

## Segmentasi Atap Bangunan dari Citra Satelit Resolusi Tinggi Menggunakan Arsitektur CBAM-UNet

### *Segmentation of Building Roofs from High-Resolution Satellite Imagery Using CBAM-UNet Architecture*

Muhammad Azhar Rafi<sup>1</sup> Samudra<sup>1</sup>, Harintaka<sup>1\*</sup>

<sup>1</sup>Magister Teknik Geomatika, Departemen Teknik Geodesi, Fakultas Teknik, Universitas Gadjah Mada, Jl. Grafika No. 2, Yogyakarta, 55281

\*Corresponding Author: [harintaka@ugm.ac.id](mailto:harintaka@ugm.ac.id)

#### Article Info:

Received: 19 – 01 - 2026

Accepted: 10 – 03 - 2026

Published: 25 – 07 - 2026

**Kata kunci:** segmentasi, atap bangunan, CBAM-UNet, IoU, F1-Score

**Keywords:** segmentation, building roof, CBAM-UNet, IoU, F1-Score

**Abstrak:** Bangunan merupakan salah satu unsur peta dasar yang berkembang pesat, sehingga memerlukan pembaruan data secara berkala. Hingga saat ini proses pembaruan informasi bangunan masih banyak mengandalkan interpretasi visual atap bangunan melalui citra satelit resolusi tinggi secara manual, yang membutuhkan waktu lama serta berpotensi menimbulkan inkonsistensi antar operator. Perkembangan teknologi penginderaan jauh dan metode *deep learning* membuka peluang untuk otomatisasi proses segmentasi atap bangunan secara lebih cepat, obyektif, dan efisien. Berbagai arsitektur *deep learning* telah dikembangkan untuk segmentasi atap bangunan, seperti FCN, SegNet, DeepLab, Mask R-CNN, dan U-Net. Di antara arsitektur tersebut, U-Net menjadi pilihan utama karena struktur *encoder-decoder* yang simetris memungkinkan ekstraksi fitur yang efektif bahkan dengan jumlah sampel data yang terbatas. Meskipun berbagai varian U-Net telah dikembangkan untuk menangani data kompleks, U-Net konvensional masih memiliki keterbatasan dalam mempertahankan detail spasial pada objek dengan geometri rumit dan variasi ukuran yang ekstrem. Penelitian ini bertujuan untuk mengoptimalkan arsitektur U-Net melalui integrasi *Convolutional Block Attention Module* (CBAM). CBAM dipilih karena kemampuannya dalam memberikan bobot secara adaptif pada *channel* dan spasial secara bersamaan. Hal ini memungkinkan model untuk lebih fokus pada fitur relevan dan mengabaikan *noise*, sehingga meningkatkan akurasi segmentasi atap bangunan pada citra satelit resolusi tinggi beresolusi spasial 50 cm. Model CBAM-UNet dilatih menggunakan data *training* yang teraugmentasi, kemudian diuji pada lima area dengan karakteristik bentuk dan kepadatan bangunan yang berbeda. Hasil penelitian menunjukkan bahwa CBAM-UNet memberikan performa yang lebih baik dibandingkan U-Net konvensional.

**Abstract:** Buildings are one of the fundamental elements of base maps that undergo rapid development, thus requiring periodic data updates. To date, the process of updating building information still largely relies on manual visual interpretation of rooftops using high-resolution satellite imagery, which is time-consuming and prone to inconsistencies between operators. Advances in remote sensing technology and deep learning methods have opened opportunities to automate rooftop segmentation processes in a faster, more objective, and more efficient manner. Various deep learning architectures have been developed for rooftop segmentation, such as FCN, SegNet, DeepLab, Mask R-CNN, and U-Net. Among these architectures, U-Net has become a primary choice due to its symmetric encoder-decoder structure, which enables effective feature extraction even with a limited number of training samples. Although various U-Net variants have been developed to handle complex data, the conventional U-Net still has limitations in preserving spatial details for objects with intricate geometries and extreme size variations. This study aims to optimize the U-Net architecture through the integration of the Convolutional Block Attention Module (CBAM). CBAM is selected for its ability to adaptively assign weights to both channel and spatial dimensions simultaneously. This capability allows the model to focus more effectively on relevant features while suppressing noise, thereby improving rooftop segmentation accuracy on high-resolution satellite imagery with a spatial resolution of 50 cm. The CBAM-UNet model was trained using augmented training data, and subsequently tested on five areas with varying building shapes and density characteristics. The results demonstrate that CBAM-UNet achieves better performance compared to the conventional U-Net.

#### How to Cite:

Samudra M. A. R., Harintaka (2026). Segmentasi Atap Bangunan dari Citra Satelit Resolusi Tinggi Menggunakan Arsitektur CBAM-UNet. *Jurnal Ilmiah Geomatika*, 6(1), 1-13. <https://doi.org/10.31315/imagi.v6i1.16388>

## 1. PENDAHULUAN

Perkembangan teknologi penginderaan jauh dalam beberapa dekade terakhir menunjukkan kemajuan yang sangat signifikan, baik dari sisi kualitas sensor maupun wahana satelit. Peningkatan resolusi spasial, spektral, dan radiometrik pada sensor satelit telah berdampak langsung pada peningkatan kualitas citra yang dihasilkan. Teknologi ini memungkinkan akuisisi data permukaan bumi dengan tingkat ketelitian yang tinggi, sebagaimana ditunjukkan oleh citra satelit resolusi tinggi yang mampu merekam detail objek secara akurat. Dengan kemampuan tersebut, citra satelit resolusi tinggi telah dimanfaatkan secara luas dalam berbagai kajian, antara lain analisis dinamika garis pantai (Apostolopoulos & Nikolakopoulos, 2020), pemetaan dan analisis penggunaan lahan (Li dkk., 2020), serta kajian kebencanaan dan mitigasi risiko (Brahmantara dkk., 2022). Selain itu, citra resolusi tinggi juga menjadi sumber data utama dalam pembentukan peta dasar (Hakim dkk., 2019).

Citra satelit mengandung beragam informasi yang dapat diekstraksi melalui proses interpretasi citra. Interpretasi citra didefinisikan sebagai proses mempelajari citra untuk mengenali objek serta memperoleh informasi atau makna yang terkait dengan objek tersebut (Seidlova dkk., 2021). Proses interpretasi dapat dilakukan secara manual oleh manusia atau secara otomatis dengan bantuan komputer. Interpretasi manual sangat bergantung pada kemampuan visual, pengalaman, serta subjektivitas operator. Faktor-faktor tersebut menyebabkan hasil interpretasi berpotensi mengalami inkonsistensi antar-operator, terutama pada area yang kompleks. Sebaliknya, interpretasi otomatis berbasis komputer menawarkan konsistensi yang lebih tinggi melalui standarisasi algoritma dan pemrosesan data. Namun, objektivitas hasilnya tetap dipengaruhi oleh desain algoritma dan bias yang mungkin terdapat dalam data *training*.

Bangunan merupakan salah satu objek utama yang diinterpretasi dari citra satelit dan berperan penting sebagai elemen peta dasar. Berdasarkan Pasal 12 Undang-Undang Nomor 4 Tahun 2011 tentang Informasi Geospasial, bangunan termasuk dalam unsur peta dasar yang wajib tersedia dan diperbarui. Sebagai salah satu jenis tutupan lahan yang mengalami perubahan relatif cepat, bangunan memiliki dinamika yang tinggi, baik dari segi jumlah, bentuk, maupun fungsi. Bangunan mencakup berbagai kategori, mulai dari permukiman, fasilitas peternakan, hingga kawasan industri. Tingginya laju pembangunan dan keragaman fungsi tersebut menjadikan data bangunan sebagai unsur peta dasar yang memerlukan pembaruan secara berkala dan berkelanjutan. Oleh karena itu, diperlukan metode ekstraksi informasi bangunan yang akurat, cepat, dan efisien dari citra satelit resolusi tinggi. Ekstraksi informasi bangunan ini dilakukan berdasar informasi visual atap bangunan dari citra.

Pada praktiknya, ekstraksi unsur-unsur peta dasar, termasuk bangunan, masih banyak dilakukan melalui digitasi manual menggunakan foto udara atau citra satelit resolusi tinggi (Kristal & Harintaka, 2022). Meskipun metode ini mampu menghasilkan detail geometrik yang baik, proses digitasi manual membutuhkan waktu yang relatif lama dan sumber daya manusia yang besar. Selain itu, kualitas hasil digitasi sangat dipengaruhi oleh keterampilan dan konsistensi operator dalam menginterpretasi objek pada citra. Kondisi ini mendorong perlunya pendekatan alternatif yang lebih objektif, cermat, dan efisien untuk ekstraksi fitur bangunan.

Pendekatan ekstraksi atap bangunan secara otomatis dikembangkan untuk mengatasi keterbatasan interpretasi manual. Otomatisasi memungkinkan proses ekstraksi dilakukan dengan efisiensi tinggi pada cakupan wilayah yang luas, memberikan hasil yang lebih seragam dibandingkan dengan interpretasi manual yang cenderung bervariasi tergantung pada tingkat ketelitian operator. Salah satu pendekatan yang banyak digunakan dalam ekstraksi objek otomatis adalah *deep learning*, yang merupakan pengembangan dari komputasi berbasis jaringan saraf tiruan (*neural network*). Pendekatan ini terbukti efektif dalam pengenalan pola, klasifikasi, dan segmentasi citra.

Sejalan dengan perkembangan tersebut, beberapa penelitian sebelumnya telah mengeksplorasi otomatisasi ekstraksi bangunan menggunakan berbagai pendekatan dan sumber data. Segmentasi multi-resolusi otomatis menggunakan data foto udara dan LiDAR telah diterapkan dan menghasilkan akurasi yang tinggi (Julian & Harintaka, 2019). Meski demikian, kombinasi data LiDAR belum tersedia secara merata di berbagai wilayah, sehingga penerapannya masih terbatas. Penelitian ekstraksi tapak bangunan

berbasis *deep learning* juga telah dilakukan, antara lain menggunakan arsitektur *Segment Anything Model* (SAM) (Baroroh & Harintaka, 2025) dan Mask R-CNN (Kristal & Harintaka, 2022), yang keduanya memanfaatkan data foto udara. Selain itu, perbandingan hasil ekstraksi objek bangunan menggunakan Mask R-CNN dari data citra satelit dan foto udara telah dilakukan oleh Susetyo, Harintaka, & Rizaldy (2023). Penelitian tersebut menunjukkan bahwa penggunaan data citra satelit resolusi tinggi menghasilkan performa yang lebih rendah dibandingkan dengan foto udara. Temuan-temuan tersebut menunjukkan bahwa meskipun foto udara mampu menghasilkan akurasi yang tinggi, tantangan masih muncul ketika menggunakan data dengan resolusi yang lebih rendah. Oleh karena itu, fokus pada penelitian ini adalah untuk menghasilkan akurasi segmentasi atap bangunan yang tinggi menggunakan data dengan resolusi lebih rendah dibandingkan foto udara, yaitu citra satelit resolusi tinggi (resolusi spasial 50 centimeter, dibandingkan foto udara yang biasanya di bawah 15 cm), sehingga diharapkan dapat menjadi solusi yang lebih praktis dan aplikatif pada skala wilayah yang luas.

Dalam beberapa tahun terakhir, teknologi *deep learning* telah memberikan kontribusi besar, khususnya dalam bidang pemrosesan citra, pengenalan pola, dan pengenalan objek (Wu dkk., 2019). Sejalan dengan itu, penerapan *deep learning* dalam segmentasi citra satelit menunjukkan performa yang sangat menjanjikan. Salah satu arsitektur yang paling banyak digunakan untuk segmentasi citra adalah U-Net. Model ini pertama kali diperkenalkan oleh Ronneberger dkk. (2015) untuk segmentasi citra medis, namun fleksibilitas dan efektivitasnya menjadikan U-Net populer dalam berbagai bidang, termasuk penginderaan jauh.

U-Net merupakan arsitektur *Convolutional Neural Network* (CNN) yang dirancang khusus untuk segmentasi citra. Struktur U-Net terdiri atas dua bagian utama, yaitu *encoder* dan *decoder* (Siddique dkk., 2021). Bagian *encoder* berfungsi untuk menangkap konteks global dengan menurunkan resolusi citra dan mengekstraksi fitur penting melalui operasi konvolusi dan *max-pooling*. Sementara itu, bagian *decoder* melakukan proses *up-sampling* untuk merekonstruksi peta segmentasi dengan resolusi tinggi. Keunggulan utama U-Net terletak pada penggunaan *skip connection* yang menghubungkan *layer encoder* dan *decoder*, sehingga memungkinkan jaringan mempertahankan detail penting yang sering hilang selama proses *encoding* (Zhou dkk., 2018) serta menjaga informasi spasial citra (Zou dkk., 2025).

Selain U-Net, berbagai arsitektur *deep learning* telah dikembangkan untuk segmentasi citra, seperti *Fully Convolutional Network* (FCN) yang memungkinkan prediksi pada tingkat piksel (Long dkk., 2015), SegNet dengan skema *encoder-decoder* dan *upsampling* untuk menjaga informasi spasial (Badrinarayanan dkk., 2017), DeepLab yang mengintegrasikan *atrous convolution* dan *Conditional Random Field* (CRF) guna meningkatkan akurasi pada objek dengan variasi ukuran (Chen dkk., 2018), serta Mask R-CNN yang menghasilkan segmentasi *instance* yang lebih presisi (He dkk., 2020). Meskipun beragam pendekatan tersebut menunjukkan luasnya eksplorasi *deep learning* dalam segmentasi citra satelit, U-Net tetap menonjol karena struktur yang sederhana, kinerja yang efektif pada data terbatas, dan fleksibilitasnya untuk dikembangkan lebih lanjut. Namun, U-Net juga memiliki keterbatasan, terutama dalam mengenali fitur penting pada objek dengan bentuk kompleks dan ukuran yang beragam, serta pada area urban yang padat (Zou dkk., 2025). Struktur *skip connection* pada U-Net hanya menghubungkan *layer encoder* dan *decoder* pada level resolusi yang sama, sehingga transfer informasi fitur masih bersifat langsung tanpa proses penyaringan ulang yang adaptif. Hal ini dapat menyebabkan *semantic gap* antara fitur level rendah dan fitur level tinggi, yang berdampak pada kurang optimalnya representasi batas objek dan detail geometrik. Selain itu, U-Net cenderung mengalami kesulitan dalam membedakan objek yang berdekatan dengan tekstur serupa, sensitif terhadap bayangan, dan noise.

Berbagai pengembangan arsitektur telah diusulkan untuk mengatasi keterbatasan tersebut. U-Net++ yang diperkenalkan oleh Zhou dkk. (2018) membentuk desain *skip connection* bertingkat (*nested skip pathways*) untuk mengurangi *semantic gap* antara *encoder* dan *decoder* melalui serangkaian *convolution block* tambahan, sehingga meningkatkan akurasi segmentasi dan stabilitas pelatihan. Pengembangan lainnya berupa U-Net v2 oleh Peng dkk. (2025) melakukan peninjauan ulang terhadap mekanisme *skip connection* dan strategi agregasi fitur guna memperbaiki aliran informasi serta efisiensi parameter. Selain itu, beberapa varian seperti CM-UNet oleh Liu dkk. (2024) dan model berbasis *attention* lainnya mengintegrasikan mekanisme *atensi* atau modul kontekstual untuk memperkuat representasi fitur

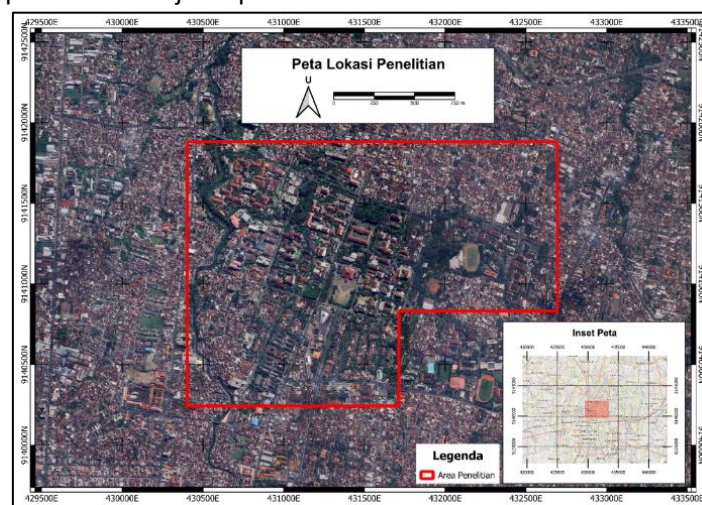
spasial dan kanal secara adaptif. Meskipun demikian, pengembangan arsitektur yang terlalu kompleks sering kali meningkatkan jumlah parameter dan kebutuhan komputasi secara signifikan, sehingga kurang efisien untuk implementasi pada skala wilayah luas dengan sumber daya terbatas. Oleh karena itu, diperlukan pendekatan yang mampu meningkatkan kemampuan seleksi fitur tanpa menambah kompleksitas model secara berlebihan.

Salah satu pendekatan yang menjanjikan adalah integrasi *Convolutional Block Attention Module* (CBAM) ke dalam arsitektur U-Net. Pemilihan CBAM didasarkan pada kemampuannya dalam mengatasi kelemahan konvolusi standar yang memberikan bobot merata pada seluruh fitur tanpa membedakan tingkat urgensi setiap fiturnya. CBAM menerapkan mekanisme *attention* pada level saluran (*channel attention*) dan spasial (*spatial attention*), sehingga jaringan mampu menekankan fitur yang relevan dan menekan informasi yang kurang penting (Woo dkk., 2018). *Channel attention* akan membantu jaringan untuk meningkatkan bobot *channel* (kanal), sehingga fitur yang sesuai dengan target akan diperkuat. *Spatial attention* akan menekankan lokasi yang relevan dan mengurangi urgensi lokasi yang kurang relevan.

CBAM dipilih karena mempertimbangkan beberapa hal. Pertama, citra satelit resolusi tinggi memiliki kompleksitas spektral dan spasial yang tinggi, sehingga diperlukan mekanisme seleksi fitur yang adaptif. Kedua, CBAM bersifat modular dan ringan sehingga tidak meningkatkan beratnya proses komputasi secara drastis. Selain itu, kombinasi antara *spatial* dan *channel attention* dapat secara relevan untuk segmentasi atap bangunan karena dapat mempertimbangkan tekstur, bentuk, dan unsur spasial di dalamnya. Integrasi CBAM ke dalam U-Net diharapkan dapat meningkatkan akurasi segmentasi atap bangunan dari citra satelit resolusi tinggi, khususnya dalam kondisi kompleks seperti keberadaan bayangan, vegetasi, dan variasi bentuk bangunan. Dengan demikian, pendekatan CBAM-UNet berpotensi menghasilkan segmentasi atap bangunan yang lebih presisi dan andal untuk mendukung pemutakhiran peta dasar secara efektif

## 2. METODE PENELITIAN

Penelitian ini dilaksanakan di kawasan Universitas Gadjah Mada, Sleman, Yogyakarta, dengan total luas sebesar 317,297 hektar. Secara geografis, area ini terletak pada koordinat  $110^{\circ}22'4,634''$  hingga  $110^{\circ}23'22,564''$  BT dan  $7^{\circ}45'45,963''$  hingga  $7^{\circ}46'39,186''$  LS. Pemilihan lokasi didasarkan pada keberagaman bentuk dan tingkat kepadatan bangunan. Keberagaman bentuk yang dimaksud mencakup variasi warna, ukuran, dan bentuk (*shape*) dari atap bangunan. Di samping itu, area ini mencakup wilayah dengan kepadatan bangunan rendah, sedang, hingga tinggi. Keberagaman ini bertujuan untuk menguji ketangguhan dan efektivitas model CBAM-UNet dalam melakukan segmentasi pada lingkungan yang kompleks. Peta lokasi penelitian disajikan pada Gambar 1.



Gambar 1. Lokasi Penelitian (Sumber: Citra Google Satellite, 2025)

## 2.1 Alat & Bahan

Alat yang digunakan terdiri atas perangkat keras dan lunak. Perangkat keras berupa komputer dengan spesifikasi CPU Intel(R) Core(TM) i5-10500H, RAM 16 GB, dengan sistem operasi Windows 11 64 bit. Adapun perangkat lunak yang digunakan terdiri atas QGIS 3.30.1 dan Jupyter Notebook untuk pemrosesan CBAM-UNet.

Bahan yang digunakan dalam penelitian ini citra satelit resolusi tinggi dari Badan Informasi Geospasial (BIG). Citra satelit resolusi tinggi yang digunakan mencakup wilayah Daerah Istimewa Yogyakarta dan sebagian wilayah di Jawa Tengah. Citra yang digunakan merupakan hasil penggabungan citra Pleiades, GeoEye, WorldView-2, dan WorldView-3 dengan rentang akuisisi data pada tahun 2018 s.d. 2021. Citra satelit resolusi tinggi yang digunakan memiliki resolusi spasial sebesar 50 cm.

## 2.2 Metode

### 2.2.1. Pre-Processing Citra

Data citra satelit resolusi tinggi yang digunakan dalam penelitian ini diperoleh dari Badan Informasi Geospasial dan telah berada pada level *orthorectified*. Citra pada area penelitian terdiri atas empat *scene* terpisah sehingga dilakukan proses *mosaicking* untuk memperoleh citra yang utuh. Selanjutnya, dilakukan *clipping* untuk membatasi area penelitian dan menghilangkan data yang tidak relevan guna meningkatkan efisiensi komputasi. Citra hasil *mosaicking* dan *clipping* memiliki format GeoTIFF dengan tipe data *UInt16*, sementara pemrosesan menggunakan arsitektur CBAM-UNet memerlukan data bertipe *byte*. Oleh karena itu, dilakukan konversi tipe data dari *UInt16* ke *byte*. Tahapan *pre-processing* ini bertujuan untuk menjamin kesesuaian format dan struktur data, meningkatkan efisiensi pembacaan data oleh model, sehingga citra siap digunakan pada tahap pelatihan model.

### 2.2.2. Pembentukan Ground Truth

*Ground truth* dibentuk sebagai data referensi untuk memberikan label kelas bangunan dan non-bangunan pada citra satelit. Data ini digunakan baik dalam proses *training model* CBAM-UNet maupun dalam evaluasi hasil segmentasi. Pembentukan *ground truth* dilakukan melalui interpretasi visual digitasi *on-screen* untuk menghasilkan data referensi yang akurat dalam bentuk *shapefile* bangunan. Proses tersebut dilakukan dengan cara mengidentifikasi atap bangunan berdasarkan unsur-unsur interpretasi citra, yaitu berdasarkan warna, rona, bentuk, ukuran, dan tekstur. Selanjutnya, *shapefile* tersebut dikonversi menjadi data raster biner, dengan nilai 1 merepresentasikan piksel bangunan dan nilai 0 merepresentasikan piksel non-bangunan.

### 2.2.3. Pembentukan Dataset

Seluruh pemrosesan menggunakan arsitektur CBAM-UNet dilakukan berbasis *tile*, sehingga data citra satelit dan *ground truth* dipecah menjadi *tile* berukuran  $128 \times 128$  piksel yang merepresentasikan area  $64 \times 64$  meter di lapangan. Pendekatan ini bertujuan untuk memperkecil ukuran data masukan dan mempercepat proses komputasi. Setiap *tile* merupakan pasangan antara citra satelit dan *ground truth*. Dataset kemudian dibagi menjadi data *training* dan data *testing*, di mana data *testing* dipilih dari lokasi yang berbeda dan belum pernah dilihat oleh model untuk menjamin kemampuan generalisasi serta mencegah *overfitting*. Pada data *training*, dilakukan augmentasi dengan penambahan *overlap* dan *sidelap* sebesar 50% untuk memperkaya variasi data. Total *tiles* yang dihasilkan pada *training dataset* adalah 2414 *tiles* setelah dilakukan augmentasi data. Di samping itu, total *testing dataset* ada di angka 80 *tiles* karena dilakukan tanpa augmentasi. *Training dataset* dirancang mencakup variasi bentuk, warna, dan kerapatan bangunan, sedangkan *testing dataset* dipilih untuk merepresentasikan keberagaman kondisi citra guna mengevaluasi performa model pada kondisi yang berbeda-beda.

### 2.2.4. Training Model

Proses *training model* dilakukan untuk membentuk model CBAM-UNet yang digunakan dalam segmentasi atap bangunan, dengan memanfaatkan library TensorFlow dan Keras. Data citra dan *ground truth* dikonversi ke dalam format numpy array agar kompatibel dengan TensorFlow dan meningkatkan

efisiensi pemrosesan. Arsitektur CBAM-UNet dibangun dengan mengintegrasikan *Convolutional Block Attention Module* (CBAM) ke dalam struktur *encoder-decoder* U-Net untuk meningkatkan kemampuan ekstraksi fitur melalui mekanisme *channel attention* dan *spatial attention*. Modul CBAM pada penelitian ini terdiri atas dua tahapan berurutan:

1. *Channel Attention*: Mengekstraksi fitur penting antar *channel* (saluran) menggunakan *Global Average Pooling* dan *Global Max Pooling*.
2. *Spatial Attention*: Memfokuskan model pada lokasi spasial (atap bangunan) yang sesuai, melalui operasi konvolusi dengan kernel  $7 \times 7$ .

Untuk menjaga stabilitas gradien dan mencegah *overfitting*, model dilengkapi dengan lapisan *Batch Normalization* dan *Dropout*. Model dikompilasi menggunakan *loss function mean squared error* (MSE) dan *optimizer Adam*. Proses *training* dilakukan menggunakan data *training* dan divalidasi dengan data *testing* selama maksimum 100 *epoch* dengan *batch size* 50, serta menerapkan *callback ModelCheckpoint* dan *EarlyStopping* untuk menyimpan bobot terbaik dan mencegah *overfitting*. Parameter lain yang digunakan yaitu *loss function* 'mean square error' dan *optimizer* 'Adam'. Model terbaik yang dihasilkan kemudian digunakan untuk melakukan segmentasi atap bangunan pada data uji. Model terbaik pada penelitian ini diambil pada data dengan nilai *validation loss* terkecil.

### 2.2.5. Segmentasi Citra

Segmentasi atap bangunan dilakukan menggunakan model CBAM-UNet terbaik hasil proses *training* dan optimalisasi untuk memprediksi kelas setiap piksel pada dataset *testing* ke dalam kelas bangunan dan non-bangunan. Proses segmentasi menghasilkan peta prediksi (format *numpy*) yang kemudian dikonversi menjadi raster biner, dengan nilai 1 merepresentasikan bangunan dan nilai 0 merepresentasikan non-bangunan. Hasil segmentasi berupa raster *tile* selanjutnya digabungkan melalui proses *mosaicking* untuk memperoleh peta segmentasi atap bangunan yang utuh dengan mempertahankan sistem koordinat dan referensi spasial citra *input*. Raster hasil segmentasi kemudian disimpan dalam format GeoTIFF.

### 2.2.6. Evaluasi Hasil Segmentasi

Evaluasi hasil segmentasi dilakukan dengan membandingkan hasil prediksi model CBAM-UNet terhadap data *ground truth* menggunakan matriks konfusi. Setiap piksel diklasifikasikan ke dalam kategori *true positive*, *true negative*, *false positive*, dan *false negative*. Berdasarkan nilai tersebut, dihitung metrik evaluasi berupa *Intersection over Union* (IoU) dan *F1-score* untuk menilai kualitas segmentasi. Evaluasi dilakukan pada seluruh *testing data*, kemudian dirata-ratakan untuk memperoleh performa keseluruhan model. Persamaan perhitungan metrik IoU dan *F1-Score* disajikan pada persamaan 1 dan 2.

$$IoU = \frac{TP}{(TP + FP + FN)} \dots\dots\dots (1)$$

$$F1-Score = \frac{2 \times TP}{2 \times TP + FP + FN} \dots\dots\dots (2)$$

Dalam hal ini:

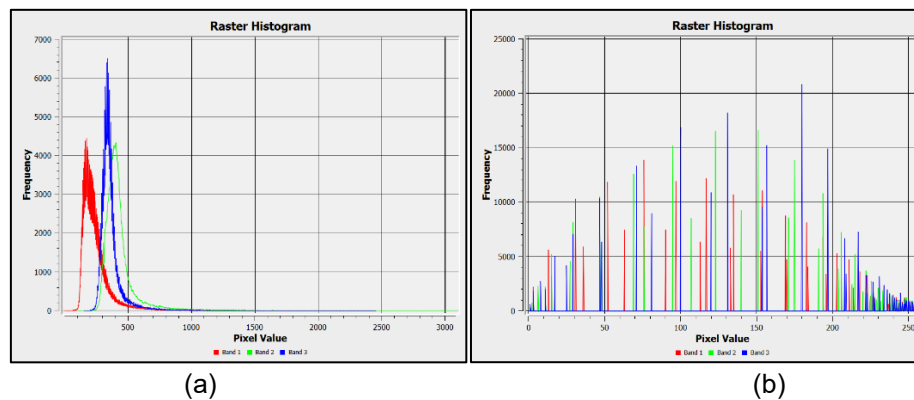
TP : True Positive  
TN : True Negative  
FP : False Positive  
FN : False Negative

## 3. HASIL DAN PEMBAHASAN

### 3.1. Pre-Processing Citra

Citra satelit resolusi tinggi yang digunakan dalam penelitian ini merupakan data dari Badan Informasi Geospasial (BIG) dengan resolusi spasial 50 cm dan telah berstatus *orthorectified imagery*, sehingga tidak memerlukan proses ortorektifikasi tambahan. Citra tersebut merupakan hasil komposit dari beberapa sensor satelit, yaitu GeoEye, WorldView-2, WorldView-3, dan Pleiades. Sebelum digunakan sebagai *input* pada model CBAM-UNet, dilakukan tahapan *pre-processing* yang meliputi *mosaicking*, *cropping*, dan konversi tipe data. Proses *mosaicking* dilakukan untuk menggabungkan empat citra terpisah

menjadi satu citra utuh, kemudian dilanjutkan dengan *cropping* guna membatasi citra pada area penelitian sehingga meningkatkan efisiensi komputasi. Selanjutnya, citra dikonversi dari format *UInt16* ke *Byte* agar sesuai dengan kebutuhan *input* CBAM-UNet, sekaligus menstabilkan proses pelatihan model melalui penskalaan nilai piksel dari rentang 0–3000 menjadi 0–255 yang disajikan pada Gambar 2.



Gambar 2. Histogram Citra (a). Sebelum Konversi (b). Setelah Konversi

### 3.2. Pembentukan Ground Truth

*Ground truth* merupakan nilai acuan/referensi yang dianggap sebagai nilai yang benar. Data *ground truth* dibentuk sesuai kondisi sebenarnya sehingga data acuan tersebut dapat dinilai valid. Fungsi *ground truth* ini adalah untuk melatih model CBAM-UNet (*training model*) guna membedakan objek bangunan terhadap objek non-bangunan. Selain itu, data *ground truth* juga digunakan untuk menilai akurasi model segmentasi dari arsitektur CBAM-UNet. Pembentukan *ground truth* pada penelitian ini dilakukan dengan cara on screen digitation. Proses digitasi dilakukan untuk menandai atap bangunan sebagai sebuah shapefile.

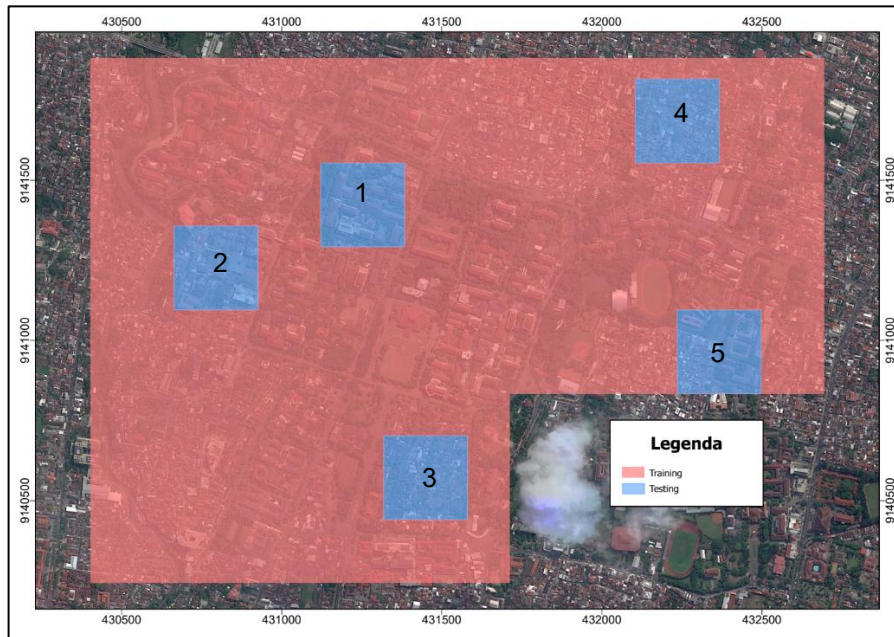
Setelah proses digitasi selesai, hasil digitasi atap bangunan disimpan dalam format *shapefile*, yang berisi geometri poligon yang menggambarkan *footprint* setiap atap bangunan. Representasi vektor ini memberikan fleksibilitas dalam pengeditan dan validasi, tetapi belum dapat digunakan secara langsung sebagai *input* bagi proses training model segmentasi berbasis piksel. Tahap selanjutnya adalah konversi *shapefile* atap bangunan tersebut menjadi format raster agar kompatibel dengan struktur data yang dibutuhkan oleh model CBAM-UNet. Proses rasterisasi mengubah poligon atap bangunan menjadi piksel raster, di mana setiap piksel diberi nilai yang merepresentasikan kelas bangunan atau non-bangunan. Konversi ini dilakukan dengan resolusi spasial yang konsisten dengan citra satelit input untuk menjaga kesesuaian antar dataset. Dengan demikian, raster *ground truth* yang dihasilkan dapat berfungsi secara optimal sebagai label pelatihan, memastikan bahwa model CBAM-UNet mampu mempelajari hubungan spasial dan tekstural secara akurat dalam proses segmentasi atap bangunan. Hasil akhir pembentukan *ground truth*, baik itu dalam format *shapefile* atau raster disajikan pada Gambar 3.



Gambar 3. Data Ground Truth (a) Shapefile (b) Raster

### 3.3. Pembentukan Dataset

Penelitian ini menggunakan *training* dan *testing dataset* dalam pengembangan model CBAM-UNet. *Training dataset* digunakan untuk melatih model, sedangkan *testing dataset* berfungsi untuk mengevaluasi performa model secara objektif dan menilai kemampuan generalisasi terhadap data yang belum pernah dilihat sebelumnya. Pemilihan lokasi *training dataset* difokuskan pada area dengan variasi dan kepadatan bangunan yang tinggi (variasi bentuk, warna, kerapatan, dan ukuran atap bangunan) agar model mampu mempelajari beragam karakteristik objek bangunan. Evaluasi model dilakukan menggunakan lima *region of interest* (ROI) pada *testing dataset* yang memiliki perbedaan bentuk, warna, dan kepadatan bangunan, sehingga performa model dapat dianalisis pada berbagai kondisi lingkungan. Pembagian *dataset* disajikan pada Gambar 4 dan deskripsi variasi *testing dataset* disajikan pada Tabel 1.



Gambar 4. Pembagian *Dataset* (Sumber: BIG, 2025)

Tabel 1. Pembagian ROI *Testing Data* (Sumber: BIG, 2025)

| ROI | Bentuk Bangunan | Kepadatan Bangunan | Tampilan |
|-----|-----------------|--------------------|----------|
| 1   | Rapi            | Rendah             |          |
| 2   | Rapi            | Sedang             |          |

| ROI | Bentuk Bangunan | Kepadatan Bangunan         | Tampilan                                                                             |
|-----|-----------------|----------------------------|--------------------------------------------------------------------------------------|
| 3   | Berpola         | Tinggi                     |   |
| 4   | Tidak Beraturan | Tinggi                     |   |
| 5   | Beragam         | Campuran Rendah dan Tinggi |  |

Setelah pembagian *dataset*, dilakukan proses *tiling* dengan ukuran  $128 \times 128$  piksel pada citra satelit dan data *ground truth*. Pada *training dataset*, *tiling* dilakukan dengan *overlap* dan *sidelap* sebesar 50% sebagai bentuk augmentasi data untuk meningkatkan keragaman data, mencegah *overfitting*, dan memperkuat kemampuan generalisasi model, sehingga dihasilkan 2414 pasangan *tile*. Sementara itu, *tiling* pada *testing dataset* dilakukan tanpa *overlap* dan *sidelap*, dengan total 80 pasangan *tile* yang digunakan dalam proses evaluasi model.

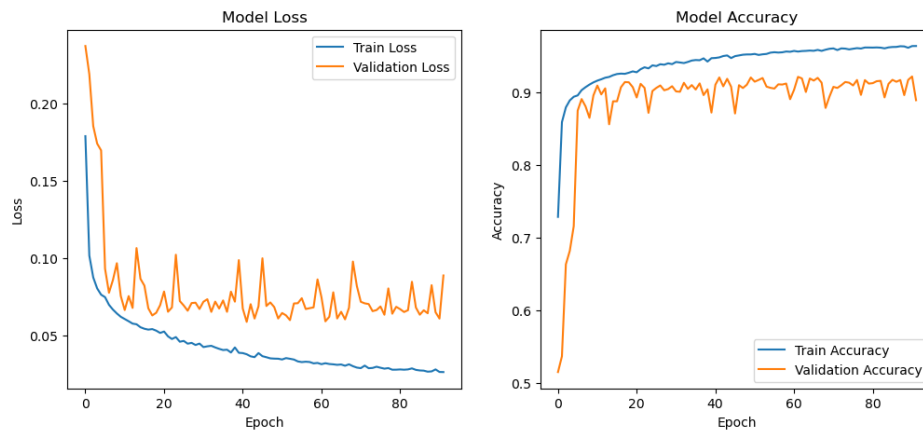
### 3.4. Training Model

Model yang dikembangkan pada penelitian ini merupakan varian U-Net yang diintegrasikan dengan *Convolutional Block Attention Module* (CBAM) untuk meningkatkan kemampuan ekstraksi fitur pada citra satelit resolusi tinggi. Model menerima input citra RGB berukuran  $128 \times 128$  piksel dan menggunakan arsitektur *encoder-decoder* dengan *skip connection* guna mempertahankan informasi spasial. Jalur *encoder* terdiri atas blok konvolusi, normalisasi, dan *dropout* untuk menekan risiko *overfitting*, sementara jalur *decoder* berfungsi merekonstruksi resolusi spasial citra secara bertahap.

Integrasi CBAM diterapkan pada *skip connection* dan tahap penggabungan fitur di decoder untuk memfokuskan perhatian model secara adaptif. Mekanisme *channel attention* digunakan untuk mempelajari bobot kepentingan setiap kanal spektral melalui *global average pooling* dan *global max pooling*, sedangkan *spatial attention* menekankan lokasi fitur penting melalui agregasi statistik spasial dan konvolusi berukuran  $7 \times 7$ . Kombinasi kedua mekanisme ini memungkinkan model menyoroti atap bangunan secara lebih presisi serta mengurangi perhatian pada area non-bangunan, sehingga meningkatkan kualitas segmentasi pada objek dengan variasi bentuk dan tekstur.

Pelatihan model dilakukan menggunakan *loss function* mean squared error dan optimizer Adam dengan dukungan *Early Stopping* dan *Model Checkpoint*. Model dilatih selama 100 epoch dengan ukuran *batch* 50 menggunakan 2414 pasangan *tile* data *training* dan 80 pasangan *tile* data *testing*. Hasil pelatihan menunjukkan konvergensi yang stabil dengan performa validasi yang tinggi. Dibandingkan dengan Basic U-Net, CBAM-UNet menghasilkan *validation loss* yang lebih rendah (0,0589) dan *validation accuracy* yang

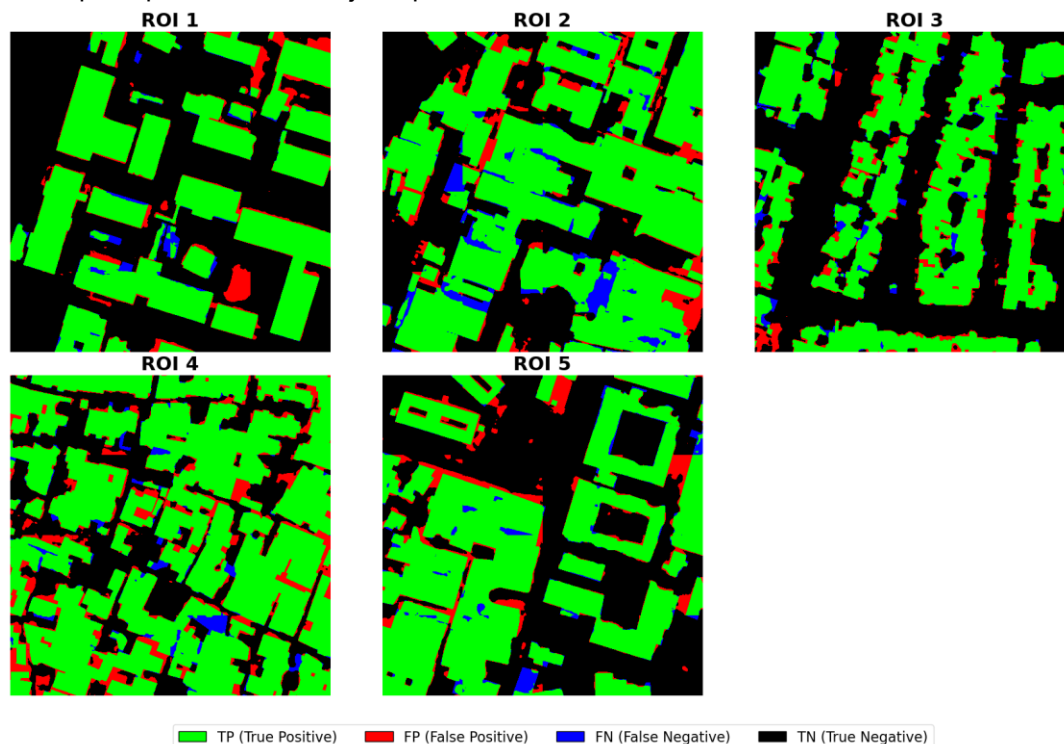
lebih tinggi (0,9208), yang menegaskan peningkatan kinerja dan kemampuan generalisasi dalam segmentasi atap bangunan. Grafik hasil *training model* disajikan pada Gambar 5.



**Gambar 5.** Hasil *Training Model* CBAM-UNet

### 3.5. Segmentasi Atap Bangunan

Proses segmentasi atap bangunan dilakukan secara otomatis menggunakan model yang dihasilkan pada *training data*. Segmentasi atap bangunan menggunakan model CBAM-UNet dilakukan untuk mengidentifikasi dan memisahkan piksel yang merepresentasikan atap bangunan dan non-bangunan pada citra satelit resolusi tinggi. Dengan kombinasi arsitektur CBAM-UNet, model mampu menonjolkan fitur-fitur penting baik secara *channel* maupun *spatial*, sehingga menghasilkan segmentasi yang lebih presisi pada area bangunan yang memiliki tekstur kompleks atau ukuran bervariasi. Hasil segmentasi pada penelitian ini disajikan pada Gambar 6.



**Gambar 6.** Hasil Segmentasi CBAM-UNet

### 3.6. Evaluasi Hasil Segmentasi

Hasil segmentasi atap bangunan menggunakan arsitektur CBAM-UNet telah mampu memisahkan objek bangunan dengan non-bangunan secara *general*. Namun demikian, masih terdapat

kesalahan di dalamnya. Oleh karena itu, perlu dilakukan uji akurasi untuk menilai seberapa baik model dalam melakukan segmentasi atap bangunan. Proses segmentasi ini dilakukan menggunakan nilai komponen pengujian akurasi, yaitu *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). Informasi hasil komponen uji akurasi tiap ROI disajikan pada Tabel 2 berikut.

**Tabel 2.** Komponen Uji Akurasi

| Arsitektur | ROI   | Jumlah Komponen (Piksel) |         |        |        |
|------------|-------|--------------------------|---------|--------|--------|
|            |       | TP                       | TN      | FP     | FN     |
| CBAM-UNet  | ROI 1 | 96.699                   | 150.398 | 11.785 | 3.262  |
|            | ROI 2 | 146.116                  | 87.911  | 16.790 | 11.327 |
|            | ROI 3 | 109.552                  | 135.810 | 12.442 | 4.340  |
|            | ROI 4 | 160.728                  | 71.188  | 24.513 | 5.715  |
|            | ROI 5 | 118.638                  | 123.558 | 14.665 | 5.283  |
| Basic UNet | ROI 1 | 99.383                   | 139.563 | 22.620 | 578    |
|            | ROI 2 | 154.551                  | 70.739  | 33.962 | 2.892  |
|            | ROI 3 | 112.333                  | 126.621 | 21.631 | 1.559  |
|            | ROI 4 | 165.136                  | 53.486  | 42.215 | 1.307  |
|            | ROI 5 | 120.786                  | 112.642 | 25.581 | 3.135  |

Lebih lanjut, data dari seluruh komponen tersebut digunakan sebagai bahan perhitungan akurasi. Akurasi dalam penelitian ini dinilai dari indeks *IoU* dan *F1-Score*. *IoU* dan *F1-Score* ini dihitung berdasarkan persamaan 1 dan 2. Hasil perhitungan akurasi hasil segmentasi disajikan di Tabel 3 berikut.

**Tabel 3.** Hasil Uji Akurasi

| Arsitektur | ROI              | Lokasi |        |        |        |        |
|------------|------------------|--------|--------|--------|--------|--------|
|            |                  | ROI 1  | ROI 2  | ROI 3  | ROI 4  | ROI 5  |
| CBAM-UNet  | Precision        | 0,8914 | 0,8969 | 0,8980 | 0,8677 | 0,8900 |
|            | Recall           | 0,9674 | 0,9281 | 0,9619 | 0,9657 | 0,9574 |
|            | IoU              | 0,8653 | 0,8386 | 0,8672 | 0,8417 | 0,8561 |
|            | F1-Score         | 0,9278 | 0,9122 | 0,9289 | 0,9140 | 0,9224 |
|            | Overall Accuracy | 0,9426 | 0,8927 | 0,9360 | 0,8847 | 0,9239 |
|            | Mean IoU         | 0,8538 |        |        |        |        |
|            | Mean F1-Score    | 0,9211 |        |        |        |        |
| Basic UNet | Precision        | 0,8146 | 0,8198 | 0,8385 | 0,7964 | 0,8252 |
|            | Recall           | 0,9942 | 0,9816 | 0,9863 | 0,9921 | 0,9747 |
|            | IoU              | 0,8108 | 0,8075 | 0,8289 | 0,7914 | 0,8079 |
|            | F1-Score         | 0,8955 | 0,8935 | 0,9064 | 0,8836 | 0,8938 |
|            | Overall Accuracy | 0,9115 | 0,8594 | 0,9115 | 0,8340 | 0,8905 |
|            | Mean IoU         | 0,8093 |        |        |        |        |
|            | Mean F1-Score    | 0,8945 |        |        |        |        |

Tabel 3 menunjukkan bahwa model CBAM-UNet memberikan kinerja segmentasi atap bangunan yang lebih unggul dibandingkan Basic UNet pada lima area uji (ROI). Hal ini terlihat dari nilai *Intersection over Union* (IoU) dan *F1-Score* yang secara konsisten lebih tinggi pada seluruh lokasi ketika menggunakan CBAM-UNet. Rata-rata IoU untuk CBAM-UNet mencapai 0.8538, lebih tinggi dibandingkan Basic UNet yang hanya 0.8093, menandakan bahwa model dengan integrasi *Convolutional Block Attention Module* (CBAM) mampu menghasilkan hasil segmentasi yang lebih akurat terhadap *ground truth*. Demikian pula, rata-rata *F1-Score* CBAM-UNet mencapai 0.9211, melampaui Basic UNet dengan nilai 0.8945, yang menunjukkan peningkatan ketepatan dalam mendeteksi dan melabeli piksel atap bangunan. Secara keseluruhan, integrasi modul CBAM terbukti meningkatkan kemampuan model dalam mengenali struktur atap bangunan pada berbagai kondisi yang berbeda, sehingga menghasilkan performa segmentasi yang lebih stabil dan presisi di seluruh ROI.

#### 4. KESIMPULAN

Hasil dari penelitian ini menunjukkan bahwa optimisasi segmentasi atap bangunan dari citra satelit resolusi tinggi menggunakan arsitektur CBAM-UNet dapat diterapkan dengan efisien. Model CBAM-UNet untuk segmentasi atap bangunan pada citra satelit resolusi tinggi dapat terbentuk dengan optimal. Model CBAM-UNet yang terbentuk mendapatkan nilai *validation loss* sebesar 0,0589. Nilai tersebut lebih baik dibandingkan model *basic UNet* yang mendapatkan *validation loss* sebesar 0,0627. Model CBAM-UNet yang terbentuk terbukti mampu digunakan untuk segmentasi atap bangunan dengan optimal. Objek bangunan dan non-bangunan dapat terpisahkan secara otomatis menggunakan model yang dihasilkan. CBAM-UNet menghasilkan akurasi yang baik dalam proses segmentasi atap bangunan dari citra satelit resolusi tinggi..

#### Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Badan Informasi Geospasial (BIG) atas dukungan dan penyediaan data citra satelit resolusi tinggi yang menjadi bahan utama dalam pelaksanaan penelitian ini. Penulis juga menyampaikan apresiasi kepada Program Studi Magister Teknik Geomatika Universitas Gadjah Mada (UGM) atas dukungan dan fasilitas, sehingga penelitian ini dapat diselesaikan dengan baik..

#### Daftar Pustaka

- Apostolopoulos, D. N., & Nikolakopoulos, K. G. (2020). Assessment and quantification of the accuracy of low-and high-resolution remote sensing data for shoreline monitoring. *ISPRS International Journal of Geo-Information*, 9(6). <https://doi.org/10.3390/ijgi9060391>
- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- Brahmantara, R. P., Ulfa, K., Rahayu, M. I., Prabowo, Y., Hestrio, Y. F., Sartika, Kustiyo, Candra, D. S., & Novresiandi, D. A. (2022). Cloud-free mosaic for high-resolution satellite to support natural resources and disaster monitoring. *IOP Conference Series: Earth and Environmental Science*, 1109(1). <https://doi.org/10.1088/1755-1315/1109/1/012053>
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)*, 801–818.
- Hakim, N. I. A., Sabri, L. M., & Sukmono, A. (2019). Kajian Akurasi Citra Satelit Worldview 4 Pada Pembuatan Peta Dasar Pendaftaran Tanah. *Jurnal Geodesi Undip*, 8(1), 308–317.
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 386–397. <https://doi.org/10.1109/TPAMI.2018.2844175>
- Kristal, A., & Harintaka, H. (2022). Analisis Kehandalan Ekstraksi Garis Tepi Bangunan dari Data Foto Udara Menggunakan Pendekatan Deep Learning Berbasis Mask R-CNN. *Journal of Geodesy and Geomatics*, 17(2), 273. <https://doi.org/10.12962/j24423998.v17i2.11401>
- Li, M., Stein, A., & de Beurs, K. M. (2020). A Bayesian characterization of urban land use configurations from VHR remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, 92(June), 102175. <https://doi.org/10.1016/j.jag.2020.102175>
- Liu, M., Dan, J., Lu, Z., Yu, Y., Li, Y., & Li, X. (2024). CM-UNet: Hybrid CNN-Mamba UNet for Remote Sensing Image Semantic Segmentation. *Computer Vision and Pattern Recognition*, 14(8), 1–5. <http://arxiv.org/abs/2405.10530>
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440. <https://doi.org/10.1109/ICCVW.2019.00113>
- Peng, Y., Chen, D. Z., & Sonka, M. (2025). U-Net V2: Rethinking the Skip Connections of U-Net for Medical Image Segmentation. *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*, 1–5.

- <https://doi.org/10.1109/ISBI60581.2025.10980742>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Computer Vision and Pattern Recognition*, 9, 16591–16603. <https://doi.org/https://doi.org/10.48550/arXiv.1505.04597>
- Seidlova, A., Kudelcikova, M., Mihalik, J., & Rekus, D. (2021). Interpretation of Remote Sensing Imagery. *IOP Conference Series: Earth and Environmental Science*, 906(1). <https://doi.org/10.1088/1755-1315/906/1/012070>
- Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. *IEEE Access*, 9, 82031–82057. <https://doi.org/10.1109/ACCESS.2021.3086020>
- Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision (ECCV)*, 33–19. <https://doi.org/10.1007/978-3-030-01234-2>
- Wu, H., Liu, Q., & Liu, X. (2019). A Review on Deep Learning Approaches to Image Classification and Object Segmentation. *Computers, Materials and Continua*, 60(2), 575–597. <https://doi.org/10.32604/cmc.2019.03595>
- Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). UNet++: A Nested U-Net Architecture. In *Deep learning in medical image analysis and multimodal learning for clinical decision support: Vol. 11045 LNCS*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-00889-5>
- Zou, W., Jiang, Y., Liao, W., Fan, S., Yang, Y., Hou, J., & Tang, H. (2025). Improved U-Net for Precise Gauge Dial Segmentation in Substation Inspection Systems: A Study on Enhancing Accuracy and Robustness. *Information (Switzerland)*, 16(5). <https://doi.org/10.3390/info16050382>