STORAGE – Jurnal Ilmiah Teknik dan Ilmu Komputer, Vol. 4 No. 4, November 2025, 337 – 347
e-ISSN: 2828-5344 | DOI: 10.55123

SEGMENTASI BANGUNAN PERKOTAAN PADA CITRA SATELIT BERESOLUSI TINGGI: CNN, U-NET (VGG16), DAN DEEPLABV3+ (RESNET-50)

[1] Putu Haryaka Setadewa, [2] Kadek Yota Ernanda Aryanto, [3] Luh Joni Erawati Dewi
[1,2]Ilmu Komputer, Teknik dan Kejuruan, Universitas Pendidikan Ganesha,
[3]Teknologi Rekayasa Perangkat Lunak, Teknik dan Kejuruan, Universitas Pendidikan Ganesha
[1] haryaka@student.undiksha.ac.id

INFO ARTIKEL

Riwayat Artikel :
Diterima : 5 Oktober 2025
Disetujui : 16 Oktober 2025

Kata Kunci :
Segmentasi Bangunan, CNN, U-Net, DeepLabV3+, Citra Satelit, Urbanisasi.

ABSTRAK

Seiring meningkatnya laju urbanisasi di Indonesia, kebutuhan pemetaan bangunan yang akurat menjadi semakin penting untuk mendukung perencanaan tata ruang, mitigasi bencana, dan pengelolaan infrastruktur perkotaan. Pendekatan konvensional berbasis survei manual dinilai kurang efisien, terutama di wilayah dengan pertumbuhan pesat. Oleh karena itu, pemanfaatan citra satelit dan *Deep learning* menjadi solusi potensial untuk identifikasi bangunan secara otomatis. Penelitian ini membandingkan performa tiga model segmentasi bangunan pada citra satelit resolusi tinggi: CNN konvensional (*CNN-K*), U-Net berbasis VGG16 (*U-VGG*), dan DeepLabV3+ dengan ResNet-50 (*DL-ResNet*). Dataset terdiri atas 1.216 *patch* citra dari kawasan Bali Selatan yang telah dilabeli dan diaugmentasi. Evaluasi dilakukan menggunakan metrik akurasi, *IoU*, *dice coefficient*, *precision*, *recall*, dan *F1-score*. Hasil menunjukkan *U-VGG* unggul (*dice* 89%, *IoU* 81%) dengan keseimbangan presisi dan efisiensi, sementara *DL-ResNet* mendekati hasilnya (*dice* 85%, *IoU* 80%) tetapi memerlukan sumber daya komputasi lebih besar. *CNN-K* mengalami *overfitting* dengan performa terendah.

ARTICLE INFO

Article History :
Received : Oct 5, 2025
Accepted : Oct 16, 2025

Keywords*:*
Building Segmentation, CNN, U-Net, DeepLabV3+, Satellite Imagery, Urbanization.

ABSTRACT

As urbanization accelerates across Indonesia, the demand for accurate building mapping has become increasingly vital for spatial planning, disaster mitigation, and urban infrastructure management. Conventional mapping methods based on manual surveys are often inefficient, especially in rapidly growing regions. Thus, satellite imagery integrated with Deep learning offers a promising solution for automated building detection. This study compares three building segmentation models on high-resolution satellite imagery: conventional CNN (CNN-K), VGG16-based U-Net (U-VGG), and DeepLabV3+ with ResNet-50 (DL-ResNet). The dataset comprises 1,216 labeled and augmented image patches from South Bali. Performance evaluation was conducted using accuracy, IoU, dice coefficient, precision, recall, and F1-score. The results show that U-VGG achieved the best performance (dice 89%, IoU 81%), balancing precision and computational efficiency, while DL-ResNet yielded comparable results (dice 85%, IoU 80%) but required more resources. CNN-K suffered from overfitting, recording the lowest accuracy among the three.

1. PENDAHULUAN

Urbanisasi global yang terus meningkat telah mendorong kebutuhan akan pemetaan bangunan yang akurat untuk mendukung perencanaan tata ruang, mitigasi bencana, dan pemantauan lingkungan. Teknologi citra satelit resolusi tinggi menjadi solusi efektif untuk pemantauan wilayah perkotaan secara efisien dan real-time (Tang et al., 2025). Namun, seiring dengan semakin besarnya volume data yang dihasilkan, diperlukan teknologi pemrosesan yang mampu menangani data dalam jumlah sangat besar secara cepat dan efisien (Putu et al., 2021). Tantangan utama dalam deteksi dan segmentasi bangunan pada citra satelit juga terletak pada variasi bentuk bangunan, keberadaan objek lain seperti vegetasi dan bayangan, serta kondisi atmosfer dan pencahayaan yang berubah-ubah (Rahman et al., 2021; Vasavi, Sri Somagani and Sai, 2023; Peng et al., 2024; Singla and Vaghela, 2024). Proses klasifikasi dan deteksi tersebut menuntut pemilihan model serta representasi data yang tepat, karena pemanfaatan data mentah dapat menimbulkan beban komputasi volumetrik yang sangat besar (Suputra et al., 2022). Oleh karena itu, pengelolaan jumlah data yang sangat besar dan beragam memerlukan teknologi yang mampu mengekstraksi serta menganalisis informasi dari berbagai sumber untuk menghasilkan solusi yang efisien dan tepat guna (Putu et al., 2021). Metode segmentasi tradisional seperti *thresholding* dan *edge detection* memiliki keterbatasan dalam menghadapi kompleksitas struktur dan kondisi lingkungan perkotaan. Segmentasi bangunan berukuran kecil masih menjadi tantangan karena hanya menempati sedikit piksel di latar belakang yang kompleks dan memiliki variasi intra-kelas yang tinggi(Chang et al., 2025). Sementara itu, pendekatan *machine learning* konvensional seperti *Random Forest* dan *Support Vector Machine (SVM)* masih memerlukan proses ekstraksi fitur secara manual, sehingga kurang efisien untuk menangani citra berskala besar dan beragam. Seiring perkembangan *deep learning*, arsitektur berbasis *Convolutional Neural Network* (CNN) telah menjadi standar dalam segmentasi citra satelit resolusi tinggi karena kemampuannya mengekstraksi fitur secara otomatis. Pendekatan *deep learning* telah menjadi metode utama dalam segmentasi citra penginderaan jauh karena kemampuannya mengekstraksi fitur secara mendalam dan melakukan optimasi *end-to-end* (Li et al., 2024). Namun, CNN konvensional masih terbatas dalam menangkap informasi spasial kompleks, sehingga arsitektur lanjutan seperti U-Net dan DeepLabV3+ dikembangkan untuk meningkatkan akurasi segmentasi (Zhang et al., 2025).

U-Net berbasis VGG16 dan DeepLabV3+ dengan ResNet-50 merupakan dua arsitektur *deep learning* yang banyak digunakan dalam segmentasi bangunan. Keunggulan utama U-Net terletak pada penggunaan *skip connections* yang membantu mempertahankan informasi resolusi tinggi selama proses segmentasi (Alsabhan, Alotaiby and Dudin, 2022). Setiap objek pada citra memiliki karakteristik bentuk yang unik dan dapat dikenali secara otomatis melalui deskriptor tiga dimensi; dengan deskriptor yang tepat, sistem komputer mampu membangun hubungan antara bentuk dan kelas objek menggunakan teknik *machine learning* (Suputra et al., 2022). U-Net berbasis VGG16 efektif dalam mempertahankan detail spasial melalui mekanisme *skip connections*, serta akurat dalam menangkap tekstur dan struktur bangunan (Ramalingam et al., 2024). VGG16 memiliki keunggulan dalam transfer learning, di mana model yang telah dilatih pada dataset besar seperti ImageNet dapat dimanfaatkan untuk meningkatkan kinerja segmentasi pada dataset citra satelit yang lebih spesifik (Gibril et al., 2024). Di sisi lain, DeepLabV3+ dengan ResNet-50 unggul dalam mengenali kontur objek di lingkungan perkotaan yang kompleks (Rahman et al., 2021). ResNet-50 menggunakan *residual learning* yang memungkinkan pelatihan model lebih stabil dan efisien dibandingkan arsitektur CNN lainnya (Li et al., 2024b). Penelitian sebelumnya melaporkan akurasi segmentasi mencapai 93,5% dengan kombinasi U-Net berbasis VGG16 (Singla and Vaghela, 2024). Meskipun telah menunjukkan performa yang baik, penerapan model-model ini masih menghadapi kendala pada wilayah tropis seperti Indonesia, terutama terkait ketersediaan dataset spesifik dan kebutuhan optimasi hyperparameter untuk kondisi citra satelit di kawasan padat penduduk. Selain itu, segmentasi bangunan

bertingkat masih menjadi tantangan akibat efek perspektif pada citra satelit beresolusi tinggi(Zhang et al., 2025). Penelitian ini bertujuan untuk membandingkan performa CNN konvensional, U-Net berbasis VGG16, dan DeepLabV3+ dengan ResNet-50 dalam segmentasi bangunan menggunakan citra satelit resolusi tinggi pada wilayah perkotaan padat. Evaluasi dilakukan menggunakan berbagai metrik akurasi untuk menganalisis efektivitas ketiga model, serta mengidentifikasi keunggulan dan keterbatasan masing-masing arsitektur. Hasil penelitian ini diharapkan dapat berkontribusi pada pengembangan sistem pemantauan perkotaan berbasis kecerdasan buatan (AI) yang mampu mendeteksi dan memetakan bangunan secara otomatis dari citra satelit resolusi tinggi. Dalam perencanaan kota, model ini dapat digunakan untuk membantu pemetaan tata ruang dan pengambilan kebijakan berbasis data spasial yang akurat (Mehta, Baz and Patel, 2024). Dengan demikian, hasil yang diperoleh dapat dimanfaatkan untuk mendukung pengelolaan tata ruang kota secara lebih akurat, efisien, dan berkelanjutan, terutama dalam perencanaan infrastruktur, mitigasi risiko bencana, serta pemantauan dinamika urbanisasi di wilayah perkotaan yang berkembang pesat.

2. METODE

2.1 Pengumpulan Data

Dataset penelitian ini mencakup 302 citra satelit resolusi tinggi dari kawasan Bali Selatan, yang diakuisisi secara manual melalui Google Earth Pro. Proses akuisisi distandardisasi dengan mengatur ketinggian pandang (*eye altitude*) secara konsisten pada 1 kilometer dan menggunakan resolusi spasial tertinggi yang tersedia. Pendekatan ini sangat penting untuk memastikan keseragaman skala dan ketajaman visual di seluruh dataset, yang menjadi dasar fundamental untuk analisis segmentasi yang akurat.

2.2 Labeling Data

Untuk menghasilkan ground truth yang akurat dan efisien, penelitian ini menerapkan alur kerja labeling semi-otomatis. Segmentasi awal dihasilkan secara otomatis menggunakan *Segment Anything Model* (SAM), yang kemudian hasilnya diperbaiki dan diperhalus secara manual dengan anotasi poligon menggunakan perangkat lunak VGG *Image Annotator* (VIA) yang menghasilkan gambar yang sudah dilabel seperti gambar 1. Untuk mengoptimalkan data pelatihan, jumlah titik pada setiap poligon dikurangi menggunakan algoritma simplifikasi kontur *Douglas-Peucker* tanpa menghilangkan detail bentuk bangunan secara signifikan. Hasil akhir dari proses ini adalah serangkaian anotasi dalam format JSON, yang memuat koordinat presisi dari setiap bangunan.

Gambar 1. Citra Asli dan *Binary mask*

2.3 Data Preprocessing

Tahap pra-pemrosesan data dimulai dengan mengonversi anotasi JSON menjadi *binary mask* (nilai 1 untuk bangunan, 0 untuk latar belakang). Untuk standardisasi dan efisiensi komputasi, seluruh citra dan *mask* diubah ukurannya menjadi 512 × 512 piksel. Proses normalisasi diterapkan secara selektif: *mask* dinormalisasi ke rentang [0, 1] guna menjaga stabilitas pelatihan, sementara citra asli dibiarkan pada nilai aslinya untuk mempertahankan detail visual. Normalisasi ini membantu mengurangi ketergantungan gradien terhadap skala nilai awal, sehingga memungkinkan penggunaan laju pembelajaran yang lebih tinggi tanpa risiko divergensi (Suputra et al., 2022).

Untuk meningkatkan *robustness* dan variasi data, diterapkan serangkaian augmentasi geometris secara acak. Representasi fitur tiga dimensi berbasis *surface curvature* digunakan untuk menggambarkan bentuk morfologis lokal suatu objek secara kuantitatif (Suputra et al.,

2022). Teknik augmentasi meliputi rotasi (10°), *flipping* (horizontal dan vertikal), *scaling* (10%), *shearing* (5%), dan translasi (5%). Seluruh proses ini memastikan model mampu beradaptasi terhadap berbagai variasi tampilan objek, sehingga meningkatkan kemampuan generalisasinya dalam tugas segmentasi.

2.4 Rancangan Model CNN Konvensional

Sebagai model dasar (*baseline*), penelitian ini mengimplementasikan arsitektur CNN Konvensional dengan 14 lapisan (10 konvolusi, 2 pooling, 2 *fully connected*), merujuk pada desain oleh Altun & Turker (2025). CNN bekerja dengan mengekstraksi fitur visual melalui serangkaian operasi konvolusi, pooling, dan aktivasi non-linear untuk mendeteksi pola dalam gambar secara hierarkis (Zeng, Guo and Li, 2022). Proses ekstraksi fitur dilakukan melalui serangkaian lapisan konvolusi 3x3 (ReLU) dan *max pooling* 2x2 untuk mereduksi dimensi spasial. Setelah proses konvolusi, lapisan pooling berfungsi untuk mengurangi dimensi data tanpa menghilangkan informasi penting, sehingga meningkatkan efisiensi komputasi dan mengurangi risiko *overfitting* (Zhang et al., 2021).

Kunci dari arsitektur ini untuk tugas segmentasi adalah penggunaan lapisan dekonvolusi (*transposed convolution*) setelah lapisan *fully connected* untuk merekonstruksi resolusi spasial dan menghasilkan peta segmentasi. Untuk menstabilkan pelatihan dan mencegah *overfitting*, model ini dilengkapi dengan *batch normalization* dan *dropout* (0.5) (Ayala et al., 2021). Meskipun CNN konvensional tidak memiliki mekanisme skip connections seperti U-Net atau ASPP seperti DeepLabV3+, optimalisasi dengan transfer learning dan *fine-tuning* pada dataset spesifik dapat meningkatkan akurasi segmentasi bangunan dari citra satelit resolusi tinggi (Sundaresan and Solomon, 2025). Output akhir berupa peta biner dihasilkan oleh lapisan konvolusi 1x1 dengan aktivasi sigmoid. Arsitektur ini sengaja tidak menyertakan mekanisme canggih seperti *skip connections* atau ASPP agar dapat berfungsi sebagai tolak ukur fundamental.

2.5 Rancangan Model VGG16 – U-Net

Model segmentasi ini dibangun menggunakan arsitektur U-Net yang memanfaatkan *backbone*VGG16 *pretrained* dari dataset ImageNet sebagai jalur encoder. Untuk memanfaatkan *transfer learning* secara efektif dan mencegah *overfitting*, 15 lapisan konvolusi awal VGG16 dibekukan (*non-trainable*), sementara lapisan yang lebih dalam dibiarkan dapat beradaptasi dengan tugas segmentasi citra satelit. Jalur decoder bertugas merekonstruksi resolusi spasial melalui operasi *upsampling*. Keunggulan utama arsitektur ini terletak pada mekanisme *skip connections* yang menggabungkan feature maps dari jalur encoder ke decoder. Mekanisme ini krusial untuk memulihkan detail spasial halus yang mungkin hilang selama proses *downsampling*.

Untuk meningkatkan stabilitas dan regularisasi, model ini menerapkan *batch normalization* dan *dropout* (0.5). Output akhir berupa *binary mask* probabilitas dihasilkan oleh lapisan konvolusi 1×1 dengan aktivasi sigmoid. Seluruh model dilatih menggunakan optimizer Adam dengan *learning rate* sebesar 1×10^{-4}

2.6 Rancangan Model ResNet-50–DeepLabV3+

Model selanjutnya adalah DeepLabV3+, sebuah arsitektur canggih yang dirancang untuk segmentasi semantik multi-skala. Jalur encoder-nya memanfaatkan *backbone*ResNet-50 *pretrained*, di mana lapisan-lapisan awal dibekukan untuk menerapkan *transfer learning*. Fitur yang diekstraksi dari ResNet-50 kemudian diproses oleh modul inti arsitektur ini, yaitu *Atrous Spatial Pyramid Pooling* (ASPP). ASPP menggunakan konvolusi dilatasi (*dilated convolution*) dengan laju berbeda untuk menangkap informasi kontekstual pada berbagai

skala, yang sangat efektif untuk mengidentifikasi bangunan dengan ukuran yang bervariasi. Jalur decoder pada DeepLabV3+ dirancang untuk merekonstruksi peta segmentasi secara detail. Proses ini menggabungkan (*concatenated*) fitur dari ASPP yang telah di-*upsampling* dengan fitur tingkat rendah (*low-level features*) yang diambil dari lapisan awal *backbone* ResNet-50. Penggabungan ini krusial untuk memulihkan detail spasial dan menghasilkan batas- batas objek yang tajam.

Output akhir untuk klasifikasi piksel-ke-piksel (bangunan atau latar belakang) menggunakan fungsi aktivasi *softmax*. Model ini dilatih menggunakan fungsi kerugian *dice loss*, yang efektif menangani potensi ketidakseimbangan kelas, serta dioptimalkan menggunakan optimizer Adam dengan *learning rate* awal 1×10−4 yang diturunkan secara bertahap

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi dan Dataset

Dataset penelitian ini berasal dari 302 citra satelit resolusi tinggi dari wilayah Bali Selatan yang diakuisisi melalui Google Earth Pro. Setiap citra awal berukuran 2048×2048 piksel dipotong menjadi empat bagian, menghasilkan total 1.216 *patch* citra berukuran 1024×1024 piksel. Dataset ini kemudian dibagi secara proporsional menjadi tiga subset: 848 *patch* (70%) untuk data latih, 120 *patch* (10%) untuk data validasi, dan 248 *patch* (20%) untuk data uji guna memastikan perbandingan kinerja model yang objektif. Pembuatan *ground truth* dilakukan dengan mengonversi anotasi poligon menjadi *binary mask*, di mana strategi labeling hibrida diterapkan untuk menyeimbangkan efisiensi dan akurasi data latih. Pendekatan ini memanfaatkan *weak* labeling (anotasi area umum) untuk mempercepat proses anotasi, sejalan dengan prinsip efisiensi dalam pengolahan data berskala besar yang bertujuan meningkatkan efektivitas kegiatan serta mengurangi biaya (Putu et al., 2021). Sementara itu, data validasi dan uji menggunakan *dense* labeling (anotasi poligon detail) untuk memastikan evaluasi model yang presisi dan dapat diandalkan. Secara keseluruhan, proses ini menghasilkan 121.166 poligon anotasi pada seluruh dataset, yang menjadi dasar *ground truth* dalam pelatihan dan evaluasi ketiga model secara konsisten.

3.2 Desain dan Impelementasi Model

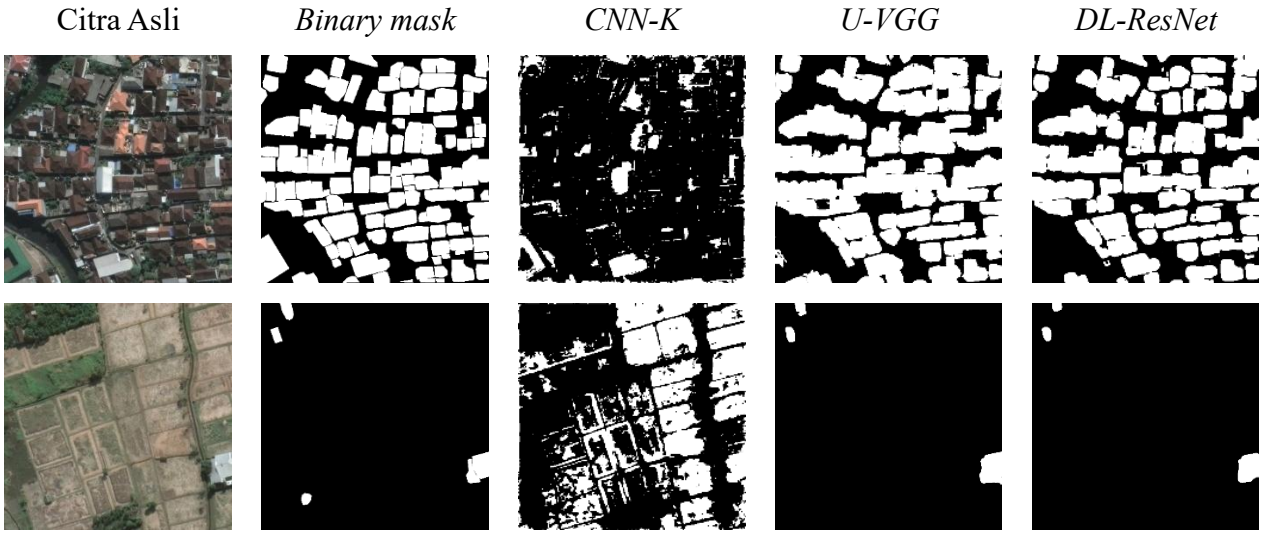
Seluruh model dilatih menggunakan ukuran input 512×512 piksel, *batch size* 2, optimizer Adam, serta teknik augmentasi data yang sama (rotasi, *flipping*, *scaling*, dll.). Model pertama, *CNN-K*, berfungsi sebagai *baseline* dengan arsitektur encoder-decoder sederhana yang menggunakan lapisan Conv2D, Conv2DTranspose, dan *skip connections*; model ini dilatih dengan *learning rate* 1×10−3 dan fungsi kerugian *dice loss*. Model kedua, *U-VGG*, mengadopsi struktur U-Net dengan *backbone* VGG16 *pretrained* sebagai encoder, memanfaatkan *skip connections* untuk mempertahankan detail spasial, dan dilatih dengan *learning rate* 1×10−4 serta fungsi kerugian *Dice Loss*. Arsitektur ketiga yang paling canggih, *DL-ResNet*, menggunakan DeepLabV3+ dengan *backbone*ResNet-50 pra-terlatih, yang keunggulannya terletak pada modul Atrous Spatial Pyramid Pooling (ASPP) untuk menangkap konteks multi-skala. *DL-ResNet* dilatih dengan *learning rate* 1×10−4 (disertai *callback ReduceLROnPlateau*) dan menggunakan fungsi *loss* gabungan (*dice* l*oss* + *binary cross entropy*) untuk hasil yang lebih seimbang. Perbandingan ketiga model ini bertujuan untuk mengukur peningkatan performa dari arsitektur dasar hingga yang paling kompleks pada tugas segmentasi bangunan.

3.3 Hasil Visualisasi Segmentasi

Evaluasi visual dilakukan pada dua skenario yang berbeda, yaitu wilayah dengan kepadatan bangunan tinggi dan wilayah dengan kepadatan

rendah, untuk menguji kemampuan generalisasi dan robustisitas setiap arsitektur secara komprehensif. Hasil dari model dasar *CNN-K* menunjukkan performa yang sangat terbatas di kedua skenario. Pada wilayah padat, model ini mengalami *under-segmentation* yang signifikan, di mana area bangunan yang terdeteksi tampak terputus-putus dan gagal merepresentasikan kepadatan urban secara akurat. Pola kesalahan ini menunjukkan bahwa *CNN-K* memiliki keterbatasan dalam mengenali struktur bangunan yang berdekatan atau saling menempel, terutama di area perkotaan dengan tingkat heterogenitas tinggi. Kondisi ini bahkan lebih parah pada wilayah dengan kepadatan rendah, di mana *CNN-K* menunjukkan kegagalan ekstrem untuk mendeteksi sebagian besar bangunan, termasuk struktur berukuran besar dan jelas yang seharusnya mudah diidentifikasi. Hal tersebut mengindikasikan bahwa model ini tidak memiliki kemampuan generalisasi yang memadai terhadap variasi tekstur dan bentuk bangunan. Secara konsisten, *CNN-K* terbukti tidak mampu memetakan bangunan secara utuh dan andal di berbagai kondisi spasial.

Sebaliknya, model *U-VGG* menunjukkan peningkatan performa yang sangat signifikan dibandingkan dengan baseline. Pada area padat, segmentasi yang dihasilkan jauh lebih lengkap dan menyeluruh, dengan bentuk dan batas antar bangunan yang relatif jelas serta minim kesalahan prediksi. Kemampuan *U-VGG* dalam mempertahankan detail spasial melalui mekanisme *skip connection* membuatnya mampu merekonstruksi kontur objek dengan presisi tinggi. Selain itu, model ini juga menunjukkan stabilitas yang lebih baik pada area dengan variasi pencahayaan dan tekstur permukaan, yang menjadikannya lebih adaptif terhadap kompleksitas citra satelit resolusi tinggi.



Gambar 2. Citra Asli dan *Binary mask*, *CNN-K*, *U-VGG*, dan *DL-ResNet* pada Wilayah Bangunan Padat dan Minim Bangunan

Meskipun demikian, pada beberapa kasus masih terlihat adanya sedikit *bleeding* atau penggabungan antar bangunan yang sangat berdekatan. Pada wilayah berkepadatan rendah, *U-VGG* mampu mengenali sebagian besar struktur bangunan utama dengan baik, namun terkadang menghasilkan sedikit *noise* atau prediksi palsu pada area vegetasi dan lahan kosong. Secara keseluruhan, *U-VGG* terbukti sebagai model yang kuat dan stabil, mampu menghasilkan segmentasi yang representatif di kedua kondisi. Model *DL-ResNet* secara

konsisten menunjukkan performa segmentasi yang paling optimal dan presisi. Pada wilayah padat, model ini unggul dalam hal detail, ketepatan kontur, dan pemisahan antar bangunan yang sangat rapat, baik untuk objek berukuran besar maupun kecil. Keunggulan ini juga terlihat pada wilayah berkepadatan rendah, di mana hasil segmentasinya tampak bersih dengan noise yang sangat minim dan mampu menghindari prediksi palsu pada area non-bangunan secara efektif. *DL-ResNet* menunjukkan kemampuan generalisasi superior dalam menangani variasi bentuk, ukuran, dan kepadatan bangunan di berbagai lingkungan. Secara keseluruhan, analisis visual ini menguatkan hasil evaluasi metrik kuantitatif. Terdapat hierarki performa yang jelas: *DL-ResNet* sebagai model terbaik, diikuti oleh *U-VGG* yang sangat kompeten, dan *CNN-K* yang berkinerja paling rendah. Temuan ini menegaskan bahwa arsitektur yang lebih kompleks, yang memanfaatkan *backbone pretrained* dan mekanisme penangkapan fitur multi-skala seperti ASPP, memiliki keunggulan signifikan untuk tugas pemetaan bangunan dari citra satelit resolusi tinggi.

3.4 Interpretasi Hasil Evaluasi

Analisis kuantitatif pada data uji, sebagaimana dirangkum dalam Tabel 1, menunjukkan perbedaan performa yang jelas di antara ketiga model. Model *U-VGG* menunjukkan kinerja terbaik dan paling stabil, dengan capaian *IoU* 81% dan *Dice Coefficient* 89%, membuktikan kemampuan generalisasi yang sangat baik tanpa *overfitting*. Diikuti oleh model *DL-ResNet*, yang juga menunjukkan performa sangat kuat dengan *IoU* 80% dan *Dice Coefficient* 85%. Sebaliknya, model dasar *CNN-K* berkinerja paling rendah dengan *IoU* hanya 26% dan *Dice Coefficient* 40%, yang mengindikasikan terjadinya *overfitting* karena kegagalan generalisasi pada data uji. Secara keseluruhan, evaluasi metrik ini mengukuhkan bahwa arsitektur canggih seperti *U-VGG* dan *DL-ResNet* jauh lebih unggul daripada CNN konvensional dan sangat cocok untuk tugas segmentasi bangunan di lingkungan urban yang kompleks.

Tabel 1. Performa Model

Model	Dataset	Accuracy (%)	*Dice* (%)	*IoU* (%)	*Loss* (%)	*Precision* (%)	*Recall* (%)	*F1-score* (%)
CNN-K	*Training*	84	83	73	16	82	86	83
	Validation	86	87	77	12	86	88	87
	Testing	68	40	26	59	69	30	40
U-VGG	*Training*	83	83	73	16	81	85	83
	Validation	86	88	79	11	82	95	88
	Testing	92	89	81	10	87	93	89
DL-RESNET	*Training*	87	83	77	42	87	87	86
	Validation	88	85	79	42	91	85	88
	Testing	91	85	80	35	91	87	88

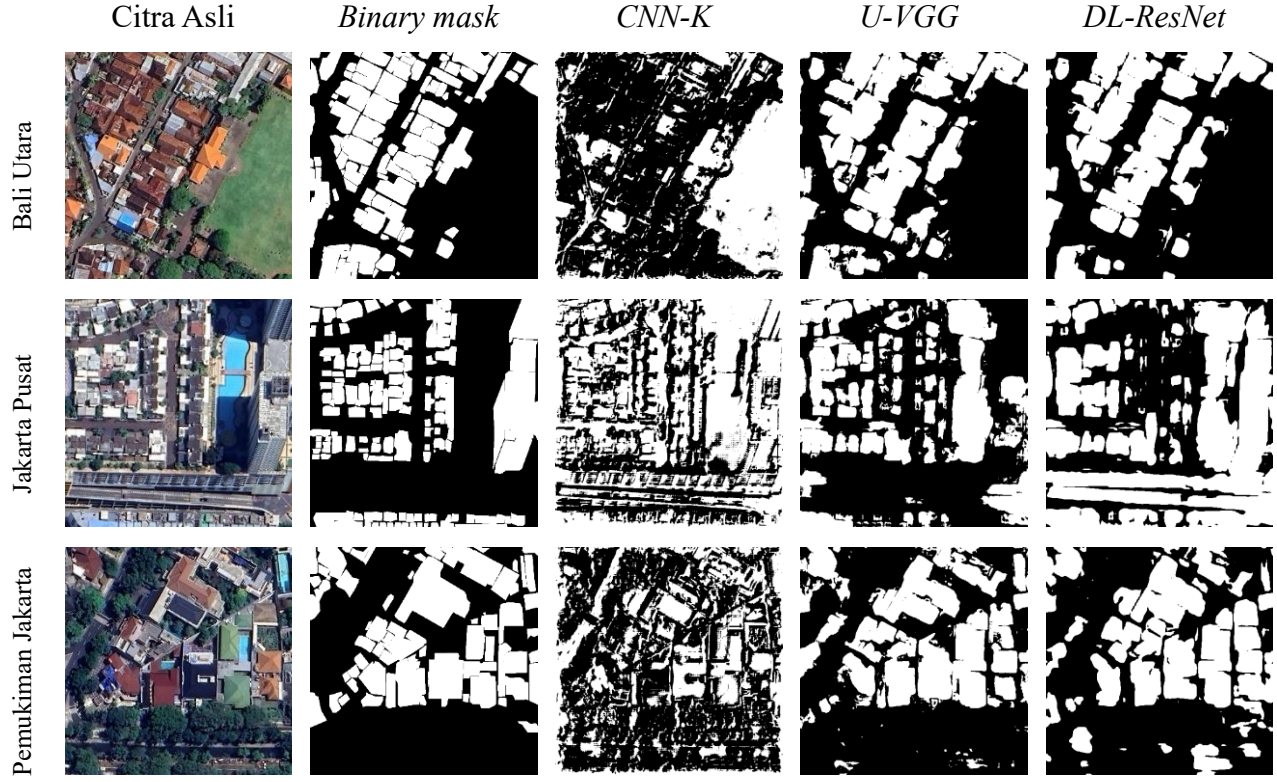
3.5 Evaluasi Generalisasi pada Wilayah Lain

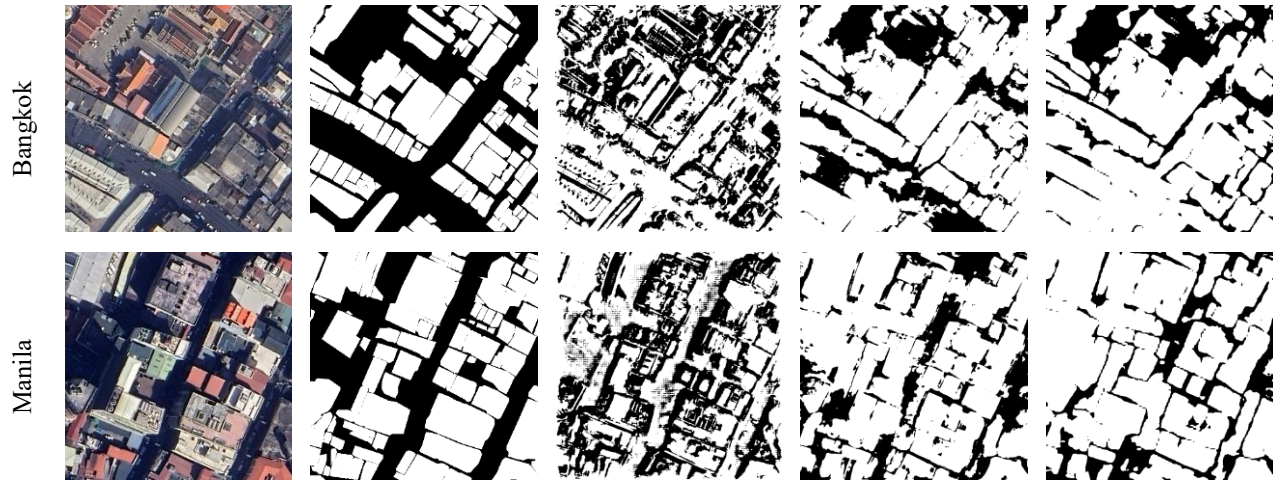
Untuk menguji kemampuan generalisasi, ketiga model dievaluasi pada lima wilayah

geografis yang beragam di luar data latih: Bali Utara, Jakarta (gedung tinggi dan permukiman padat), Bangkok, dan Manila, guna menilai kemampuan adaptasi terhadap berbagai tipologi bangunan, kondisi pencahayaan, serta variasi tekstur permukaan. Hasil pengujian menunjukkan pola kinerja yang konsisten untuk setiap model, mencerminkan perbedaan mendasar dalam kemampuan representasi spasial dan semantik masing-masing arsitektur.

Model dasar *CNN-K* secara konsisten gagal melakukan generalisasi, ditandai dengan *under-segmentation* yang signifikan di seluruh lokasi uji. Model ini tampak tidak mampu menangkap konteks spasial antar objek, sehingga sering kali mengabaikan bangunan yang saling berdekatan atau tertutup bayangan. Sebaliknya, *U-VGG* menunjukkan kemampuan generalisasi yang kuat dan stabil, menghasilkan segmentasi dengan kontur yang rapi serta representasi bentuk yang utuh, meskipun terkadang mengalami sedikit *bleeding* (penggabungan bangunan) di area padat. Sementara itu, *DL-ResNet* juga memperlihatkan generalisasi yang sangat baik dengan tingkat detail tertinggi, terutama dalam mempertahankan struktur bangunan di area berbayang atau dengan pencahayaan kontras, namun memiliki kecenderungan minor untuk melakukan *over-segmentation* di sekitar vegetasi.

Secara keseluruhan, baik *U-VGG* maupun *DL-ResNet* terbukti mampu beradaptasi dengan baik pada lingkungan baru yang belum pernah dilatihkan, jauh melampaui performa *CNN-K*. Kedua model ini dapat dianggap sebagai kandidat yang *robust* untuk aplikasi segmentasi di berbagai wilayah geografis, dengan *U-VGG* unggul dalam konsistensi spasial dan *DL-ResNet* menonjol dalam ketepatan detail morfologis.





Gambar 3. Hasil Segmentasi pada Berbagai Daerah

4. PENUTUP

4.1 Kesimpulan

Arsitektur U-Net dengan *backbone* VGG16 (*U-VGG*) menunjukkan performa terbaik secara keseluruhan, diikuti oleh DeepLabV3+ dengan ResNet-50 (*DL-ResNet*), sedangkan CNN Konvensional (*CNN-K*) menampilkan kinerja terendah. Model *U-VGG* unggul dalam hal stabilitas, akurasi metrik yang tinggi, serta konsistensi dalam mempertahankan kontur bangunan yang tajam di berbagai wilayah uji. Sementara itu, *DL-ResNet* juga menunjukkan kemampuan generalisasi yang kuat dengan tingkat detail segmentasi yang sangat baik, meskipun masih terdapat kecenderungan kecil untuk melakukan *over-segmentation* pada area berbayang. Meskipun kedua model canggih tersebut mengalami sedikit penurunan performa ketika diuji di luar wilayah pelatihan, keduanya tetap menunjukkan kemampuan adaptasi yang baik berbeda dengan *CNN-K* yang gagal melakukan generalisasi. Oleh karena itu, penelitian ini merekomendasikan *U-VGG* sebagai model yang paling seimbang dan andal untuk tugas segmentasi bangunan, sementara *DL-ResNet* menjadi alternatif yang sangat kompetitif ketika prioritas utama adalah ketelitian detail pada kawasan padat.

Hasil segmentasi yang diperoleh dari kedua model ini memiliki potensi penerapan langsung dalam berbagai konteks dunia nyata, seperti perencanaan tata ruang, pemantauan perkembangan permukiman, serta analisis risiko bencana di kawasan perkotaan padat. Dengan demikian, penelitian ini tidak hanya berkontribusi terhadap pengembangan metode segmentasi citra satelit berbasis *deep learning*, tetapi juga mendukung penerapan kecerdasan buatan dalam pengelolaan dan perencanaan wilayah perkotaan di Indonesia secara lebih akurat dan berkelanjutan.

4.2. Saran

Disarankan beberapa arah pengembangan untuk meningkatkan *robustness* dan efisiensi model. Pertama, perlu dilakukan perluasan dataset dengan cakupan yang lebih luas dan beragam, mencakup berbagai kondisi cuaca, waktu pengambilan citra, serta tipologi kawasan urban yang berbeda, guna memperkuat kemampuan generalisasi model. Karena dataset dalam penelitian ini masih terbatas pada wilayah Bali Selatan, penelitian lanjutan dapat mempertimbangkan penambahan data dari berbagai wilayah tropis di Indonesia seperti Sumatera, Kalimantan, dan Sulawesi. Keragaman karakteristik bangunan, vegetasi, serta kondisi atmosfer di wilayah-wilayah tersebut berpotensi meningkatkan kemampuan adaptasi model terhadap konteks geografis yang lebih luas, sehingga hasil segmentasi menjadi lebih representatif secara nasional.

Selain perluasan dataset, penelitian selanjutnya juga dapat berfokus pada eksplorasi optimasi *hyperparameter* secara lebih sistematis, seperti penyesuaian *learning rate scheduler*, pemilihan *optimizer* alternatif, serta pengaturan *batch size* yang lebih adaptif terhadap kompleksitas data. Upaya tersebut diharapkan dapat meningkatkan efisiensi proses pelatihan

sekaligus menjaga stabilitas konvergensi model, terutama pada dataset berskala besar dan heterogen.

5. DAFTAR PUSTAKA

Alsabhan, W., Alotaiby, T. and Dudin, B., 2022. Detecting Buildings and Nonbuildings from Satellite Images Using U-Net. *Computational Intelligence and Neuroscience*, 2022. https://doi.org/10.1155/2022/4831223.

Altun, M. and Turker, M., 2025. Integration of convolutional neural networks with parcel-based image analysis for crop type mapping from time-series images. *Earth Science Informatics*, [online] 18(3). https://doi.org/10.1007/s12145-025-01819-8.

Ayala, C., Sesma, R., Aranda, C. and Galar, M., 2021. A deep learning approach to an enhanced building footprint and road detection in high-resolution satellite imagery. *Remote Sensing*, 13(16), pp.1–21. https://doi.org/10.3390/rs13163135.

Chang, F., Ma, T., Wang, D., Zhu, S., Li, D., Feng, S. and Fan, X., 2025. Method for building segmentation and extraction from high-resolution remote sensing images based on improved YOLOv5ds. *PLoS ONE*, 20(3 March). https://doi.org/10.1371/JOURNAL.PONE.0317106.

Gibril, M.B.A., Al-Ruzouq, R., Shanableh, A., Jena, R., Bolcek, J., Shafri, H.Z.M. and Ghorbanzadeh, O., 2024. Transformer-based semantic segmentation for large-scale building footprint extraction from very-high resolution satellite images. *Advances in Space Research*, [online] 73(10), pp.4937–4954. https://doi.org/10.1016/j.asr.2024.03.002.

Li, J., Cai, Y., Li, Q., Kou, M. and Zhang, T., 2024a. A review of remote sensing image segmentation by deep learning methods. *International Journal of Digital Earth*, [online] 17(1). https://doi.org/10.1080/17538947.2024.2328827.

Li, Z.H., Shi, A.C., Xiao, H.X., Niu, Z.H., Jiang, N., Li, H.B. and Hu, Y.X., 2024b. Robust Landslide Recognition Using UAV Datasets: A Case Study in Baihetan Reservoir. *Remote Sensing*, 16(14). https://doi.org/10.3390/rs16142558.

Mehta, Y., Baz, A. and Patel, S.K., 2024. Semantic segmentation of optical satellite images for the illegal construction detection using transfer learning. *Results in Engineering*, [online] 24, p.103383. https://doi.org/10.1016/J.RINENG.2024.103383.

Peng, F., Yao, S., Chen, Y. and Li, W., 2024. Unsupervised Domain Adaptive Transfer Learning for Urban Built-Up Area Extraction. p.10. https://doi.org/10.3390/proceedings2024110010.

Putu, N., Dewi, N.P., Budhi, P. and Purwanta, D., 2021. *Big Data for Indonesian Marine Fisheries A Preliminary Research Plan*.

Rahman, A.M., Zaber, M., Cheng, Q., Nayem, A.B.S., Sarker, A., Paul, O. and Shibasaki, R., 2021. Applying state-of-the-art deep-learning methods to classify urban cities of the developing world. *Sensors*, 21(22), pp.1–22. https://doi.org/10.3390/s21227469.

Ramalingam, A., Srivastava, V., George, S., Alagala, S. and Martin Leo Manickam, J., 2024. Building rooftop extraction from aerial imagery using low complexity UNet variant models. *Journal of Spatial Science*, 69, pp.1–28. https://doi.org/10.1080/14498596.2024.2302166.

Singla, J.G. and Vaghela, B., 2024. Semantic segmentation on multi-resolution optical and microwave data using deep learning. [online] pp.0–3. Available at: <https://arxiv.org/pdf/2411.07581>.

Sundaresan, A.A. and Solomon, A.A., 2025. Post-disaster flooded region segmentation using DeepLabv3+ and unmanned aerial system imagery. *Natural Hazards Research*, [online] 5(2), pp.363–371. https://doi.org/10.1016/j.nhres.2024.12.003

Suputra, P.H., Sensusiati, A.D., Artaria, M.D., Verkerke, G.J., Yuniarno, E.M. and Purnama, I.K.E., 2022. Automatic 3D Cranial Landmark Positioning based on Surface Curvature Feature using Machine Learning. *Knowledge Engineering and Data*

Science, 5(1), p.27. https://doi.org/10.17977/um018v5i12022p27-40.

Tang, Y., Zhang, G., Liu, J.K. and Qin, R., 2025. Weakly supervised land-cover classification of high-resolution images with low-resolution labels through optimized label refinement. *International Journal of Remote Sensing*, [online] 46(5), pp.1913–1937. https://doi.org/10.1080/01431161.2024.2443612.

Vasavi, S., Sri Somagani, H. and Sai, Y., 2023. Classification of buildings from VHR satellite images using ensemble of U-Net and ResNet. *Egyptian Journal of Remote Sensing and Space Science*, [online] 26(4), pp.937–953. https://doi.org/10.1016/j.ejrs.2023.11.008.

Zeng, Y., Guo, Y. and Li, J., 2022. Recognition and extraction of high-resolution satellite remote sensing image buildings based on deep learning. *Neural Computing and Applications*, [online] 34(4), pp.2691–2706. https://doi.org/10.1007/s00521-021-06027-1.

Zhang, J., Li, Y., Yang, X., Jiang, R. and Zhang, L., 2025. RSAM-Seg: A SAM-Based Model with Prior Knowledge Integration for Remote Sensing Image Semantic Segmentation. *Remote Sensing*, 17(4), pp.1–26. https://doi.org/10.3390/rs17040590.

Zhang, T., Tang, H., Ding, Y., Li, P., Ji, C. and Xu, P., 2021. Fsrss-net: High-resolution mapping of buildings from middle-resolution satellite images using a super-resolution semantic segmentation network. *Remote Sensing*, 13(12). https://doi.org/10.3390/rs13122290.