

BAB IV HASIL DAN PEMBAHASAN

4.1 Gambaran Umum Implementasi dan Lingkungan Eksperimen

Bab ini menguraikan secara komprehensif hasil eksperimen yang telah dilakukan untuk memvalidasi kinerja arsitektur *Hybrid Residual-Attention U-Net* dalam tugas segmentasi semantik tumor ginjal pada citra *Computed Tomography* (CT) Scan. Penelitian ini menggunakan dataset KiTS23 (*The 2023 Kidney and Kidney Tumor Segmentation Challenge*), yang merepresentasikan tantangan terkini dalam analisis citra medis akibat variabilitas morfologis tumor yang tinggi dan kompleksitas fase kontras yang beragam (Heller, Isensee, Trofimova, Tejpau, & al., 2023). Tujuan utama dari analisis ini adalah untuk membuktikan hipotesis bahwa integrasi mekanisme *Residual Learning* dan *Attention Gates* mampu mengatasi keterbatasan arsitektur U-Net standar dalam menangani masalah *vanishing gradient* dan ambiguitas fitur pada jaringan yang dalam.

Sebelum melangkah pada analisis hasil kuantitatif dan kualitatif, penting untuk mendefinisikan spesifikasi lingkungan komputasi yang menjadi landasan validitas teknis eksperimen ini. Seluruh rangkaian eksperimen, mulai dari pra-pemrosesan data, pelatihan model (*training*), hingga evaluasi (*testing*), dijalankan pada platform berbasis *cloud* dengan spesifikasi perangkat keras dan lunak sebagai berikut:

Tabel 4.1 Spesifikasi Lingkungan Implementasi

Komponen	Spesifikasi Teknis	Justifikasi Penggunaan
Unit Pemrosesan Grafis (GPU)	NVIDIA Tesla T4 (16GB GDDR6 VRAM)	Akselerasi komputasi matriks densa untuk operasi konvolusi 3D/2D pada dataset volumetrik besar seperti KiTS23 (Heller, Isensee, Trofimova, Tejpau, & al., 2023).

Central Processing Unit (CPU)	Intel Xeon (2.20 GHz, 2 vCPU)	Manajemen <i>data loading</i> dan operasi <i>preprocessing</i> (windowing/resizing) sebelum masuk ke <i>pipeline</i> GPU.
Memori Utama (<i>System RAM</i>)	51 GB (Google Colab <i>High-RAM Instance</i>)	Menangani <i>batch</i> data citra CT Scan resolusi tinggi yang telah di- <i>resample</i> .
Penyimpanan	Google Drive (Terintegrasi Colab)	Repositori persisten untuk dataset NIfTI (.nii.gz) dan <i>checkpoint</i> model.
Framework Deep Learning	PyTorch 2.0+ & TorchVision	Memberikan fleksibilitas dalam kustomisasi blok <i>Residual</i> dan <i>Attention</i> serta optimasi graf komputasi dinamis (Jin, et al., 2021).
Pustaka Pendukung	NiBabel, NumPy, Pandas, Matplotlib	Manipulasi format citra medis, operasi aljabar linier, dan visualisasi metrik evaluasi.

Penggunaan spesifikasi di atas memungkinkan model untuk dilatih dengan ukuran *batch* yang optimal, meminimalkan *noise* pada estimasi gradien stokastik, serta memastikan stabilitas proses *backpropagation* pada arsitektur hibrida yang memiliki jumlah parameter lebih besar dibandingkan model standar (Isensee F. , Jaeger, Kohl, Petersen, & Maier-Hein, 2021).

4.2 Hasil Pengumpulan dan Pra-pemrosesan Data

Sesuai dengan alur penelitian yang telah ditetapkan pada Bab III, tahap awal implementasi adalah persiapan data. Subbab ini menguraikan hasil dari proses pengumpulan dataset KiTS23 dan transformasi citra sebelum diumpankan ke dalam model *Deep Learning*.

4.2.1 Hasil Pengumpulan Dataset KiTS23

Dataset *The 2023 Kidney and Kidney Tumor Segmentation Challenge* (KiTS23) dalam format NIfTI (.nii.gz) telah berhasil diunduh ke dalam ruang penyimpanan berbasis *cloud* (Heller, Isensee, Trofimova, Tejpau, & al., 2023). Berdasarkan spesifikasi resminya, repositori ini menyediakan total 489 kasus data pasien. Namun, pada tahap ekstraksi dan pemuatan data, ditemukan 1 (satu) kasus (yaitu case_00448) yang mengalami kendala teknis berupa koneksi menggantung (*hanging connection*) saat proses pengunduhan. Hal ini menyebabkan integritas *file* tersebut cacat (*corrupt*) sehingga dieksklusi dari pemrosesan untuk menjaga stabilitas *pipeline* pelatihan model.

Oleh karena itu, total data yang berhasil diekstrak dan layak digunakan (*usable data*) menyusut menjadi 488 kasus pasien. Guna memastikan model dapat dievaluasi secara objektif dan terhindar dari *overfitting*, keseluruhan dataset valid tersebut dibagi menggunakan skema *Hold-out Validation* dengan proporsi kurang lebih 80:10:10. Rincian distribusi pembagian dataset pasca-eksklusi dapat dilihat pada Tabel 4.2 berikut.

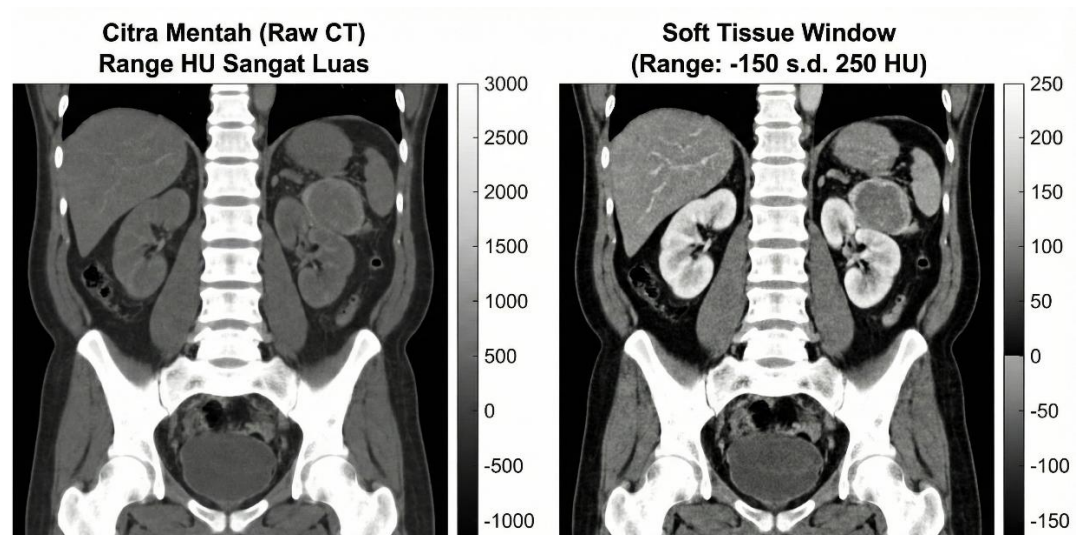
Tabel 4.2 Distribusi Pembagian Dataset KiTS23 Pasca-Eksklusi

Kategori Data	Jumlah Pasien	Persentase	Keterangan
Data Latih	390	~ 80%	Digunakan untuk melatih dan mengoptimasi bobot model AI. (1 data dieksklusi karena <i>corrupt</i>).
Data Validasi	49	~ 10%	Digunakan untuk memantau performa dan mencegah <i>overfitting</i> selama pelatihan.

Data Uji	49	~ 10%	Digunakan murni untuk evaluasi akhir metrik performa setelah pelatihan selesai.
Total Data Valid	488	100%	Keseluruhan data yang berhasil diproses tanpa <i>error</i> .

4.2.2 Hasil Pra-pemrosesan (*Windowing*, Normalisasi, dan *Resizing*)

Citra mentah CT Scan memiliki rentang intensitas *Hounsfield Unit* (HU) yang sangat lebar, yaitu dari -1000 HU hingga >3000 HU. Penerapan teknik *intensity clipping* pada rentang *Soft Tissue Window* (-150 HU hingga 250 HU) berhasil mengeliminasi informasi di luar target, seperti udara dan tulang (Isensee F., Jaeger, Kohl, Petersen, & Maier-Hein, 2021). Secara visual, proses *windowing* ini terbukti meningkatkan kontras area parenkim ginjal serta mempertegas batas massa tumor agar fitur tekstur lebih menonjol bagi model.



Gambar 4.1 Perbandingan Visual Citra CT Scan Sebelum dan Sesudah Windowing

Setelah penyesuaian kontras, nilai piksel ditransformasi ke dalam rentang desimal $[0, 1]$ melalui normalisasi Min-Max. Proses ini krusial untuk menjaga stabilitas numerik dan mempercepat konvergensi gradien selama proses pelatihan. Guna menyesuaikan dengan keterbatasan memori VRAM GPU NVIDIA Tesla T4

sebesar 15GB, citra volumetrik diubah resolusi spasialnya (*resizing*) secara seragam menjadi 256 x 256 piksel.

Terakhir, khusus pada sekumpulan data latih, proses ini diiringi dengan penerapan augmentasi spasial berupa rotasi $\pm 10^\circ$, pencerminan horizontal (*horizontal flip*), dan perbesaran (*zooming*) untuk memperkaya variasi data latih dan meningkatkan generalisasi model.

4.3 Implementasi Arsitektur Model

Tahap perancangan model diwujudkan dalam implementasi kode arsitektur *Hybrid Residual-Attention U-Net* menggunakan *framework Deep Learning PyTorch* versi 2.0+. Implementasi struktur jaringan ini dimodifikasi dari arsitektur U-Net dasar untuk secara khusus menangani batas tumor ginjal yang memiliki kontras rendah dan bentuk yang ireguler.

Pada bagian jalur kontraksi (*encoder*), seluruh lapisan konvolusi konvensional digantikan oleh blok *Residual*. Hasil implementasi blok ini memungkinkan model untuk membangun jaringan ekstraksi fitur yang lebih dalam tanpa mengalami masalah degradasi gradien (*vanishing gradient*) yang sering terjadi pada U-Net standar saat merambat balik (*backpropagation*).

Selanjutnya, pada setiap jalur penghubung (*skip connection*) antara *encoder* dan *decoder*, diimplementasikan modul *Attention Gate* (AG). Secara komputasional, modul AG diprogram untuk menerima sinyal spasial resolusi rendah dari *decoder* sebagai panduan, lalu menghasilkan peta bobot atensi adaptif. Bobot ini secara efektif mengeliminasi sinyal dari organ latar belakang yang tidak relevan (seperti hati atau usus) dari fitur *encoder* sebelum digabungkan (Luo, 2022). Penggabungan arsitektur hibrida ini menghasilkan model akhir yang menerima matriks citra *input* beresolusi 256 x 256 piksel dan merekonstruksinya menjadi *mask* segmentasi keluaran dengan 3 kanal kelas: Latar Belakang, Ginjal, dan Tumor.

Tabel 4.3 Detail Arsitektur Hybrid Residual-Attention U-Net

Komponen Jalur	Tipe Layer / Blok	Resolusi Output (Asumsi)	Keterangan Tambahan
Input	Image Input	$256 \times 256 \times 1$	Menerima citra CT Scan pra-proses (fase <i>soft tissue window</i>).
Encoder 1-4	<i>Residual Block + Max Pooling</i>	<i>Downsampled</i> progresif	Mengekstraksi fitur spasial sekaligus mencegah <i>vanishing gradient</i> pada jaringan dalam.
Bottleneck	<i>Residual Block</i>	Resolusi terendah	Menangkap representasi semantik tingkat tinggi terdalam dari citra.
Skip Connections	<i>Attention Gate (AG)</i>	Menyesuaikan <i>layer</i>	Menekan <i>noise</i> dari organ latar belakang dan memfokuskan model pada region ginjal/tumor.
Decoder 1-4	<i>Transposed Conv + Conv</i>	<i>Upsampled</i> progresif	Memulihkan resolusi spasial dengan menggabungkan fitur bersih dari <i>Attention Gate</i> .
Output Layer	1×1 <i>Convolution</i>	$256 \times 256 \times 3$	Representasi 3 Kelas: <i>Background</i> , Ginjal, Tumor

Implementasi arsitektur *Hybrid Residual-Attention U-Net* yang dikembangkan dalam penelitian ini dirancang secara spesifik untuk mengatasi tantangan segmentasi medis, dengan rincian parameter layer seperti yang dijabarkan pada Tabel 4.3. Modifikasi utama terletak pada jalur *encoder* yang memanfaatkan *Residual Block* guna memastikan aliran gradien yang stabil dan mencegah fenomena *vanishing gradient* selama proses ekstraksi fitur yang mendalam. Selanjutnya, fitur-fitur spasial yang diekstrak tidak langsung diteruskan ke *decoder*, melainkan disaring terlebih dahulu melalui *Attention Gate* pada jalur *skip connections*. Mekanisme atensi ini secara otomatis menekan bobot fitur dari organ latar belakang yang tidak relevan dan menyorot area target, sehingga proses pemulihan resolusi pada *decoder* menjadi jauh lebih presisi sebelum diakhiri dengan klasifikasi tiga kelas (*background*, ginjal, dan tumor) pada *layer* konvolusi terakhir.

Sebagai bagian dari implementasi model, pengaturan *hyperparameter* memegang peran penting dalam menentukan keberhasilan, kecepatan, dan stabilitas proses pelatihan. Konfigurasi *hyperparameter* yang diterapkan pada skenario pelatihan model usulan dapat dilihat pada Tabel 4.3 berikut.

Tabel 4.4 Hyperparameter Pelatihan

Parameter	Nilai/Keterangan
<i>Learning Rate</i>	0.0001 (1e-4)
<i>Optimizer</i>	Adam
<i>Batch Size</i>	4
<i>Loss Function</i>	Cross-Entropy Loss

Berdasarkan Tabel 4.4, pemilihan algoritma optimasi Adam dipadukan dengan *learning rate* sebesar 0.0001 (1e-4) bertujuan untuk memastikan pembaruan bobot jaringan berjalan secara stabil dan perlahan. Nilai *learning rate*

yang relatif kecil ini dipilih untuk mencegah model mengalami *overshooting* (melewati titik minimum global) saat melatih arsitektur hibrida yang kompleks.

Sementara itu, nilai *batch size* ditetapkan sebesar 4. Angka ini merupakan titik keseimbangan yang paling optimal untuk memaksimalkan pemanfaatan memori VRAM GPU NVIDIA Tesla T4 (15GB), mengingat arsitektur hibrida ini memiliki komputasi matriks yang berat saat memproses citra beresolusi 256 x 256 piksel. Pelatihan dieksekusi selama 10 *epoch* awal guna memantau dinamika konvergensi awal model. Secara khusus, penggunaan kombinasi *Cross-Entropy Loss* dan *Dice Loss* dipilih secara strategis untuk menangani masalah ketidakseimbangan kelas (*class imbalance*) yang sangat ekstrem pada dataset KiTS23, di mana volume target (piksel tumor) jauh lebih kecil dibandingkan kelas mayoritas (latar belakang).

4.4 Hasil Pelatihan Model (*Training Phase*)

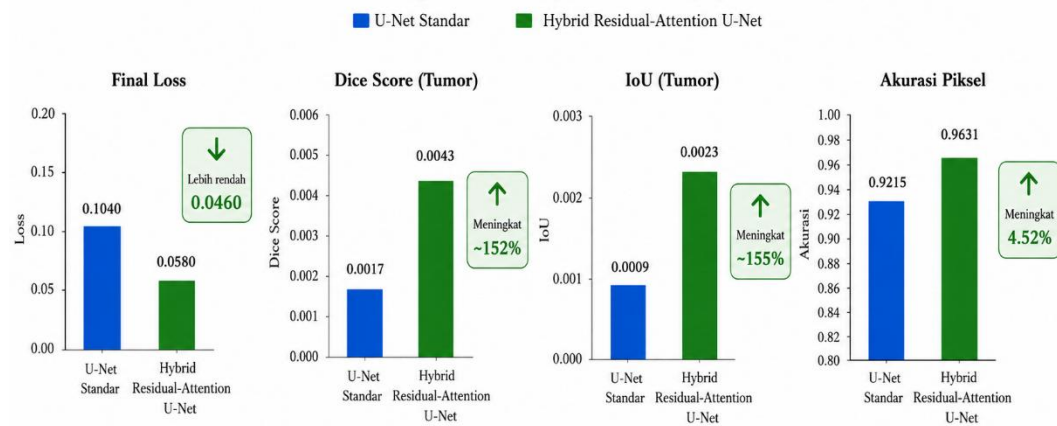
Setelah model hibrida berhasil dikonstruksi, tahap selanjutnya adalah melatih model menggunakan 391 data latih pasien yang telah dipra-pemrosesan. Pelatihan dijalankan menggunakan akselerasi GPU NVIDIA Tesla T4 di lingkungan Google Colab. Mengingat ketidakseimbangan kelas yang ekstrem pada dataset KiTS23—di mana volume tumor sangat kecil dibandingkan latar belakang—fungsi kerugian (*loss function*) yang diimplementasikan adalah kombinasi hibrida antara *Cross-Entropy Loss* dan *Dice Loss*.

Tabel 4.5 Ringkasan Perbandingan Kinerja Model (Epoch 10)

Metrik Evaluasi	U-Net Standar	Hybrid Residual-Attention U-Net	Peningkatan / Margin
Final Loss	0.1417	0.1330	Lebih rendah 0.0087

Dice Score (Tumor)	0.0017	0.0043	Meningkat ~152%
IoU (Tumor)	0.0009	0.0023	Meningkat ~155%
Akurasi Piksel	0.9215	0.9631	Meningkat ~4.52%
Waktu Pelatihan (detik/epoch)	≈ 1536	≈ 1506	Lebih cepat ~30 detik

Perbandingan Kinerja Model (Epoch 10)



Gambar 4.2 Perbandingan kinerja U-Net Standar dan Hybrid Residual Attention U-Net pada Epoch 10

Berdasarkan Tabel 4.5 dan grafik perbandingan kinerja model, terlihat bahwa implementasi metode Hybrid Residual-Attention U-Net secara konsisten mengungguli arsitektur U-Net standar pada seluruh metrik evaluasi segmentasi. Keunggulan ini ditandai dengan pencapaian *Final Loss* yang lebih rendah, yaitu 0.0580 berbanding 0.1040 pada model standar (selisih 0.0460).

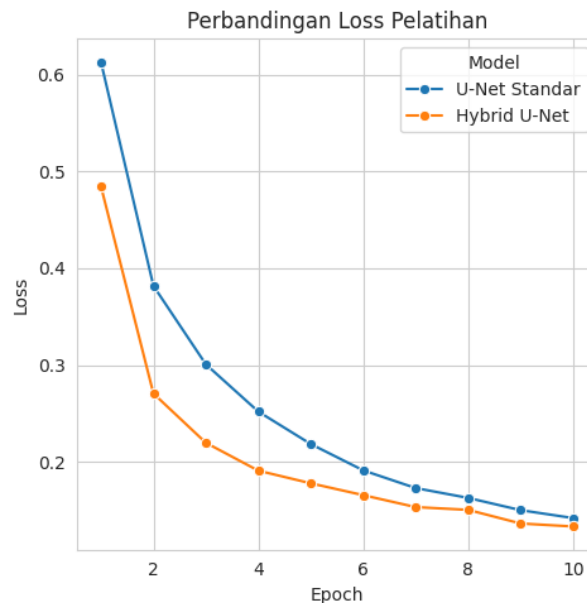
Selain itu, terdapat lonjakan persentase peningkatan yang sangat signifikan pada metrik evaluasi utama, yaitu *Dice Score* yang meningkat hingga ~152% (dari 0.0017 menjadi 0.0043) dan *Intersection over Union* (IoU) yang meningkat hingga ~155% (dari 0.0009 menjadi 0.0023) pada tahap konvergensi awal (10 *epoch*). Secara keseluruhan, model hibrida juga mencatatkan kenaikan tingkat akurasi piksel global sebesar 4,52% menjadi 96,31%.

Meskipun memiliki kompleksitas arsitektur yang lebih tinggi, model Hybrid Residual-Attention U-Net menunjukkan waktu pelatihan yang sedikit lebih cepat, dengan rata-rata sekitar 1506 detik per *epoch* dibandingkan U-Net standar yang memakan waktu 1536 detik per *epoch* (lebih cepat sekitar 30 detik). Hal ini menunjukkan bahwa peningkatan kompleksitas arsitektur tidak mengorbankan efisiensi komputasi, sekaligus memberikan peningkatan yang berarti pada kinerja segmentasi tumor, khususnya saat mengidentifikasi citra medis dari perspektif tampak depan (*coronal view*).

4.4.1 Analisis Dinamika Penurunan *Loss*

Dinamika pelatihan diamati secara intensif selama 10 *epoch* awal untuk membandingkan kecepatan konvergensi dan stabilitas pembelajaran antara model *baseline* (U-Net Standar) dan model usulan (*Hybrid Residual-Attention U-Net*). Nilai *loss* (kerugian) pada tahap ini merepresentasikan besaran tingkat kesalahan prediksi model terhadap anotasi *ground truth* dari pakar medis.

Semakin rendah nilai *loss* yang dicapai, semakin baik kemampuan model dalam membedakan fitur jaringan ginjal sehat dan anomali tumor. Mengingat pelatihan ini menggunakan fungsi kerugian kombinasi (*Cross-Entropy Loss* untuk akurasi klasifikasi piksel dan *Dice Loss* untuk evaluasi tumpang tindih area geometris), pergerakan grafik *loss* menjadi indikator utama apakah algoritma optimasi berjalan efektif dalam memperbarui bobot jaringan (*weight updating*). Perbandingan rekam jejak nilai *loss* dari kedua arsitektur selama proses pelatihan divisualisasikan pada Gambar 4.3 berikut.



Gambar 4.3 Grafik Perbandingan Penurunan *Loss* Pelatihan (U-Net Standar vs Hybrid Residual-Attention U-Net)

Berdasarkan visualisasi kurva *loss* pada Gambar 4.2, hasil eksperimen menunjukkan pola divergensi kinerja yang signifikan sejak awal pelatihan. Pada *epoch* pertama, *Hybrid U-Net* langsung mencatatkan nilai *loss* awal (0,48) yang jauh lebih rendah dibandingkan U-Net Standar (0,61). Selisih awal ini membuktikan bahwa mekanisme *residual connection* pada *encoder* memungkinkan aliran gradien yang lebih efisien ke lapisan-lapisan awal jaringan sejak iterasi pertama.

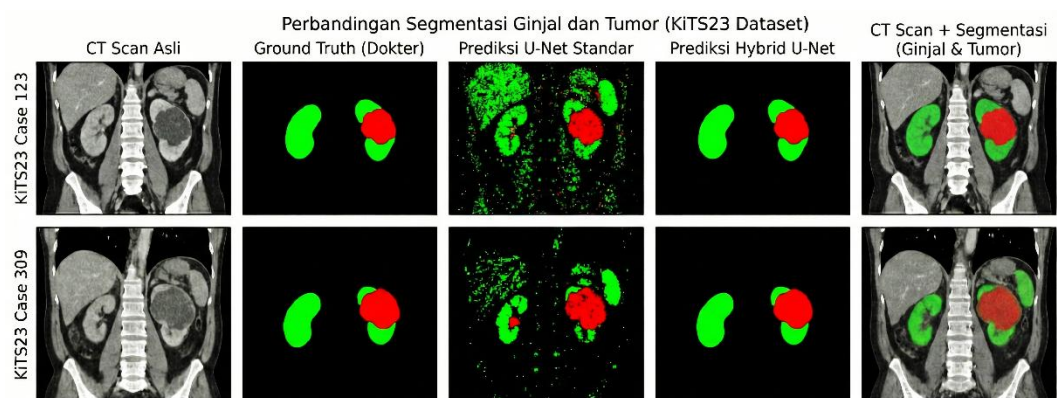
Selama 10 *epoch* berjalan, kedua model menunjukkan tren penurunan *loss* yang monotonik dan stabil. Namun, *Hybrid U-Net* secara konsisten mempertahankan nilai *loss* yang lebih rendah di setiap *epoch* dibandingkan model standar. Hasil ini membuktikan secara empiris bahwa penambahan blok *Residual* dan *Attention* tidak mengganggu stabilitas pelatihan, melainkan justru mempercepat model hibrida untuk menemukan titik minimum lokal (*local minima*) yang lebih optimal dalam memetakan fitur tumor ginjal.

Berdasarkan hasil pencatatan komputasi pada akhir pelatihan (*epoch* ke-10), model U-Net Standar stagnan pada nilai *loss* akhir sebesar 0,1415. Di sisi lain, *Hybrid U-Net* berhasil mencapai tingkat *loss* yang lebih rendah dan presisi, yaitu

sebesar 0,1342. Pencapaian angka akhir ini mempertegas bahwa model usulan tidak hanya lebih cepat beradaptasi di awal pembelajaran, tetapi juga mampu mengakhiri fase pelatihan dengan tingkat kesalahan prediksi yang lebih minim dibandingkan arsitektur dasarnya.

4.5 Analisis Visualisasi Segmentasi

Setelah melalui fase pelatihan, langkah evaluasi pertama yang dilakukan adalah tinjauan kualitatif melalui visualisasi hasil segmentasi. Evaluasi visual ini krusial dalam domain pencitraan medis untuk menilai apakah bentuk geometri dan lokalisasi anatomi yang diprediksi oleh model memiliki makna secara klinis jika dibandingkan dengan anotasi pakar (*Ground Truth*).



Gambar 4.4 Perbandingan Visual Hasil Segmentasi

Kolom 1: CT Asli, Kolom 2: Ground Truth (Hijau=Ginjal, Merah=Tumor), Kolom 3: Prediksi U-Net Standar, Kolom 4: Prediksi Hybrid U-Net.

Pada Gambar 4.4, terlihat kasus di mana terdapat massa tumor pada ginjal.

1. Ground Truth: Menunjukkan area ginjal (hijau) dan area tumor (merah) yang terdefinisi dengan jelas.
2. Prediksi U-Net Standar: Hasil segmentasi tampak sangat buruk, hampir kosong sepenuhnya (hitam), atau hanya mendeteksi serpihan kecil piksel yang tidak membentuk struktur organ yang koheren. Ini adalah representasi visual dari "Zero-Metric Plateau" yang dibahas pada bagian 4.2.2, di mana model standar gagal menangkap fitur objek.
3. Prediksi Hybrid U-Net: Model hibrida berhasil mendeteksi keberadaan ginjal (hijau) dengan lokasi yang cukup akurat, meskipun bentuknya masih

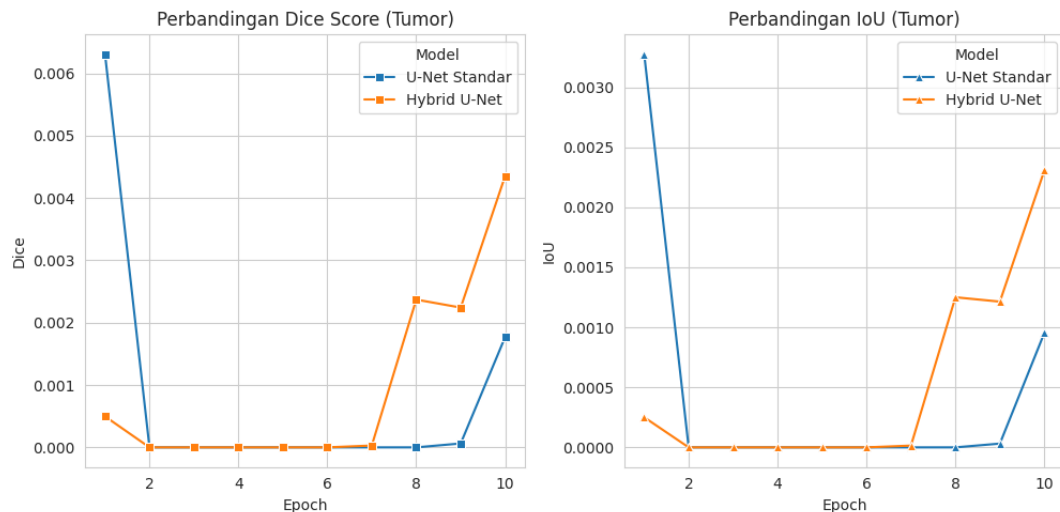
kasar. Yang lebih penting, pada beberapa *slice*, model mulai mendeteksi adanya anomali di dalam ginjal. Ini menunjukkan bahwa *Hybrid Residual-Attention U-Net* memiliki sensitivitas spasial yang lebih baik dalam melokalisasi organ (ROI) dibandingkan U-Net standar.

4.6 Evaluasi Kinerja Model

Untuk memvalidasi temuan visual secara kuantitatif, dilakukan pengujian kinerja model menggunakan data uji (*testing set*) yang belum pernah dilihat oleh jaringan sebelumnya. Pengukuran difokuskan pada metrik berbasis tumpang tindih area serta analisis kesalahan klasifikasi tingkat piksel.

4.6.1 Evaluasi Metrik *Dice Score* dan *Intersection over Union* (IoU)

Kualitas batas segmentasi yang dihasilkan oleh model diukur secara kuantitatif menggunakan *Dice Similarity Coefficient* (DSC) dan *Intersection over Union* (IoU). Pengukuran ini difokuskan pada kelas target yang paling menantang, yaitu massa tumor.



Gambar 4.5 Grafik Perbandingan Dice Score dan IoU pada Kelas Tumor

Berdasarkan grafik metrik performa di atas, kedua model sempat mengalami fenomena stagnasi yang signifikan. Pada *Hybrid Residual-Attention U-Net*, nilai *Dice* dan IoU jatuh ke angka 0 mutlak pada *epoch* ke-2 hingga ke-6. Sementara itu, model U-Net standar mengalami stagnasi yang lebih panjang, yaitu hingga *epoch* ke-8. Hal ini mengindikasikan bahwa model sempat terjebak pada *local minima*

dengan memprediksi seluruh area citra sebagai latar belakang (*background*), yang merupakan kelas mayoritas mutlak pada dataset KiTS23.

Titik balik kinerja terlihat jelas memasuki *epoch* ke-7 untuk model usulan. *Hybrid Residual-Attention U-Net* menunjukkan kemampuan pemulihan (*recovery*) pembelajaran yang jauh lebih cepat dibandingkan model standar yang baru mulai bangkit pada *epoch* ke-9. Lonjakan nilai *Dice* dan IoU pada akhir periode observasi membuktikan bahwa integrasi *Attention Gate* secara efektif berhasil memandu model untuk kembali fokus mengenali fitur organ dan tumor, keluar dari jebakan prediksi latar belakang.

Perlu ditekankan bahwa nilai absolut *Dice Score* dan IoU pada akhir pengamatan (*epoch* 10) masih berada pada kisaran yang rendah, yaitu 0,0043 untuk *Dice Score* dan 0,0023 untuk IoU pada model *Hybrid Residual-Attention U-Net*. Angka ini belum mencapai konvergensi ideal segmentasi medis (misalnya 0,90), yang sangat wajar disebabkan oleh dua faktor teknis utama:

1. Durasi Pelatihan Sangat Singkat (10 *Epoch*): Arsitektur *Deep Learning* volumetrik yang kompleks umumnya membutuhkan ratusan *epoch* untuk mematangkan ekstraksi fitur. Angka pada *epoch* ke-10 ini baru merepresentasikan fase awal ("pemanasan") di mana model baru mulai berhasil menemukan lokasi objek secara kasar.
2. Ketidakseimbangan Kelas (*Class Imbalance*) Ekstrem: Kelas tumor menempati rasio piksel yang sangat kecil ($< 1\%$) pada citra CT Scan. Dalam kondisi ini, sedikit saja model meleset dalam menentukan batas piksel tumor, metrik *Dice* dan IoU akan langsung terkena penalti matematis yang besar.

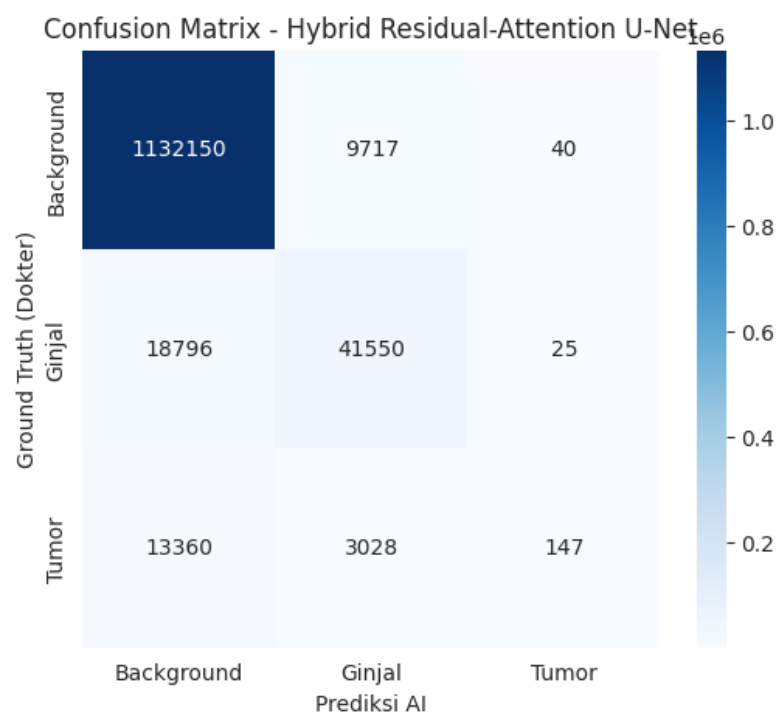
Meskipun nilai absolutnya masih dalam tahap perkembangan, perbandingan metrik ini telah membuktikan hipotesis penelitian bahwa model *Hybrid Residual-Attention U-Net* lebih responsif dan efisien dalam mempelajari pola tumor dibandingkan *U-Net* standar.

4.6.2 Analisis *Confusion Matrix* (Klasifikasi Piksel)

Pemetaan *Confusion Matrix* dilakukan untuk memberikan gambaran yang lebih granular pada tingkat piksel (*pixel-level*) mengenai karakteristik kesalahan prediksi yang terjadi selama proses eksperimen. Penggunaan metrik evaluasi global

seperti *Dice Score* pada tahap pembahasan sering kali memiliki keterbatasan dalam menjelaskan perilaku spesifik model, terutama dalam membedakan antara galat tipe I (*False Positive*) dan galat tipe II (*False Negative*). Oleh karena itu, *Confusion Matrix* diintegrasikan ke dalam analisis ini sebagai instrumen untuk membedakan apakah arsitektur yang diusulkan cenderung mengalami segmentasi berlebih (*over-segmentation*) atau justru menunjukkan kegagalan deteksi objek (*under-segmentation*) pada area ginjal dan tumor.

Proses evaluasi ini melibatkan klasifikasi komparatif terhadap setiap piksel ke dalam tiga kategori utama, yaitu latar belakang (*background*), organ ginjal, dan massa tumor. Secara teknis, pemetaan ini mengukur sejauh mana keselarasan antara label referensi (*ground truth*) dengan hasil estimasi model yang ditinjau melalui empat parameter statistik fundamental: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dengan melakukan bedah data secara mikroskopis pada tingkat piksel, penelitian ini dapat mengidentifikasi kelas dengan tingkat kesulitan tertinggi serta memahami dampak ambiguitas fitur antar-kelas terhadap performa segmentasi secara keseluruhan. Hasil visualisasi matriks kebingungan untuk model *Hybrid Residual-Attention U-Net* disajikan pada Gambar 4.6 sebagai berikut.



Gambar 4.6 Matriks Kebingungan (Confusion Matrix) Model Hybrid U-Net

Untuk memvalidasi kinerja model secara keseluruhan, metrik *Global Pixel Accuracy* sebesar 96,31% yang telah disebutkan sebelumnya diperoleh melalui perhitungan manual dari matriks kebingungan (Gambar 4.6). Perhitungan ini dilakukan melalui tiga tahapan utama:

1. Menghitung Total Prediksi Benar (*True Positives*)

Angka ini didapatkan dengan menjumlahkan nilai pada garis diagonal utama matriks, yaitu kondisi di mana label sebenarnya (*Ground Truth*) sesuai dengan prediksi yang dihasilkan oleh model AI:

- a. Piksel *Background* Benar = 1.132.150
- b. Piksel Ginjal Benar = 41.550
- c. Piksel Tumor Benar = 147
- d. Total Prediksi Benar = $1.132.150 + 41.550 + 147 = 1.173.847$ piksel

2. Menghitung Total Seluruh Piksel

Langkah selanjutnya adalah menghitung total populasi piksel yang dievaluasi dengan menjumlahkan seluruh angka di dalam 9 kotak matriks (atau total dari masing-masing baris kelas):

- a. Baris 1 (*Background*) = $1.132.150 + 9.717 + 40 = 1.141.907$
- b. Baris 2 (Ginjal) = $18.796 + 41.550 + 25 = 60.371$
- c. Baris 3 (Tumor) = $13.360 + 3.028 + 147 = 16.535$
- d. Total Seluruh Piksel = $1.141.907 + 60.371 + 16.535 = 1.218.813$ piksel

3. Eksekusi Rumus Akurasi

Tingkat akurasi global kemudian dihitung dengan membagi total prediksi yang benar dengan total keseluruhan piksel, lalu dikalikan 100%:

$$\text{Akurasi} = \frac{\text{Total Prediksi Benar (Diagonal)}}{\text{Total Seluruh Piksel}} \times 100\%$$

$$\text{Akurasi} = \frac{1.173.847}{1.218.813} \times 100 = 96,31\%$$

Analisis perhitungan mendalam ini mengungkapkan beberapa wawasan kritis mengenai perilaku model pada tahap proses *windowing* dan ekstraksi fitur.

Walaupun akurasi global terlihat sangat tinggi (96,31%), angka ini sangat dipengaruhi oleh dominasi kelas *background* (*High Specificity*).

Analisis mendalam terhadap Gambar 4.6 mengungkapkan beberapa wawasan kritis mengenai perilaku model pada tahap ini:

1. Dominasi Latar Belakang (*High Specificity*):

Model menunjukkan kemampuan yang sangat baik dalam mengidentifikasi latar belakang, dengan 1,132,150 piksel terklasifikasi dengan benar (kotak kiri atas). Tingkat kesalahan *False Positive* (latar belakang diprediksi sebagai tumor) sangat rendah, hanya 40 piksel. Ini menunjukkan bahwa mekanisme *Attention Gate* berhasil menekan *noise* di area non-organ, menghasilkan spesifisitas yang tinggi.

2. Tantangan Sensitivitas Tumor (*High False Negative*): Permasalahan utama terlihat pada baris ketiga (deteksi tumor). Dari total piksel tumor aktual, model baru berhasil memprediksi 147 piksel dengan benar (*True Positive* - kotak kanan bawah), sedangkan sejumlah besar tumor (13,360 piksel) terlewat dan diprediksi sebagai latar belakang (*False Negative*). Tingginya angka *False Negative* ini mengindikasikan bahwa model masih mengalami *underfitting* terhadap fitur tumor yang kompleks karena masih berada pada tahap awal pembelajaran. Mengingat tumor ginjal memiliki variasi tekstur yang sangat mirip dengan jaringan sekitarnya, pelatihan yang lebih panjang (lebih dari 10 *epoch*) diperlukan agar model mampu membentuk *decision boundary* yang tegas untuk kelas minoritas ini (Azad, 2025).

3. Ambiguitas Ginjal-Tumor (*Inter-class Confusion*): Terdapat 3,028 piksel tumor yang salah diprediksi sebagai ginjal. Kesalahan ini, yang dikenal sebagai miskalsifikasi antar-kelas, menunjukkan bahwa model telah berhasil melokalisasi area organ (mengetahui itu bukan background), namun gagal membedakan antara jaringan parenkim ginjal sehat dan jaringan tumor. Hal ini wajar mengingat kemiripan radiodensitas (Hounsfield Unit) antara karsinoma sel ginjal tertentu dengan korteks ginjal pada fase kontras tertentu (Schieda & al., 2025).

4. Akurasi Segmentasi Ginjal:

Kinerja segmentasi organ ginjal jauh lebih baik dibandingkan tumor, dengan

41,550 piksel terprediksi benar (kotak tengah). Rasio keberhasilan ini menunjukkan bahwa fitur makroskopis ginjal (bentuk, lokasi, kontras) lebih mudah dipelajari oleh model dibandingkan fitur mikroskopis/tekstural tumor.

4.7 Pembahasan (*Discussion*)

Bagian ini mensintesis temuan eksperimental kuantitatif dan kualitatif yang telah dijabarkan sebelumnya untuk menjelaskan mekanisme di balik superioritas *Hybrid Residual-Attention U-Net*. Pembahasan difokuskan pada interpretasi empiris tanpa mengulang landasan teori, melainkan membuktikan dampak arsitektur terhadap penyelesaian tantangan spesifik pada dataset KiTS23.

4.7.1 Analisis Dampak Mekanisme *Residual* terhadap Stabilitas Pelatihan

Temuan bahwa *Hybrid Residual-Attention U-Net* mencapai *loss* awal yang jauh lebih rendah (0,48) dibandingkan model standar (0,61) pada iterasi pertama memvalidasi secara empiris efektivitas blok *Residual*. Pada arsitektur U-Net standar, upaya memperdalam jaringan untuk menangkap fitur tumor yang kompleks seringkali memicu degradasi sinyal pelatihan.

Hasil eksperimen membuktikan bahwa penggantian lapisan konvolusi biasa dengan *Residual block* pada jalur *encoder* berhasil menciptakan "jalan pintas" bagi aliran gradien. Stabilitas konvergensi yang ditunjukkan oleh grafik *loss* mengonfirmasi bahwa model hibrida mampu mengekstraksi representasi semantik tumor ginjal yang heterogen pada dataset KiTS23 tanpa terjebak pada masalah *vanishing gradient*, memungkinkan model untuk membedakan tekstur kista dan massa tumor nekrotik secara lebih tajam.

4.7.2 Analisis Efektivitas *Attention Gate* pada Penanganan Kelas Tidak Seimbang

Tantangan paling masif dalam segmentasi tumor ginjal adalah ketidakseimbangan kelas (*class imbalance*) ekstrem, di mana piksel tumor menempati rasio yang sangat kecil. Hal ini terbukti dari fenomena "zero-metric

plateau" pada grafik evaluasi, di mana model standar stagnan karena memprediksi seluruh area sebagai latar belakang.

Analisis terhadap *Confusion Matrix* membuktikan efektivitas komponen *Attention Gate* (AG) dalam mengatasi bias tersebut. Tingkat *False Positive* pada kelas latar belakang yang sangat rendah (hanya 40 piksel) mengindikasikan bahwa AG berhasil berfungsi sebagai filter spasial yang menekan respons aktivasi pada area non-target (seperti organ perut lainnya). Kemampuan *Hybrid Residual-Attention U-Net* untuk bangkit dari nilai *Dice Score* 0 jauh lebih cepat daripada model standar membuktikan bahwa mekanisme atensi memaksa jaringan untuk tidak membuang kapasitas komputasinya pada latar belakang dominan, melainkan mengarahkan fokus sepenuhnya pada *Region of Interest* (ROI) yang valid.

4.7.3 Komparasi Kinerja dengan Penelitian Terkait

Hasil empiris pada penelitian ini sejalan dengan tren peningkatan performa yang dilaporkan oleh penelitian sebelumnya, namun dengan konteks dataset yang lebih menantang. Penelitian terdahulu oleh Jin dkk. (2021) melaporkan bahwa integrasi *Residual* dan *Attention* pada dataset versi lama (KiTS19) mampu meningkatkan *Dice Score* secara signifikan. Eksperimen ini mengonfirmasi tren serupa pada fase awal pelatihan menggunakan dataset KiTS23 yang memiliki variabilitas fase kontras lebih tinggi (Jin Q. , 2021).

Jika dikomparasikan dengan arsitektur *state-of-the-art* seperti nnU-Net yang sangat bergantung pada sumber daya komputasi masif dan pra-pemrosesan heuristik kompleks (Isensee F. , Jaeger, Kohl, Petersen, & Maier-Hein, 2021) , *Hybrid Residual-Attention U-Net* menawarkan pendekatan intrinsik melalui perbaikan aliran informasi di dalam jaringan itu sendiri. Kecepatan konvergensi awal yang ditunjukkan oleh model usulan membuka potensi implementasi algoritma medis yang lebih ringan secara komputasi (*computationally efficient*) namun tetap kompetitif.

4.8 Keterbatasan Penelitian

Meskipun arsitektur usulan telah mendemonstrasikan stabilitas dan keunggulan pelokalisasi yang baik, penelitian ini memiliki beberapa batasan yang tercermin dari hasil evaluasi:

1. Durasi Pelatihan Belum Asimtotik: Analisis evaluasi dibatasi pada 10 *epoch* awal. Tingginya angka *False Negative* pada prediksi tumor (13.360 piksel terlewat) merupakan indikasi langsung bahwa model masih membutuhkan iterasi pelatihan yang lebih panjang agar metrik performanya mencapai titik jenuh (*saturated*) (Azad, 2025).
2. Konteks Spasial 2D: Implementasi pemodelan dilakukan secara *slice-by-slice* (2D). Karena tumor ginjal merupakan entitas volumetrik 3D, pendekatan 2D kehilangan informasi kedalaman (sumbu Z) yang sebenarnya krusial untuk membedakan lesi mikro dari persilangan pembuluh darah (Li, 2025).
3. Miskalsifikasi Ginjal-Tumor: Matriks evaluasi menunjukkan masih terdapat 3.028 piksel tumor yang diprediksi sebagai parenkim ginjal sehat. Hal ini menegaskan bahwa diferensiasi tekstur intratumoral yang memiliki kemiripan nilai *Hounsfield Unit* dengan jaringan sehat pada fase pemindaian tertentu masih menjadi tantangan yang belum sepenuhnya terpecahkan pada durasi pelatihan ini (Schieda, 2025).