

SEGMENTASI SINYAL PADA SPEKTROGRAM RF MENGGUNAKAN ATTENTION U-NET DAN PRA-PELATIHAN SINTETIS

Andreo Yustiantoro Anjaya, Arief Setyanto

Program Studi Magister Informatika, Universitas Amikom Yogyakarta
Jl. Ring Road Utara, Condong Catur, Sleman, Yogyakarta
andreodeo@students.amikom.ac.id

ABSTRAK

Pemantauan spektrum frekuensi radio (RF) pita lebar memerlukan pelokalan sinyal yang presisi pada citra spektrogram waktu-frekuensi. Namun, banyak sinyal muncul sebagai jejak vertikal sangat tipis (1–3 piksel) dan berada pada latar belakang derau yang bervariasi akibat perbedaan perangkat dan kondisi pengukuran. Kondisi ini menyebabkan pendekatan deteksi berbasis *bounding box* kurang optimal karena proses *downsampling* berpotensi menghilangkan detail objek kecil. Penelitian ini bertujuan meningkatkan akurasi pelokalan sinyal melalui segmentasi tingkat piksel menggunakan arsitektur *Attention U-Net*. Untuk mengatasi keterbatasan data riil berlabel dan kesenjangan distribusi antara data sintetis dan data lapangan, diterapkan strategi pelatihan dua tahap berupa *pretraining* pada dataset sintetis yang diselaraskan secara statistik dengan data riil, kemudian dilanjutkan *fine-tuning* pada dataset riil multi-band. Hasil evaluasi pada data uji riil menunjukkan peningkatan performa bertahap dari *U-Net* (*Dice* 0.3789) ke *Attention U-Net* (*Dice* 0.6196), hingga konfigurasi *pretraining-fine-tuning* yang mencapai *Dice* 0.8786 dan *IoU* 0.7908 dengan konvergensi lebih stabil. Temuan ini menunjukkan bahwa kombinasi mekanisme *attention* dan pelatihan dua tahap efektif meningkatkan presisi dan generalisasi segmentasi sinyal pada lingkungan pemantauan yang heterogen.

Kata kunci : Segmentasi Spektrogram, Attention U-Net, Pra-pelatihan Sintetis, Adaptasi Domain, Pemantauan Spektrum RF

1. PENDAHULUAN

Perkembangan pesat sistem komunikasi nirkabel dan meningkatnya pemanfaatan spektrum frekuensi radio secara bersama mendorong kebutuhan akan pemantauan spektrum frekuensi radio (RF) pita lebar yang semakin andal dan otomatis. Sistem pemantauan modern dituntut untuk beroperasi pada rentang frekuensi yang luas dan dalam lingkungan yang sangat heterogen, di mana perbedaan perangkat, konfigurasi sensor, serta karakteristik derau latar dapat bervariasi secara signifikan. Variabilitas ini menghasilkan pola spektrogram yang kompleks dan dinamis, sehingga menyulitkan proses identifikasi sinyal semakin menantang [1], [2], [3].

Berbagai pendekatan tradisional, seperti deteksi energi dan metode berbasis ambang (*thresholding*), relatif ringan secara komputasi namun cenderung mengalami penurunan performa pada kondisi rasio sinyal terhadap derau (SNR) rendah, interferensi tinggi, serta derau yang tidak stasioner [3], [4]. Seiring perkembangan *deep learning*, pemantauan spektrum frekuensi radio (RF) mulai diformulasikan sebagai persoalan visi komputer melalui representasi spektrogram waktu-frekuensi. Model deteksi objek seperti *Faster R-CNN* dan *YOLO* telah menunjukkan peningkatan ketangguhan dalam mengenali sinyal pada spektrogram pita lebar [5], [6], [7]. Namun demikian, pendekatan berbasis *bounding box* memiliki keterbatasan dalam mempertahankan detail spasial sinyal yang sangat tipis, khususnya ketika lebar sinyal hanya 1–3 piksel.

Selain itu, sebagian besar studi terdahulu dievaluasi pada kondisi yang relatif terkontrol dan

belum sepenuhnya merepresentasikan kompleksitas pemantauan pita lebar di lapangan, yang melibatkan perangkat heterogen, tekstur latar beragam, serta variasi domain antar lokasi. Ketersediaan dataset riil berlabel dalam jumlah besar juga masih terbatas, sementara dataset sintetis yang digunakan dalam penelitian sebelumnya sering kali belum diselaraskan secara statistik dengan distribusi data lapangan. Kondisi ini menimbulkan kesenjangan domain (*domain gap*) yang berpotensi menurunkan kemampuan generalisasi model.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengembangkan pendekatan segmentasi tingkat piksel untuk meningkatkan presisi pelokalan sinyal pada spektrogram RF pita lebar yang heterogen, khususnya pada sinyal berukuran sangat kecil atau tipis. Untuk mencapai tujuan tersebut, penelitian ini mengusulkan penggunaan arsitektur *Attention U-Net* yang mampu mempertahankan struktur spasial halus melalui mekanisme *attention*, serta menerapkan strategi pelatihan dua tahap berupa *pretraining* pada dataset spektrogram sintetis berbasis distribusi statistik data riil, diikuti *fine-tuning* pada dataset lapangan. Melalui kombinasi arsitektur dan strategi pelatihan tersebut, diharapkan model mampu meningkatkan akurasi segmentasi sekaligus memperbaiki kemampuan generalisasi terhadap variasi domain pada sistem pemantauan spektrum dunia nyata.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Beberapa penelitian telah memanfaatkan *deep learning* untuk analisis sinyal pada spektrogram RF.

Nguyen et al. [1] dan Benazzouza et al. [8] mengembangkan sistem identifikasi sinyal pita lebar berbasis deep learning dengan pendekatan end-to-end yang menunjukkan akurasi tinggi dan performa real-time. Keduanya berfokus pada klasifikasi jenis sinyal dalam konteks spektrum lebar, dengan representasi spektrogram sebagai pola global untuk pengenalan kelas. Pendekatan ini efektif untuk identifikasi spektral, namun belum menekankan pelokalan spasial presisi tingkat piksel pada struktur sinyal yang sangat kecil dan muncul dalam jumlah banyak pada satu citra.

Alammar et al. [9] menerapkan *deep learning* untuk mengestimasi area okupansi sinyal pada spektrogram. Pendekatan ini menunjukkan peningkatan akurasi estimasi dibanding metode berbasis *threshold* tradisional. Model difokuskan pada perhitungan luasan sinyal sebagai indikator okupansi, dengan struktur objek yang relatif utuh dan terdefinisi jelas. Kompleksitas morfologi seperti sinyal sangat tipis, terfragmentasi, atau saling berhimpitan pada resolusi tinggi belum menjadi fokus utama dalam konfigurasi eksperimen tersebut.

Su et al. [2] mengembangkan sistem identifikasi spektrum untuk membedakan berbagai jenis sinyal nirkabel berbasis fitur spektral yang dipelajari secara otomatis. Model menunjukkan peningkatan akurasi klasifikasi dibanding pendekatan berbasis fitur manual. Fokus penelitian berada pada pengenalan kelas sinyal secara global, dengan dataset yang relatif konsisten dalam karakteristik visualnya. Aspek variasi domain lintas perangkat serta kompleksitas struktur sinyal berskala sangat kecil tidak menjadi variabel utama dalam evaluasi.

Pendekatan deteksi berbasis bounding box dikembangkan oleh Truong et al. [6] menggunakan YOLOv5, serta oleh Li et al. [3] untuk deteksi sinyal HF pada wideband spectrogram. Kedua penelitian menunjukkan peningkatan performa deteksi dibanding metode klasik dan efisiensi komputasi yang baik. Namun, karakteristik deteksi berbasis kotak pembatas dengan *backbone ber-downsampling* progresif lebih sesuai untuk objek dengan representasi spasial cukup jelas terhadap grid fitur. Pada struktur sinyal yang sangat tipis (1–3 piksel) dalam spektrogram resolusi tinggi, pendekatan ini berpotensi menghadapi keterbatasan representasi detail spasial.

Berbagai penelitian terdahulu tersebut telah menunjukkan kemajuan dalam klasifikasi dan deteksi sinyal berbasis spektrogram, terutama pada skenario dengan karakteristik visual yang relatif terdefinisi. Namun, pemantauan pita lebar beresolusi tinggi dengan sinyal sangat kecil dan latar heterogen memerlukan pendekatan yang lebih presisi. Penelitian ini mengusulkan segmentasi berbasis Attention U-Net dengan strategi pretraining–fine-tuning menggunakan dataset sintesis yang diselaraskan secara statistik untuk meningkatkan presisi dan generalisasi model.

2.2. Spektrogram

Sinyal nirkabel pada pemantauan spektrum frekuensi radio (RF) bersifat non-stasioner, sehingga karakteristik frekuensinya berubah terhadap waktu. Analisis *Fourier* klasik hanya memberikan informasi frekuensi secara global dan tidak mampu merepresentasikan dinamika temporal sinyal. Oleh karena itu, digunakan pendekatan analisis waktu–frekuensi yang mampu menggambarkan evolusi sinyal secara simultan dalam domain waktu dan frekuensi [10].

Spektrogram memvisualisasikan distribusi energi sinyal pada bidang waktu–frekuensi, dengan intensitas piksel merepresentasikan level daya lokal. Representasi dua dimensi ini membentuk pola visual khas untuk setiap jenis emisi, sehingga pemantauan spektrum dapat diformulasikan sebagai permasalahan analisis citra dan secara alami sesuai untuk diproses menggunakan pendekatan deep learning berbasis pembelajaran fitur spasial.

2.3. Deteksi Objek dan Segmentasi Semantik

Pelokalan objek pada citra umumnya dilakukan melalui dua pendekatan utama, yaitu deteksi objek dan segmentasi semantik. Deteksi objek memprediksi lokasi dalam bentuk *bounding box* beserta kelasnya, sedangkan segmentasi semantik memberikan label pada setiap piksel sehingga menghasilkan delineasi spasial yang lebih detail. Perkembangan *Convolutional Neural Networks (CNNs)* secara signifikan meningkatkan kinerja kedua pendekatan tersebut karena kemampuannya mempelajari representasi fitur spasial secara hierarkis langsung dari data [11].

Dalam deteksi objek modern, metode diklasifikasikan menjadi dua-tahap, seperti *R-CNN*, *Fast R-CNN*, dan *Faster R-CNN* [12], serta satu-tahap seperti YOLO [13]. Metode dua-tahap umumnya menawarkan akurasi pelokalan lebih tinggi dengan biaya komputasi lebih besar, sedangkan metode satu-tahap unggul dalam efisiensi dan kecepatan inferensi. Namun, keduanya menggunakan *downsampling progresif* pada *backbone* yang dapat mengurangi resolusi fitur, sehingga kurang optimal untuk objek berukuran sangat kecil. Pada spektrogram RF, sinyal sering muncul sebagai jejak vertikal sangat tipis di tengah latar belakang kompleks. Dalam kondisi tersebut, segmentasi semantik dengan prediksi padat (*dense prediction*) lebih mampu mempertahankan detail spasial dan lebih sesuai untuk struktur tipis atau berdekatan.

2.4. U_Net dan Attention U-Net

Arsitektur segmentasi berbasis *fully convolutional networks (FCNs)* telah menjadi pendekatan utama dalam tugas prediksi padat yang memerlukan pelokalan presisi tinggi. Salah satu arsitektur yang paling luas digunakan adalah *U-Net*, yang mengadopsi struktur *encoder–decoder* simetris. *Encoder* mengekstraksi fitur melalui konvolusi dan

downsampling, sedangkan *decoder* melakukan *upsampling* untuk merekonstruksi resolusi spasial sehingga prediksi dapat dilakukan pada tingkat piksel. Keunggulan utama *U-Net* terletak pada penggunaan *skip connections* yang menggabungkan fitur resolusi tinggi dari *encoder* dengan fitur semantik dari *decoder*, sehingga detail struktur kecil tetap dapat dipertahankan selama proses rekonstruksi [14].

Pengembangan lebih lanjut dari arsitektur ini adalah *Attention U-Net*, yang menambahkan *attention gate* pada jalur *skip connection*. Mekanisme perhatian tersebut berfungsi menyaring fitur encoder sebelum digabungkan dengan decoder, sehingga jaringan lebih fokus pada area yang relevan dan menekan respons latar belakang yang kurang informatif. Pendekatan ini terbukti meningkatkan ketahanan model terhadap noise dan kompleksitas latar belakang [15], serta lebih efektif dalam menangani struktur objek berskala kecil seperti sinyal tipis pada spektrogram RF.

2.5. Domain Gap & Data Sintetis

Dalam pembelajaran mesin, model yang dilatih pada satu dataset sering mengalami penurunan performa ketika diterapkan pada data dengan karakteristik berbeda, fenomena yang dikenal sebagai *domain shift* ($P_s(x, y) \neq P_t(x, y)$). Pada pemantauan spektrum RF, pergeseran ini dapat disebabkan oleh variasi perangkat, konfigurasi sensor, dan kondisi derau. Untuk mengatasinya, dikembangkan pendekatan *transfer learning* dan adaptasi domain yang mentransfer pengetahuan dari domain sumber ke domain target [16], [17], [18].

Strategi yang umum digunakan adalah *pretraining* diikuti *fine-tuning*, di mana model terlebih dahulu dilatih pada dataset berskala besar, kemudian disesuaikan menggunakan data target berlabel [17]. Data sintetis sering dimanfaatkan untuk menyediakan sampel berlabel dalam jumlah besar, namun pelatihan murni pada data sintetis berisiko membuat model mempelajari artefak yang tidak sepenuhnya merepresentasikan kondisi riil [19]. Oleh karena itu, penyesuaian distribusi statistik antara data sintetis dan

data riil, baik dari sisi tampilan visual maupun distribusi objek, menjadi penting [20]. Pendekatan generasi sintetis berbasis data memungkinkan pembentukan representasi morfologi sinyal yang lebih relevan pada tahap pra-pelatihan, yang kemudian disempurnakan melalui *fine-tuning* pada data lapangan untuk meningkatkan generalisasi pada kondisi operasional nyata.

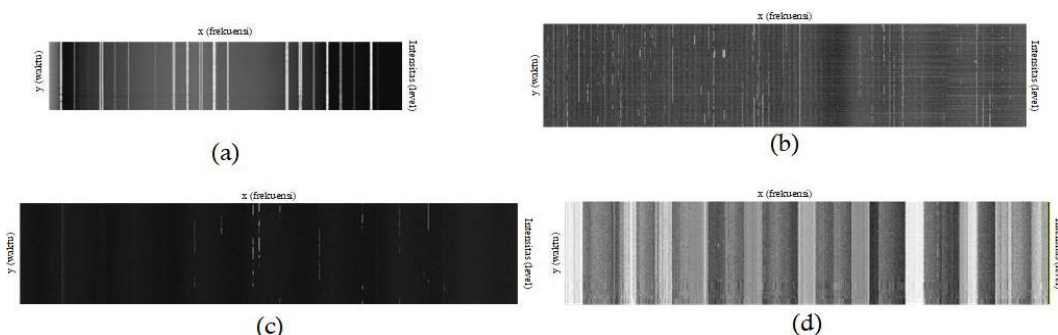
3. METODE PENELITIAN

Dalam penelitian ini terdapat beberapa tahapan penelitian yang dilakukan, mulai dari pengumpulan dataset riil, data, proses generasi dataset sintetis, persiapan dan pra-pengolahan data, implementasi arsitektur model dan strategi pelatihan, hingga metode evaluasi kinerja. Penyusunan metodologi dilakukan secara sistematis untuk memastikan bahwa pendekatan yang dikembangkan mampu menangani kompleksitas spektrogram pita lebar dengan variasi morfologi sinyal dan karakteristik latar yang heterogen..

3.1. Pengumpulan Dataset Riil

Dataset riil diperoleh dari sistem pemantauan spektrum RF pada beberapa lokasi dengan tiga jenis perangkat berbeda, sehingga mencerminkan variasi derau, tekstur latar, dan dinamika visual antar domain operasional. Setiap sampel berupa citra spektrogram grayscale, dengan sumbu frekuensi (horizontal), waktu (vertikal), dan intensitas yang merepresentasikan level daya. Pemantauan mencakup pita *FM* (87–108 MHz), *airband* (108–137 MHz), *two-way radio* (150–174 MHz), dan *DVB-T2* (478–694 MHz), menghasilkan citra beresolusi horizontal besar (>1500 piksel) di mana banyak sinyal hanya menempati area kecil.

Secara morfologi, sinyal meliputi pola kontinu, intermiten, berdempetan, hingga objek sangat kecil (1–3 piksel) yang mendominasi dataset dan rentan terhadap kehilangan detail spasial. Setiap citra dianotasi menggunakan bounding box serta mask segmentasi biner untuk eksperimen pelokalan tingkat piksel.



Gambar 1. Citra Spektrogram

(a) *FM Broadcast*, (b) Komunikasi Radio (*two-way radio*), (c) Penerbangan, dan (d) *DVB-T2*

Gambar 1 menunjukkan variasi karakteristik visual antar pita frekuensi. *FM broadcast* membentuk blok kontinu yang lebar, *two-way radio* menampilkan

sinyal sempit berulang, *airband* cenderung berupa burst pendek, sedangkan *DVB-T2* memiliki struktur spektrum yang lebar dan padat. Perbedaan morfologi

dan tekstur latar ini mencerminkan kompleksitas dataset serta pentingnya segmentasi yang presisi terhadap objek kecil.

Untuk memberikan gambaran menyeluruh mengenai komposisi dan kompleksitas data yang digunakan, ringkasan karakteristik utama dataset riil disajikan pada Tabel 1.

Tabel 1. Ringkasan Karakteristik dataset riil

Deskripsi	Keterangan	Deskripsi	Keterangan
Format data	Spektrogram grayscale	Jumlah objek/sinyal	15815 sinyal
Representasi	x: frekuensi, y: waktu, intensitas: level daya	Pola sinyal	kontinu, intermiten, berdempetan, dot
Jenis perangkat	3 tipe perangkat monitoring	Ketebalan sinyal	1 – 87 piksel
Lokasi Pengukuran	21 lokasi	Objek per citra	1 – 166 objek per citra
Jenis sinyal/rentang frekuensi	FM 87–108 MHz; Airband 108–137 MHz; Two-way 150–174 MHz; DVB-T2 478–694 MHz	Karakteristik latar	Heterogen, noise bervariasi antar lokasi
Ukuran citra	843 x 227 – 1583 x 346 piksel (maks w:1583piksel; maks. h: 460 piksel)	Anotasi	bounding box + mask segmentasi
Jumlah citra/dataset	521 citra spektrogram	Pembagian dataset	train / validation / test

Tabel 1 menunjukkan kompleksitas dataset dengan 15.815 objek pada 521 citra dan variasi ketebalan 1–87 piksel. Jumlah objek hingga 166 per citra serta latar heterogen menjadikan pelokalan, khususnya pada sinyal tipis, lebih menantang. Karakteristik ini mendasari pemilihan segmentasi berbasis attention dan strategi pelatihan bertahap.

3.2. Generasi Dataset Sintetis

Keterbatasan jumlah data riil, variasi perangkat monitoring, serta tingginya biaya anotasi manual mendorong pengembangan dataset sintetis berbasis data (*data-driven synthetic generation*). Proses sintesis tidak dilakukan secara acak, melainkan dirancang untuk meniru distribusi statistik dataset riil guna mengurangi domain gap dan meningkatkan kemampuan generalisasi model.

Proses generasi diawali dengan analisis dataset riil untuk mengekstraksi distribusi ukuran kanvas, jumlah dan proporsi jenis sinyal, ketebalan, pola temporal, serta karakteristik latar belakang. Parameter-parameter tersebut dirumuskan dalam berkas konfigurasi (*YAML*) yang mengatur blueprint ukuran citra, sintesis *background* menyerupai *noise* spektrum riil, penempatan objek tanpa tumpang tindih, serta proses *rendering* berbagai tipe sinyal (*FM*, *DVB-T2*, dan *two-way radio*). Sistem juga menghasilkan anotasi otomatis berupa *bounding box* dan *mask* segmentasi sehingga konsistensi label tetap terjaga dalam skala besar. Pendekatan ini memungkinkan kontrol terhadap variasi morfologi sinyal sekaligus menjaga kesesuaian distribusi dengan data lapangan.

Dataset sintetis digunakan pada tahap *pretraining* untuk membangun representasi awal geometri dan pola sinyal, kemudian model disempurnakan melalui *fine-tuning* menggunakan data riil. Ringkasan komposisi dataset sintetis ditampilkan pada Tabel 2.

Tabel 2 menunjukkan bahwa dataset sintetis mempertahankan kompleksitas multi-objek serta dominasi sinyal tipis (1–3 piksel), khususnya pada

kelas *two-way radio*. Distribusi jumlah objek per citra dan resolusi kanvas disesuaikan dengan data riil sehingga tetap representatif terhadap kondisi operasional pemantauan spektrum pita lebar.

Tabel 2. Ringkasan dataset sintetis

Karakteristik	Keterangan
Jumlah citra sintetis	15000
Jumlah total objek sinyal	412320
Two-way radio (sinyal lebar sempit 1-3 piksel)	223359 (54,17%)
FM Broadcast	142422 (34,54%)
DVB-T2	46359 (11,29%)
Rata-rata objek per citra	27,5 sinyal/citra
Minimum objek per citra	1
Maksimum objek per citra	166
Resolusi kanvas	mengikuti distribusi data riil

Selain meniru proporsi jenis sinyal, generator juga memodelkan variasi pola temporal pada sinyal *two-way radio* untuk merepresentasikan dinamika transmisi yang lebih realistis. Pola yang dihasilkan mencakup sinyal kontinu (tajam dan buram), berbagai bentuk intermiten (periodik, buram, *non-uniform*, dan *sparse*), serta *burst* pendek (dot). Distribusi terbesar berasal dari pola kontinu buram ($\pm 28\%$), *intermiten sparse* ($\pm 22\%$), dan *intermiten non-uniform* ($\pm 20\%$), sementara pola lainnya melengkapi variasi morfologi secara proporsional. Keberagaman ini menghasilkan struktur spasial yang lebih kompleks dan mendekati kondisi transmisi nyata, sehingga dataset sintetis tidak hanya mereplikasi komposisi kelas, tetapi juga karakter temporal dan morfologi sinyal untuk mendukung proses *pretraining* yang lebih representatif.

3.3. Persiapan dan Pengolahan Data

Seluruh citra spektrogram dianotasi menggunakan *bounding box* format *YOLO* pada resolusi asli untuk menjaga presisi koordinat,

mengingat sebagian besar sinyal memiliki ketebalan hanya 1–3 piksel. Untuk eksperimen segmentasi, *bounding box* dikonversi menjadi *mask biner* berbentuk persegi panjang (*rectangular masking*) yang mencakup seluruh area objek. Pendekatan ini dipilih karena anotasi kontur presisi piksel pada struktur sangat tipis dan terfragmentasi berpotensi menimbulkan ketidakstabilan label, sehingga representasi persegi panjang dinilai lebih konsisten dalam merepresentasikan area okupansi sinyal.

Untuk menjaga konsistensi input model, seluruh citra dikonversi ke *grayscale* satu kanal dan ditempatkan pada kanvas tetap berukuran 1583×460 piksel menggunakan *padding* tanpa *resize*. Strategi ini mempertahankan resolusi spasial asli dan mencegah distorsi ketebalan sinyal tipis akibat interpolasi. Warna *padding* disesuaikan dengan rata-rata intensitas latar (abu-abu netral) guna menghindari kontras buatan pada tepi citra.

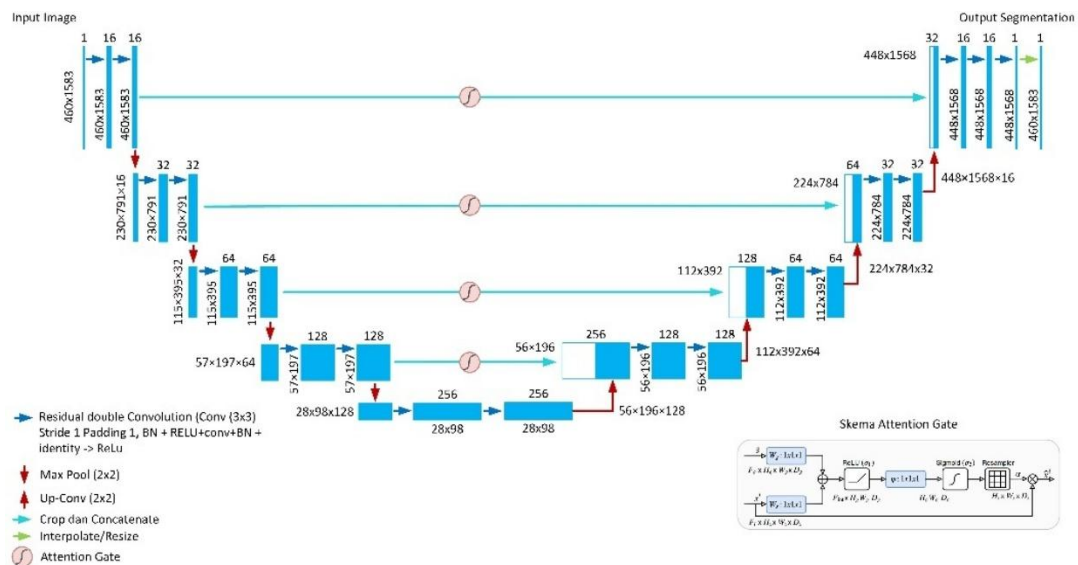
Pembagian dataset dilakukan menggunakan *stratified split* dengan rasio 80/10/10 untuk data pelatihan, validasi, dan pengujian. Pada dataset sintetis diperoleh 12.000/1.500/1.500 sampel, sedangkan dataset riil menghasilkan 417/52/52 sampel. Stratifikasi dilakukan berbasis tag otomatis yang mencakup ketebalan, pola temporal, panjang sinyal, tingkat kontras relatif, dan posisi spasial. Pendekatan

ini memastikan distribusi karakteristik sinyal tetap proporsional pada setiap subset, sehingga evaluasi performa model mencerminkan kompleksitas data operasional secara representatif.

3.4. Implementasi Arsitektur Model dan Strategi Pelatihan

Pada bagian ini menjelaskan arsitektur model segmentasi yaitu *U-Net* dan *Attention U-Net* yang digunakan, strategi pelatihan bertahap untuk mengatasi domain gap antara data sintetis dan data riil, serta detail implementasi teknis yang digunakan selama eksperimen.

Sebagaimana ditunjukkan pada Gambar 2 Arsitektur model *Attention U-Net*, hampir sama dengan *Attention U-net*, *U-Net* menggunakan arsitektur encoder-decoder simetris dengan *skip connection* tanpa *attention gate* untuk mempertahankan detail spasial. *Encoder* mengekstraksi fitur melalui empat tingkat *downsampling*, sedangkan *decoder* melakukan rekonstruksi melalui *upsampling* bertahap. Model menerima input grayscale satu kanal dengan *base channel* 16 dan menggunakan blok konvolusi ganda (*Conv-BatchNorm-ReLU*). Output berupa logits satu kanal untuk segmentasi biner.



Gambar 2. Arsitektur model Attention U-Net

Pada Gambar 2, *Attention U-Net* menambahkan *attention gate* pada setiap *skip connection* untuk menyaring fitur *encoder* sebelum digabungkan dengan *decoder*. Mekanisme ini meningkatkan fokus pada struktur tipis dan kontras rendah serta menekan respons latar belakang yang tidak relevan. Selain itu, blok konvolusi diganti dengan *residual convolution* untuk memperlancar aliran *gradien*. Berdasarkan evaluasi awal, *Attention U-Net* menunjukkan peningkatan *IoU* dan *Dice* dibanding *U-Net standar*, sehingga digunakan sebagai model utama.

Strategi pelatihan dilakukan dalam dua tahap. Pada tahap *pretraining*, model dilatih menggunakan dataset sintetis untuk mempelajari morfologi dasar sinyal dalam jumlah besar dan distribusi terkontrol. Selanjutnya, tahap *fine-tuning* dilakukan menggunakan dataset riil untuk menyesuaikan karakteristik tekstur, noise, dan kontras perangkat pemantauan aktual. Pendekatan ini meningkatkan stabilitas pelatihan dan generalisasi dibanding pelatihan langsung pada data riil terbatas.

Implementasi dilakukan menggunakan *Python* dan *PyTorch* dengan GPU NVIDIA RTX 3060. Citra

diproses dalam format *grayscale* tanpa *resize* dan *dipadding* menjadi 1583×460 piksel. Pelatihan menggunakan kombinasi *BCEWithLogits* dan *Dice Loss* untuk mengatasi ketidakseimbangan piksel, dengan *optimizer Adam/AdamW* dan *scheduler ReduceLROnPlateau*. Augmentasi dibatasi pada transformasi ringan agar morfologi sinyal tetap terjaga.

Tabel 3. Konfigurasi parameter model

Parameter	U-Net	Attention U-Net	Attention U-Net + Pretrain-Finetune
Input	460×1583×1 (padding, grayscale)		
Batch size	2		
Epoch	80	40 & 100	
Learning rate	1x10 ⁻³	1x10 ⁻³ - 1x10 ⁻⁵	
Optimizer	AdamW (wd=1e-4)	Adam	
Scheduler	ReduceLROnPlateau (patience=3)	ReduceLROnPlateau (patience=5)	
Loss	BCEWithLogits + Dice		
Normalisasi	mean=0.5, std=0.5	Tidak digunakan	
Augmentasi	flip + jitter + noise + blur	Dataset sintetis: flip+jitter+noise; Dataset riil: flip saja	
Inisialisasi bobot	Random	Pretrained	
Dataset awal	Riil	Sintetis - Riil	

Tabel 3 merangkum konfigurasi eksperimen yang mencakup variasi jumlah *epoch*, *rentang learning rate*, strategi inisialisasi bobot, dan dataset awal. *Attention U-Net* dengan skema *pretraining-fine-tuning* menggunakan bobot prelatih sintetis sebagai inisialisasi, sedangkan dua skenario lainnya dilatih dari bobot acak pada data riil. Perbedaan konfigurasi ini memungkinkan evaluasi komprehensif terhadap pengaruh arsitektur dan strategi pelatihan terhadap kinerja segmentasi.

3.5. Evaluasi dan Metode Pengukuran Kinerja

Evaluasi dilakukan pada data validasi dan uji menggunakan pendekatan segmentasi piksel. Keluaran model berupa peta probabilitas (*logits*) dikonversi menjadi *mask biner* melalui fungsi *sigmoid* dan *thresholding*, kemudian disempurnakan dengan *post-processing fill_all* untuk menutup lubang kecil sebelum perhitungan metrik.

Kinerja dinilai berdasarkan kesesuaian spasial antara *mask prediksi* dan *ground-truth* menggunakan metrik *IoU*, *Dice*, *Precision*, *Recall*, dan *F1-score*. *IoU* dan *Dice* mengukur tingkat tumpang tindih, dengan *Dice* lebih sensitif terhadap objek kecil, sedangkan *Precision* dan *Recall* mengevaluasi kesalahan deteksi positif dan negatif. Seluruh metrik dihitung pada tingkat piksel dan dirata-ratakan pada seluruh dataset.

Selama pelatihan, *loss* dan metrik validasi dicatat setiap *epoch*, model terbaik dipilih berdasarkan

validation loss terendah, dan kurva pelatihan divisualisasikan untuk memantau stabilitas konvergensi. Pendekatan ini memungkinkan perbandingan yang adil antar skenario berdasarkan kualitas segmentasi piksel.

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil evaluasi kuantitatif serta interpretasi performa dari tiga skenario pelatihan yang diuji, yaitu *U-Net*, *Attention U-Net*, serta *Attention U-Net* dengan strategi *pretraining* pada dataset sintetis yang dilanjutkan *fine-tuning* pada dataset riil. Evaluasi dilakukan pada data uji (test set) yang sepenuhnya terpisah dari proses pelatihan dan validasi, sehingga hasil yang diperoleh merefleksikan kemampuan generalisasi model terhadap data baru.

4.1. Evaluasi Performa

Tabel 4. menyajikan hasil evaluasi segmentasi piksel pada data uji riil untuk ketiga pendekatan yang dibandingkan.

Tabel 4. Evaluasi performa model

Model	IoU	Dice	Precision	Recall
U-Net	0.2719	0.3789	0.4713	0.4604
Attention U-Net	0.5141	0.6196	0.7217	0.7551
Attunet U-Net (pretrain-finetune)	0.7908	0.8786	0.9017	0.8664

Seperti ditunjukkan dalam Tabel 4. Model *U-Net* standar menunjukkan performa terendah dengan *IoU* 0.2719 dan *Dice* 0.3789. Nilai *Precision* dan *Recall* yang berada di kisaran 0.46–0.47 mengindikasikan bahwa model masih mengalami kesulitan dalam membedakan area sinyal dan latar belakang secara konsisten.

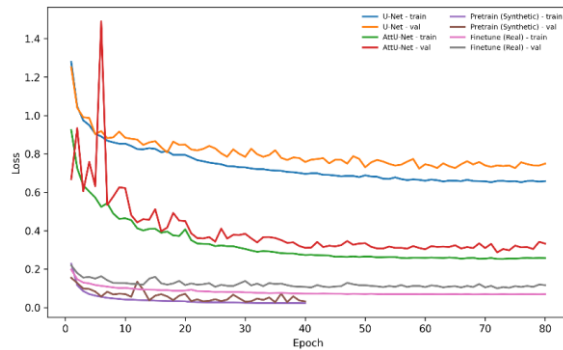
Penambahan mekanisme *attention* pada *Attention U-Net* menghasilkan peningkatan yang signifikan. *IoU* meningkat dari 0.2719 menjadi 0.5141 (peningkatan relatif ±89%), sedangkan *Dice* meningkat dari 0.3789 menjadi 0.6196 (peningkatan relatif ±63%). *Precision* dan *Recall* juga meningkat secara seimbang, menunjukkan bahwa model tidak hanya mendeteksi lebih banyak sinyal, tetapi juga mengurangi kesalahan deteksi palsu.

Peningkatan terbesar diperoleh pada skenario *Attention U-Net* dengan strategi *pretraining-fine-tuning*. Model ini mencapai *IoU* 0.7908 dan *Dice* 0.8786, dengan *Precision* 0.9017 serta *Recall* 0.8664. Dibandingkan *U-Net* standar, peningkatan relatif *IoU* mencapai lebih dari 190% dan *Dice* lebih dari 130%. Bahkan dibandingkan *Attention U-Net* tanpa *pretraining*, *IoU* masih meningkat sekitar 54% dan *Dice* sekitar 42%.

Secara keseluruhan, hasil ini menunjukkan bahwa peningkatan arsitektur dan strategi pelatihan memberikan kontribusi kumulatif terhadap kualitas segmentasi. Mekanisme *attention* membantu model

lebih selektif terhadap fitur sinyal, sementara tahap pretraining berbasis data sintetis memperkaya representasi awal sebelum adaptasi pada data riil. Kombinasi keduanya menghasilkan model yang lebih stabil, presisi, dan mampu beradaptasi terhadap variasi domain pada pemantauan spektrum pita lebar.

4.2. Analisa Dinamika Pelatihan



Gambar 3. Kurva loss *training* dan validasi

Gambar 3. memperlihatkan kurva *training* dan *validation loss* untuk ketiga skenario. Secara umum, seluruh model menunjukkan tren penurunan *loss* seiring bertambahnya *epoch*, yang mengindikasikan proses optimisasi berjalan stabil. Namun, terdapat perbedaan karakteristik konvergensi yang cukup jelas.

U-Net menunjukkan nilai *loss* tertinggi dengan penurunan yang relatif lambat serta jarak yang cukup besar antara kurva *training* dan *validation*. Pola ini mengindikasikan keterbatasan kapasitas model dalam menangkap struktur sinyal yang sangat tipis di tengah latar belakang yang heterogen.

Attention U-Net memperlihatkan *loss* yang lebih rendah dan kurva yang lebih stabil, menandakan peningkatan kualitas representasi fitur. Mekanisme *attention* membantu model memfokuskan aktivasi pada area yang relevan sehingga pembelajaran menjadi lebih efisien.

Pada tahap *pretraining* sintetis, *loss* menurun sangat cepat dan mencapai nilai rendah dalam beberapa epoch awal. Hal ini menunjukkan bahwa data sintetis yang dirancang berbasis distribusi riil mampu menyediakan sinyal pembelajaran yang konsisten. Ketika dilanjutkan ke *fine-tuning* pada data riil, model memulai pelatihan dari kondisi yang lebih baik dibanding inisialisasi acak, sehingga konvergensi berlangsung lebih cepat dan stabil.

Kurva ini secara empiris mendukung hipotesis bahwa *pretraining* memberikan inisialisasi bobot yang lebih informatif, sehingga mempercepat optimisasi pada domain target.

4.3. Interpretasi Hasil dan Implikasi

Hasil evaluasi menunjukkan peningkatan performa yang konsisten dari *U-Net* ke *Attention U-Net*, hingga konfigurasi *Attention U-Net* dengan strategi *pretraining-fine-tuning*. Temuan ini

menegaskan bahwa baik perbaikan arsitektur maupun strategi pelatihan berperan penting dalam meningkatkan kualitas segmentasi sinyal spektrogram.

U-Net standar menunjukkan keterbatasan dalam menangani sinyal yang sangat tipis, karena detail spasial mudah hilang selama proses ekstraksi fitur. Penambahan mekanisme *attention* pada *Attention U-Net* membantu model lebih fokus pada area sinyal dan menekan respons latar belakang, sehingga menghasilkan peningkatan yang jelas pada metrik *IoU*, *Dice*, *Precision*, dan *Recall*.

Peningkatan terbesar diperoleh melalui strategi *pretraining-fine-tuning*. Pelatihan awal pada dataset sintetis memungkinkan model mempelajari morfologi dasar sinyal sebelum dihadapkan pada kompleksitas data riil. Saat *fine-tuning*, model memulai dari bobot yang telah terinformasi dengan baik, sehingga konvergensi menjadi lebih cepat dan stabil. Konfigurasi ini mencapai *Dice* 0.8786 dan *IoU* 0.7908 pada data uji riil.

Secara keseluruhan, kombinasi *Attention U-Net* dan *pretraining* berbasis data sintetis efektif dalam meningkatkan akurasi, mengurangi *domain gap*, serta mempertahankan stabilitas pelatihan, khususnya pada kondisi sinyal tipis dan latar belakang heterogen yang menjadi ciri pemantauan spektrum frekuensi radio pita lebar.

5. KESIMPULAN DAN SARAN

Penelitian ini mengevaluasi pendekatan segmentasi berbasis *deep learning* untuk pelokalan sinyal pada spektrogram pemantauan spektrum frekuensi radio pita lebar dengan membandingkan *U-Net*, *Attention U-Net*, serta *Attention U-Net* dengan strategi *pretraining-fine-tuning*. Tantangan utama yang dihadapi mencakup dominasi sinyal berukuran sangat kecil ($\pm 1-3$ piksel), latar belakang yang heterogen, variasi perangkat pemantauan, serta kesenjangan distribusi antara data sintetis dan data riil. Hasil eksperimen menunjukkan bahwa *U-Net* standar kurang mampu mempertahankan detail spasial sinyal tipis, sementara penambahan mekanisme *attention* secara signifikan meningkatkan kemampuan model dalam memfokuskan fitur relevan dan menekan gangguan latar belakang. Performa terbaik diperoleh melalui strategi *pretraining* pada dataset sintetis yang diselaraskan secara statistik dengan data riil, kemudian dilanjutkan *fine-tuning* pada data riil, dengan capaian *Dice* 0.8786 dan *IoU* 0.7908 pada data uji riil. Temuan ini menegaskan bahwa kombinasi *Attention U-Net* dan *pretraining* sintetis efektif dalam meningkatkan akurasi, stabilitas pelatihan, dan generalisasi pada lingkungan pemantauan yang heterogen. Untuk kedepannya, penelitian dapat diarahkan pada integrasi estimasi parameter spektral seperti bandwidth dan level daya secara end-to-end, serta evaluasi pada domain operasional yang lebih beragam untuk menguji ketahanan model terhadap pergeseran distribusi yang lebih ekstrem.

DAFTAR PUSTAKA

- [1] H. N. Nguyen, M. Vomvas, T. D. Vo-Huu, dan G. Noubir, "WRIST: Wideband, Real-Time, Spectro-Temporal RF Identification System Using Deep Learning," *IEEE Trans. Mob. Comput.*, vol. 23, no. 2, hal. 1550–1567, 2024, doi: 10.1109/TMC.2023.3240971.
- [2] J. Su, J. Yao, R. Xie, dan K. Wu, "DeepSpectrum: A Deep-Learning-based Spectrum Identification for Wireless Signals," in *2023 19th International Conference on Mobility, Sensing and Networking (MSN)*, 2023, hal. 636–643. doi: 10.1109/MSN60784.2023.00094.
- [3] W. Li, K. Wang, L. You, dan Z. Huang, "A New Deep Learning Framework for HF Signal Detection in Wideband Spectrogram," *IEEE Signal Process. Lett.*, vol. 29, hal. 1342–1346, 2022, doi: 10.1109/LSP.2022.3179958.
- [4] N. West, T. Roy, dan T. O'Shea, "Wideband signal localization with spectral segmentation," *arXiv Prepr. arXiv2110.00583*, 2021.
- [5] K. N. R. S. V Prasad, K. B. Dsouza, V. K. Bhargava, S. Mallick, dan H. Boostanimehr, "A Deep Learning Framework for Blind Time-Frequency Localization in Wideband Systems," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, hal. 1–6. doi: 10.1109/VTC2020-Spring48590.2020.9128779.
- [6] T. Le Truong, N. T. Le, dan W. Benjapolakul, "An Application of Deep Learning YOLOv5 Framework to Intelligent Radio Spectrum Monitoring," in *2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Jul 2022, hal. 1–4. doi: 10.1109/ITC-CSCC55581.2022.9894862.
- [7] B. Li, S. Huang, dan G. Zhong, "LTEA-YOLO: An Improved YOLOv5s Model for Small Object Detection," *IEEE Access*, vol. 12, hal. 99768–99778, 2024, doi: 10.1109/ACCESS.2024.3429282.
- [8] S. Benazzouza, M. Ridouani, F. Salahdine, dan A. Hayar, "A novel spectrogram based lightweight deep learning for IoT spectrum monitoring," *Phys. Commun.*, vol. 64, hal. 102364, 2024, doi: <https://doi.org/10.1016/j.phycom.2024.102364>.
- [9] M. M. Alammam dan M. López-Benítez, "Signal area estimation based on deep learning," *Phys. Commun.*, vol. 59, hal. 102113, 2023, doi: <https://doi.org/10.1016/j.phycom.2023.102113>.
- [10] L. Cohen, *Time-frequency Analysis*. in Electrical engineering signal processing. Prentice Hall PTR, 1995. [Daring]. Tersedia pada: <https://books.google.co.id/books?id=CSKLQgAACAAJ>
- [11] Y. Guo, Y. Liu, T. Georgiou, dan M. S. Lew, "A review of semantic segmentation using deep neural networks," *Int. J. Multimed. Inf. Retr.*, vol. 7, no. 2, hal. 87–93, 2018, doi: 10.1007/s13735-017-0141-z.
- [12] J. Terven, D.-M. Córdova-Esparza, dan J.-A. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, hal. 1680–1716, 2023. doi: 10.3390/make5040083.
- [13] J. Redmon, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.
- [14] O. Ronneberger, P. Fischer, dan T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation BT - Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015," N. Navab, J. Hornegger, W. M. Wells, dan A. F. Frangi, Ed., Cham: Springer International Publishing, 2015, hal. 234–241.
- [15] Y. Wang *et al.*, "Attention-enhanced U-Net based network for cancerous tissue segmentation," *Biomed. Signal Process. Control*, vol. 106, hal. 107728, 2025, doi: <https://doi.org/10.1016/j.bspc.2025.107728>.
- [16] S. J. Pan dan Q. Yang, "A Survey on Transfer Learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, hal. 1345–1359, 2010, doi: 10.1109/TKDE.2009.191.
- [17] M. Wang dan W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, hal. 135–153, 2018, doi: <https://doi.org/10.1016/j.neucom.2018.05.083>.
- [18] G. Wilson dan D. J. Cook, "A survey of unsupervised deep domain adaptation," *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 5, hal. 1–46, 2020.
- [19] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, dan R. Webb, "Learning from Simulated and Unsupervised Images through Adversarial Training," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, hal. 2242–2251. doi: 10.1109/CVPR.2017.241.
- [20] A. Kar *et al.*, "Meta-Sim: Learning to Generate Synthetic Datasets," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, hal. 4550–4559. doi: 10.1109/ICCV.2019.00465.