilatih menghasilkan jalan total seluas 498.379,687 m² di lokasi penelitian.

Hasil ekstraksi jaringan jalan pada penelitian ini menunjukkan bahwa model Mask R-CNN menghasilkan kinerja yang baik. Penerapan *post-processing* hasil ekstraksi jaringan jalan dilakukan *dissolve* poligon untuk menyatukan segmen-segmen jalan yang masih saling tumpang tindih menjadi satu *record* jaringan jalan, regularisasi *olygon* menggunakan algoritma Douglas-Peucker untuk membuat jalan lebih teratur, dan penghalusan *olygon* untuk memperbaiki geometri jalan khususnya pada tikungan dan persimpangan jalan. Berdasarkan uji akurasi yang dilakukan, ekstraksi jalan menghasilkan nilai nilai presisi 90,28%; kelengkapan (*recall*) 85,85%; skor-F1 88,01%; dan IoU 78,59%; serta *Overall accuracy* 95,25 % dan nilai *kappa* 90,5%.

Ucapan Terimakasih

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

- Bonafilia, D., Gill, J., Basu, S., & Yang, D. (2019). *Building High Resolution Maps for Humanitarian Aid and Development with Weakly- and Semi-Supervised Learning*. 1–9.
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2016). Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *ArXiv:1412.7062 [Cs]*. <http://arxiv.org/abs/1412.7062>
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *ArXiv:1606.00915 [Cs]*. <http://arxiv.org/abs/1606.00915>
- Cheng, G., Yang, C., Yao, X., Guo, L., & Han, J. (2018). When Deep Learning Meets Metric Learning: Remote Sensing Image Scene Classification via Learning Discriminative CNNs. *IEEE Transactions on Geoscience and Remote Sensing*, 56(5), 2811–2821. <https://doi.org/10.1109/TGRS.2017.2783902>
- Cira, C.-I., Alcarria, R., Manso-Callejo, M.-Á., & Serradilla, F. (2020). A Deep Learning-Based Solution for Large-Scale Extraction of the Secondary Road Network from High-Resolution Aerial Orthoimagery. *Applied Sciences*, 10(20), 7272. <https://doi.org/10.3390/app10207272>
- Claussmann, L., Revilloud, M., Gruyer, D., & Glaser, S. (2019). A Review of Motion Planning for Highway Autonomous Driving. *IEEE Transactions on Intelligent Transportation Systems*, 21(5), 1826–1848. <https://doi.org/10.1109/TITS.2019.2913998>
- Gao, X., Sun, X., Zhang, Y., Yan, M., Xu, G., Sun, H., Jiao, J., & Fu, K. (2018). An End-to-End Neural Network for Road Extraction From Remote Sensing Imagery by Multiple Feature Pyramid Network. *IEEE Access*, 6, 39401–39414. <https://doi.org/10.1109/ACCESS.2018.2856088>
- Griffiths, D., & Boehm, J. (2019). Improving public data for building segmentation from Convolutional Neural Networks (CNNs) for fused airborne lidar and image data using active contours. *ISPRS Journal of Photogrammetry and Remote Sensing*, 154, 70–83. <https://doi.org/10.1016/j.isprsjprs.2019.05.013>
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2018). Mask R-CNN. *ArXiv:1703.06870 [Cs]*. <http://arxiv.org/abs/1703.06870>
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *2017 IEEE International Conference on Computer Vision (ICCV)*, 2980–2988. <https://doi.org/10.1109/ICCV.2017.322>
- Huang, B., Zhao, B., & Song, Y. (2018). Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sensing of Environment*, 214, 73–86. <https://doi.org/10.1016/j.rse.2018.04.050>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Levin, N., & Duke, Y. (2012). High spatial resolution night-time light images for demographic and socio-economic studies. *Remote Sensing of Environment*, 119, 1–10. <https://doi.org/10.1016/j.rse.2011.12.005>
- Li, J., Liu, Y., Zhang, Y., & Zhang, Y. (2021). Cascaded Attention DenseUNet (CADUNet) for Road Extraction from Very-High-Resolution Images. *ISPRS International Journal of Geo-Information*, 10(5), 329. <https://doi.org/10.3390/ijgi10050329>
- Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- Pasquali, G., Iannelli, G. C., & Dell’Acqua, F. (2019). Building Footprint Extraction from Multispectral, Spaceborne Earth Observation Datasets Using a Structurally Optimized U-Net Convolutional Neural Network. *Remote Sensing*, 11(23), 2803. <https://doi.org/10.3390/rs11232803>
- Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *ArXiv:1506.01497 [Cs]*. <http://arxiv.org/abs/1506.01497>
- Republik Indonesia. (2020). Undang-Undang Republik Indonesia No 11 Tahun 2020 Tentang Cipta Kerja. Jakarta.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Vol. 9351, pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28
- Shao, Z., Zhou, Z., Huang, X., & Zhang, Y. (2021). MRENet: Simultaneous Extraction of Road Surface and Road Centerline in Complex Urban Scenes from Very High-Resolution Images. *Remote Sensing*, 13(2), 239. <https://doi.org/10.3390/rs13020239>
- Xin, J., Zhang, X., Zhang, Z., & Fang, W. (2019). Road Extraction of High-Resolution Remote Sensing Images Derived from DenseUNet. *Remote Sensing*, 11(21), 2499. <https://doi.org/10.3390/rs11212499>

- Xu, Y., Xie, Z., Feng, Y., & Chen, Z. (2018). Road Extraction from High-Resolution Remote Sensing Imagery Using Deep Learning. *Remote Sensing*, *10*(9), 1461. <https://doi.org/10.3390/rs10091461>
- Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid Scene Parsing Network. *ArXiv:1612.01105 [Cs]*. <http://arxiv.org/abs/1612.01105>
- Zhu, Q., Zhang, Y., Wang, L., Zhong, Y., Guan, Q., Lu, X., Zhang, L., & Li, D. (2021). A Global Context-aware and Batch-independent Network for road extraction from VHR satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *175*, 353–365. <https://doi.org/10.1016/j.isprsjprs.2021.03.016>



This article is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).