

Perbandingan *Super-Resolution* dan CBAM untuk Optimasi Deteksi Objek Drone Termal

Helty Susilawati¹, Akhmad Fauzi Ikhsan¹, Firman¹, Arief Suryadi Satyawan², Chandra Rahmana¹

¹ Program Studi Teknik Elektro, Fakultas Teknik, Universitas Garut, Garut, Jawa Barat 44116, Indonesia

² Pusat Riset Telekomunikasi, Badan Riset dan Inovasi Nasional (BRIN), Bandung, Jawa Barat 40135, Indonesia

[Diserahkan: 27 Oktober 2025, Direvisi: 1 Februari 2026, Diterima: 16 Maret 2026]

Penulis Korespondensi: Helty Susilawati (email: helty.susilawati@uniga.ac.id)

INTISARI — Pendeteksian manusia menggunakan kamera termal sangat bermanfaat dalam kondisi tertentu, misalnya untuk mendeteksi orang yang tersesat di daerah pegunungan. Sulitnya medan wilayah pegunungan untuk dijelajahi menyebabkan operasi penyelamatan biasanya dilakukan dengan mengerahkan tim pencarian dan penyelamatan (*search and rescue*, SAR) ke lokasi kejadian. Namun, pendekatan ini tidak selalu efektif karena operasi SAR hanya dapat dilakukan dalam kondisi tertentu dan berpotensi membahayakan keselamatan tim SAR itu sendiri. Salah satu pendekatan alternatif teknik pencarian adalah penggunaan *drone* yang dilengkapi dengan kemampuan deteksi dan pengenalan manusia. Dalam konteks ini, kamera termal digunakan karena mampu menembus lingkungan yang menantang, sehingga cocok untuk operasi SAR. Metode deteksi objek yang digunakan dalam penelitian ini adalah YOLOv8. Penelitian ini bertujuan untuk membandingkan efektivitas antara integrasi ESRGAN dengan YOLOv8 dan penambahan *convolutional block attention module* (CBAM) pada bagian *neck* pada arsitektur *neck* YOLOv8. Perbandingan ini dilakukan untuk menentukan pendekatan yang memiliki kinerja paling baik pada aplikasi SAR. Kinerja ESRGAN dengan YOLOv8 dan CBAM dengan YOLOv8 dievaluasi menggunakan parameter presisi, *mean average precision* (mAP), dan *training loss*. Berdasarkan hasil eksperimen, ditemukan bahwa kombinasi ESRGAN dengan YOLOv8 menunjukkan kinerja yang lebih baik dibandingkan dengan modifikasi berbasis CBAM. Hal ini ditunjukkan oleh nilai presisi dan mAP yang lebih tinggi, serta nilai *training loss* yang lebih rendah pada kerangka deteksi YOLOv8 yang ditingkatkan dengan ESRGAN. Temuan hasil eksperimen menunjukkan bahwa kualitas citra dengan menggunakan ESRGAN lebih efektif daripada modifikasi dengan menggunakan CBAM untuk meningkatkan kinerja deteksi manusia berbasis citra termal untuk SAR.

KATA KUNCI — SAR, YOLOv8, *Super-Resolution*, CBAM, Deteksi Objek.

I. PENDAHULUAN

Kegiatan pendakian gunung memiliki risiko kecelakaan atau kondisi darurat, seperti pendaki yang tersesat atau dinyatakan hilang. Salah satu contoh terbaru adalah meningkatnya kasus orang yang hilang di kawasan pegunungan. Laporan terkini mencatat adanya kasus pendaki yang dilaporkan hilang di Gunung Binaya [1] dan Gunung Buthak [2]. Selama proses pencarian dan penyelamatan (*search and rescue*, SAR) terhadap pendaki yang hilang di daerah pegunungan, berbagai tantangan sering dihadapi, seperti medan yang sulit, cuaca ekstrem, jarak pandang terbatas akibat kabut atau kondisi malam hari, serta keterbatasan jumlah personel penyelamat. Kondisi ini sering menyebabkan keterlambatan proses evakuasi karena pelaksanaan operasi SAR juga dapat membahayakan keselamatan tim SAR itu sendiri. Untuk mengatasi permasalahan tersebut, diperlukan wahana udara tanpa awak (*unmanned aerial vehicle*, UAV) yang tidak hanya berfungsi untuk melakukan survei wilayah terdampak, tetapi juga mampu membantu membangun koneksi komunikasi nirkabel yang penting antara korban dan infrastruktur seluler terdekat yang tersedia [3]. Sistem robotik yang dikendalikan oleh kecerdasan buatan (*artificial intelligence*, AI) kini mulai berkembang untuk mengatasi berbagai tantangan dalam operasi SAR yang bertujuan menyelamatkan nyawa di wilayah bencana, yang sering kali mengalami kesulitan dalam implementasinya akibat lingkungan berbahaya dan karakteristik medan yang kompleks [4].

Salah satu teknologi yang dapat digunakan untuk mendukung tim SAR adalah *drone* yang dilengkapi dengan

kamera termal sebagai alat pencarian. Kamera termal telah banyak digunakan dalam berbagai aplikasi, antara lain pada *drone* untuk survei geofisika jarak dekat guna mendeteksi anomali permukaan dan bawah permukaan [5]; memprediksi dan memantau pertumbuhan tanaman secara akurat melalui pengukuran suhu permukaan tanaman yang berkorelasi dengan kandungan air [6]; sebagai sensor utama dalam mendeteksi, memetakan, dan melokalisasi kebakaran lahan gambut [7]; serta dalam estimasi populasi rusa sika Formosa secara efisien dan akurat melalui pemantauan udara berbasis *drone* [8].

Beberapa penelitian terkait penggunaan *drone* dan kamera termal dalam proses SAR meliputi pemanfaatan kamera termal sebagai sensor utama untuk mendeteksi dan melacak manusia pada misi SAR berbasis *drone* menggunakan YOLOv5, dengan parameter evaluasi berupa *total track life* (TTL), *mean track life* (MTL), dan *track purities* (TP) [9]. Selain itu, penelitian lain juga mengkaji penggunaan *drone* multirotor, dengan kamera termal yang terbukti sangat efektif untuk menemukan orang hilang dalam kondisi sulit seperti kegelapan dan kabut atau adanya hambatan visual. Dengan penerapan metode *track segment association* (TSA), akurasi pelacakan dilaporkan meningkat secara signifikan hingga lebih dari 200% [10]. Penelitian lain mengenai peningkatan deteksi manusia menggunakan kamera termal dan *drone* juga dilakukan dengan memanfaatkan YOLOv8m dan YOLOv5s, dengan parameter perbandingan berupa presisi, *recall*, *mean average precision* (mAP)@0,5, dan mAP@0,5–0,95 [11]. Selain itu, terdapat pula penelitian mengenai deteksi manusia menggunakan kamera termal pada *drone* dengan pendekatan YOLOv3 [12].

Kamera termal digunakan karena memiliki berbagai keunggulan dibandingkan kamera visual konvensional, terutama kemampuannya dalam mendeteksi objek pada kondisi gelap serta mengukur perbedaan suhu [13]. Banyak penelitian memanfaatkan *dataset* citra termal yang diambil menggunakan *drone* dalam berbagai kondisi lingkungan, seperti area terbuka maupun kawasan perkotaan. Meskipun tidak secara langsung merepresentasikan konteks pegunungan, *dataset* tersebut tetap relevan karena menyediakan citra termal manusia dari sudut pandang udara yang sesuai dengan skenario SAR.

Penggunaan *drone* untuk deteksi manusia di wilayah pegunungan dapat menjadi sangat efektif apabila sistem mampu mengenali objek dengan baik. Namun, permasalahan yang sering muncul pada citra hasil tangkapan *drone* adalah rendahnya resolusi citra, sehingga kemampuan model dalam mengekstraksi fitur-fitur penting menjadi terbatas. Penelitian ini memanfaatkan *dataset* citra termal *drone* yang tersedia untuk pengembangan, pelatihan, dan pengujian model, yang selanjutnya digunakan untuk membandingkan pendekatan prapemrosesan menggunakan *super-resolution* (SR) dengan pendekatan modifikasi arsitektur model menggunakan *convolutional block attention module* (CBAM).

SR dipilih karena mampu meningkatkan ketelitian (presisi) dalam mendeteksi objek berukuran sangat kecil [14]. Penggunaan SR tanpa dilakukannya modifikasi arsitektur model merupakan pendekatan yang sederhana dan efisien secara komputasi, tetapi tetap dapat meningkatkan tingkat deteksi [15]. Penerapan SR pada sistem deteksi objek, khususnya untuk mendeteksi objek kecil seperti kendaraan pada citra udara dan lalu lintas, dilaporkan mampu meningkatkan kinerja secara signifikan [16] serta memperbaiki kualitas visual objek pada video beresolusi rendah [17]. SR merupakan teknik yang tidak hanya memperbesar ukuran citra, tetapi juga menambahkan detail baru yang dipelajari dari data pelatihan, seperti tekstur, tepi objek, dan pola-pola halus. Teknik SR yang digunakan dalam penelitian ini adalah *enhanced super-resolution generative adversarial network* (ESRGAN), yang kemudian diintegrasikan dengan YOLOv8.

CBAM merupakan mekanisme atensi yang ditambahkan pada *convolutional neural network* (CNN) untuk membantu model memfokuskan perhatian pada bagian fitur citra yang paling penting. CBAM bekerja dengan menerapkan dua jenis atensi, yaitu atensi kanal (*channel attention*) dan atensi spasial (*spatial attention*). CBAM dipilih karena mampu mengintegrasikan informasi kanal dan spasial untuk meningkatkan fokus model terhadap konteks fitur dan lokasi objek [18]. Selain itu, CBAM juga digunakan untuk meningkatkan kemampuan diskriminasi fitur melalui mekanisme atensi kanal dan spasial [19]. CBAM telah banyak diterapkan pada citra termal, antara lain untuk mendeteksi kebocoran pada permukaan bendungan [20] serta untuk meningkatkan representasi fitur asap pada bagian *backbone* jaringan [21]. Dalam penelitian ini, CBAM diterapkan pada *backbone* YOLOv8 dengan tujuan menghasilkan fitur yang lebih bersih dan informatif sebelum masuk ke bagian *neck* dan *head*, sehingga bagian *head* tidak perlu lagi memproses derau, tetapi dapat langsung memanfaatkan fitur-fitur penting untuk proses deteksi. Setelah penerapan CBAM, modul *spatial pyramid pooling-fast* (SPPF) juga digunakan untuk memperkaya informasi multiskala sebelum diteruskan ke bagian *head* (*neck* dan deteksi).

Penelitian ini menyajikan perspektif perbandingan strategis yang baru serta bukti empiris terkontrol yang memberikan

wawasan mengenai efektivitas relatif dari dua pendekatan dalam skenario UAV. Pendekatan pertama adalah kombinasi ESRGAN dengan YOLOv8, yang diharapkan mampu menghasilkan parameter deteksi objek yang lebih tinggi melalui metode SR. Sementara itu, pendekatan kedua adalah penerapan CBAM pada *backbone* YOLOv8, dengan harapan dapat meningkatkan kualitas atensi dalam deteksi objek tanpa mengubah data masukan. Berbeda dengan penelitian sebelumnya yang umumnya mengevaluasi kedua pendekatan tersebut secara terpisah, penelitian ini secara langsung mengeksplorasi pengaruh tingkat intervensi yang berbeda terhadap kinerja deteksi objek berukuran kecil pada citra *drone* beresolusi rendah. Dari sisi kontribusi ilmiah, penelitian ini membedakan dua paradigma peningkatan kinerja yang berbeda pada deteksi objek berbasis UAV menggunakan YOLOv8, yaitu peningkatan citra pada tingkat data melalui SR dan penyempurnaan fitur pada tingkat model melalui mekanisme atensi, dengan menggunakan *dataset* publik MONET. Sementara itu, dampak praktis dari penelitian ini terletak pada potensi penerapan model deteksi berbasis ESRGAN-YOLOv8 untuk misi SAR di wilayah pegunungan. Dengan kemampuan deteksi manusia yang ditingkatkan melalui kamera termal, sistem ini diharapkan dapat mempercepat proses identifikasi korban, meningkatkan keselamatan tim SAR, serta secara signifikan mengurangi waktu pencarian.

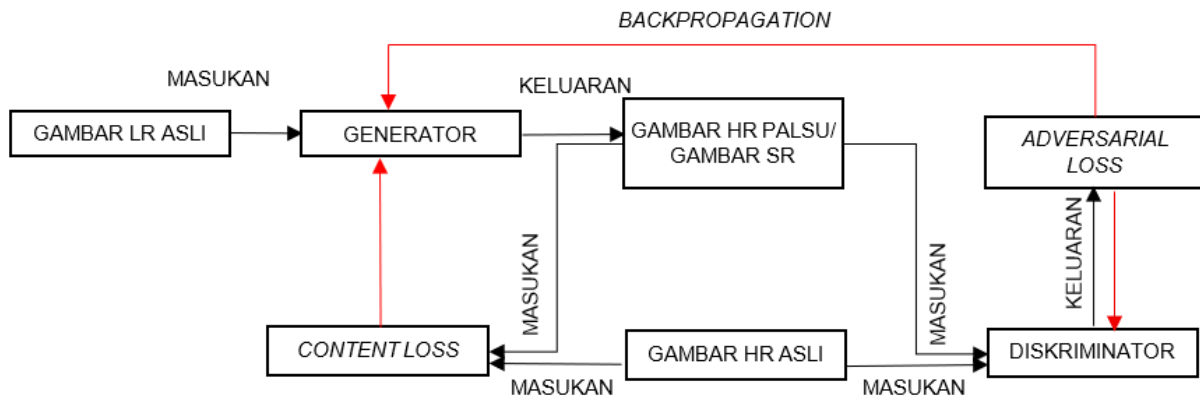
II. PENELITIAN TERKAIT

A. THERMAL IMAGING UNTUK SAR

Kemampuan pendeteksian emisi inframerah tubuh manusia telah mendorong penggunaan kamera termal secara luas dalam misi SAR karena kamera termal dapat digunakan bahkan pada kondisi cahaya rendah, berkabut, atau di area hutan lebat. Hal ini berbeda dengan kamera *red, green, blue* (RGB) yang sangat bergantung pada pencahayaan. Kamera termal mampu memberikan kontras yang tinggi antara objek hidup dan lingkungan sekitarnya. Beberapa penelitian yang menggunakan kamera termal untuk misi SAR telah membandingkan YOLOv7-Tiny dengan YOLOv8x untuk mengevaluasi kinerja model dan menentukan model yang paling sesuai untuk SAR. Hasil penelitian menunjukkan bahwa YOLOv8x mencapai nilai *average precision* (AP) sebesar 95%, sedangkan varian terkecilnya, YOLOv8n, masih mampu mempertahankan akurasi sebesar 91%. Sementara itu, YOLOv7-Tiny menunjukkan kinerja yang setara dengan YOLOv8n, tetapi memiliki ukuran model yang lebih ringan (48% lebih kecil) [22]. Penelitian lain memanfaatkan *dataset* citra termal yang diambil secara langsung menggunakan *drone* sebagai sistem deteksi korban pada operasi SAR pascabencana, yang dilatih menggunakan YOLOv3 dan diuji pada perangkat NVIDIA Jetson TX2 dengan kinerja *real-time* sebesar 26,6 FPS [23]. Selain itu, terdapat pula penelitian yang menggunakan RTMDet, PP-YOLOE+, YOLOv5s, YOLOv8s, dan DINO untuk SAR, dengan memanfaatkan *dataset* UAV berbasis kamera termal inframerah bernama *partially occluded person* (POP). *Dataset* tersebut menghasilkan nilai AP yang stabil dan akurasi yang tinggi hingga tingkat oklusi sebesar 70% [24].

B. DETEKSI OBJEK PADA CITRA TERMAL

Berbagai penelitian telah dilakukan terkait deteksi objek pada citra termal, termasuk penggunaan metode gabungan *deblur-SRRGAN* dan Mask R-CNN, yang berhasil meningkatkan kualitas citra termal serta akurasi deteksi objek dan refleksi termal dibandingkan metode lain, meskipun masih



Gambar 1. Alur SRGAN.

memiliki keterbatasan dalam kecepatan pemrosesan untuk aplikasi *real-time* [25]. Selain itu, penelitian mengenai sistem persepsi otomatis berbasis multisensor, yaitu kamera termal dan *light detection and ranging* (LiDAR) untuk pencarian korban di laut dengan menggunakan YOLOv8 menunjukkan kinerja yang sangat baik, dengan nilai presisi sebesar 93,5% dan *recall* sebesar 94,2% [26].

C. ATTENTION MECHANISMS PADA DEEP LEARNING

Attention mechanism dapat didefinisikan sebagai suatu proses yang memetakan sebuah vektor kueri dan sekumpulan pasangan vektor kunci-nilai menjadi satu vektor keluaran [27]. Salah satu penelitian terkait mekanisme atensi adalah penggunaan *channel-wise attention mechanism* yang dikombinasikan dengan *generative adversarial network* (GAN) dan ResNet, yang menekankan fitur-fitur penting pada dimensi kanal untuk memfokuskan perhatian pada karakteristik wajah yang krusial [28]. Hasil penelitian tersebut menunjukkan nilai presisi sebesar 0,79, *recall* sebesar 0,88, *F1-score* sebesar 0,83, akurasi sebesar 0,83, serta *Receiver operating characteristic - area under curve* (ROC-AUC) sebesar 0,825. Penelitian selanjutnya mengkaji penggunaan Efficient-ViT dengan penambahan mekanisme atensi untuk deteksi dini osteoarthritis lutut, dengan membandingkan hasil presisi, *recall*, *F1-score*, dan mAP dari YOLOv5n, YOLOv8n, YOLOv9n, dan YOLOv8-ViT. Hasil penelitian menunjukkan bahwa YOLOv8-ViT menghasilkan nilai kinerja tertinggi [29].

D. YOLOV8 UNTUK DETEKSI OBJEK

Penelitian mengenai penggunaan YOLOv8 telah banyak dilakukan, termasuk penerapannya untuk deteksi objek berukuran kecil [30], [31], deteksi rambu lalu lintas [32], serta deteksi potensi longsor [33]. Keempat penelitian tersebut menunjukkan bahwa penggunaan YOLOv8 dalam deteksi objek mampu meningkatkan akurasi secara signifikan.

E. SUPER-RESOLUTION UNTUK DETEKSI OBJEK KECIL

Penelitian terkait SR digunakan, antara lain, untuk meningkatkan resolusi dan tekstur objek nanas berukuran kecil, yang menunjukkan bahwa pengambilan citra menggunakan *drone* menghasilkan peningkatan nilai presisi sebesar 17% [34]. Penelitian lain menggunakan model CSRGAN dengan pendekatan *classification-oriented super-resolution*, yang secara signifikan meningkatkan kualitas deteksi objek kecil pada citra UAV [35]. Dari sudut pandang persepsi, sebuah penelitian menunjukkan bahwa rekonstruksi *super-resolution generative adversarial network* (SRGAN) dengan faktor pembesaran yang tinggi menghasilkan citra yang empat kali lebih fotorealistik [36].

III. METODOLOGI

Penelitian ini menggunakan dua pendekatan untuk meningkatkan kinerja deteksi objek pada citra termal yang diambil menggunakan *drone*. Kedua pendekatan tersebut kemudian dibandingkan untuk menentukan pendekatan yang menghasilkan nilai parameter terbaik. Pendekatan pertama adalah penerapan SR pada tahap prapemrosesan. Pendekatan kedua adalah modifikasi arsitektur model dengan menambahkan CBAM pada bagian *backbone*. *Dataset* yang digunakan dalam penelitian ini adalah *dataset* MONET [37], yang terdiri atas 7.610 citra termal untuk proses pelatihan, 1.676 citra termal untuk proses validasi, dan 3.443 citra termal untuk proses pengujian.

Pendekatan SR dipilih karena citra yang dihasilkan oleh *drone* umumnya memiliki ukuran yang relatif kecil. Dalam proses deteksi objek, citra berukuran kecil memberikan kesulitan tersendiri. Oleh karena itu, diperlukan suatu metode untuk memperbesar ukuran citra tanpa menghilangkan detail penting, salah satunya melalui penggunaan SR. Teknik SR yang digunakan dalam penelitian ini adalah ESRGAN. ESRGAN merupakan pengembangan dari SRGAN yang masih mengikuti alur kerja yang sama dengan SRGAN. Alur kerja ESRGAN ditunjukkan pada Gambar 1.

Gambar 1 menunjukkan bahwa proses SRGAN menggunakan dua komponen utama, yaitu generator dan diskriminator. SR merujuk pada proses menghasilkan citra beresolusi tinggi (*high resolution*, HR) dari citra beresolusi rendah (*low resolution*, LR). Pada SRGAN, citra LR digunakan sebagai masukan bagi generator. Selanjutnya, *discriminator* membandingkan keluaran generator dengan citra HR asli dan nilai *loss* dihitung untuk menyesuaikan bobot dan *bias* jaringan. Jaringan generator secara berkelanjutan berupaya menghasilkan citra LR yang makin mendekati citra HR sebenarnya. Sebaliknya, jaringan *discriminator* menghasilkan probabilitas diskriminasi, yang digunakan untuk mengevaluasi akurasi dengan membedakan antara citra hasil generasi dan citra HR asli.

Meskipun alur kerja SRGAN dan ESRGAN relatif serupa karena merupakan bagian dari GAN dan teknik SR, terdapat perbedaan di antara keduanya. Perbedaan utama antara SRGAN dan ESRGAN terletak pada penggunaan *residual-in-residual dense blocks* (RRDB) pada bagian generator serta penghilangan *batch normalization* pada RRDB. Pendekatan ini mampu meningkatkan kualitas berkaitan dengan persepsi, sehingga menghasilkan kualitas visual yang lebih baik secara konsisten dengan tekstur yang lebih realistis dan alami dibandingkan dengan SRGAN [38].

Selain menggunakan ESRGAN, penelitian ini juga memanfaatkan CBAM sebagai pendekatan untuk mengevaluasi kinerja deteksi objek pada citra kamera termal. CBAM merupakan blok atensi yang menerapkan dua tahap berurutan, yaitu atensi kanal dan diikuti oleh atensi spasial. Atensi kanal berfungsi untuk menghitung bobot setiap kanal fitur yang merepresentasikan informasi paling penting (*what*), sedangkan atensi spasial menghasilkan peta lokasi penting pada fitur yang merepresentasikan posisi informasi krusial (*where*).

Dalam penelitian ini, CBAM ditempatkan pada bagian *backbone* YOLOv8, tepatnya pada tingkat peta fitur terdalam setelah blok C2f, dengan jumlah kanal sebanyak 1.024. Selanjutnya, CBAM diintegrasikan ke dalam arsitektur YOLOv8. Arsitektur CBAM pada YOLOv8 ditunjukkan pada Gambar 2. Gambar tersebut menunjukkan bahwa algoritma YOLOv8 terdiri atas empat bagian utama, yaitu masukan, *backbone*, *neck*, dan keluaran. Modul C2f pada YOLOv8 memiliki jumlah parameter yang lebih sedikit serta kemampuan ekstraksi fitur yang lebih unggul. Pendekatan ini mempertahankan karakteristik model yang ringan sekaligus mampu menangkap aliran gradien yang lebih kaya, sehingga secara signifikan meningkatkan kinerja model.

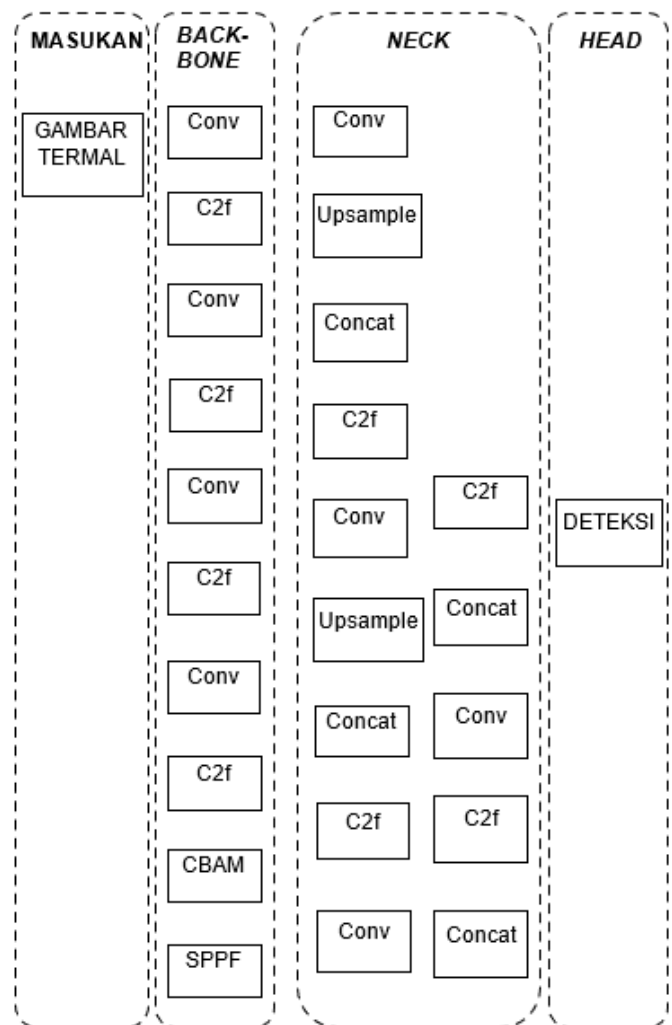
Penelitian ini bertujuan untuk membandingkan hasil penggunaan ESRGAN dan CBAM pada arsitektur YOLOv8 untuk citra termal. Alur penelitian dimulai dengan melakukan deteksi objek menggunakan citra termal asli dengan YOLOv8. Tahap selanjutnya adalah menerapkan SR menggunakan ESRGAN. Data yang digunakan pada proses SR adalah data citra pelatihan, sedangkan data validasi dan pengujian tidak dikenai proses SR agar sistem tetap mampu mengenali citra asli. Setelah proses SR dilakukan, deteksi objek kembali dilakukan menggunakan YOLOv8. Objek yang dideteksi meliputi manusia, kendaraan, dan hewan. Setelah proses deteksi oleh YOLOv8, tahap berikutnya adalah menghitung nilai *training loss*, *validation loss*, serta parameter presisi, *recall*, dan mAP.

Setelah proses SR selesai, tahap berikutnya adalah melakukan modifikasi model pada bagian *backbone* dan *neck* dengan menggunakan CBAM. Model yang telah dimodifikasi kemudian diterapkan pada YOLOv8 untuk proses deteksi objek. Sama seperti pada pendekatan SR, tahap evaluasi dilakukan dengan menghitung nilai *training loss*, *validation loss*, serta parameter presisi, *recall*, dan mAP.

IV. HASIL DAN PEMBAHASAN

Data loss yang dianalisis dalam penelitian ini meliputi *training loss* dan *validation loss*. *Training loss* terdiri atas *box loss*, *class loss*, dan *distribution focal loss* (DFL), sedangkan *validation loss* terdiri atas *val box loss*, *val class loss*, dan *val DFL loss*. *Box loss* bertujuan untuk mengukur tingkat keberhasilan prediksi *bounding box* (x, y, w, h) dibandingkan dengan *ground truth*. *Class loss* digunakan untuk mengukur kemampuan model dalam mengklasifikasikan objek dengan benar. *DFL loss* merupakan komponen khusus pada YOLOv7/YOLOv8 yang bertujuan untuk meningkatkan ketepatan prediksi koordinat *bounding box*.

Secara konseptual, *box loss* lebih berfokus pada lokasi dan ukuran *bounding box* atau kualitas *bounding box*, *class loss* berfokus pada kualitas prediksi kelas objek, sedangkan *DFL loss* berfokus pada presisi distribusi koordinat *bounding box* atau akurasi posisi objek. Dalam proses deteksi objek, penelitian ini menggunakan konfigurasi *default* YOLOv8,



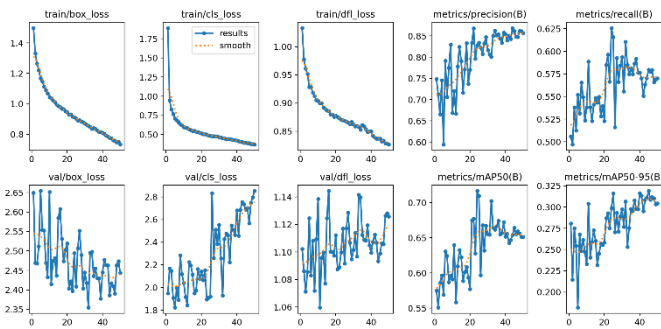
Gambar 2. Arsitektur CBAM-YOLOv8.

yaitu *optimizer* bernama *stochastic gradient descent* (SGD) dengan laju belajar awal sebesar 0,001, *final LR ratio* sebesar 0,1, *scheduler* berupa *cosine decay*, *warmup* sebanyak 5 *epoch*, *batch size* sebesar 4, serta jumlah *epoch* sebanyak 50, yang diterapkan pada data asli, data hasil SR, dan data dengan modifikasi CBAM.

A. DETEKSI OBJEK MENGGUNAKAN DATA ASLI DENGAN YOLOv8

Eksperimen pertama dilakukan menggunakan *dataset* MONET untuk memperoleh nilai *training loss*, *validation loss*, presisi, *recall*, dan mAP dengan menggunakan YOLOv8.

Gambar 3 menunjukkan hasil parameter yang diperoleh menggunakan YOLOv8. Berdasarkan hasil penelitian, nilai akurasi (presisi vs *recall*) yang diperoleh tergolong tinggi (mendekati 0,869), yang menunjukkan bahwa model mampu mengurangi jumlah *false positive*. Nilai *recall* yang diperoleh lebih rendah (maksimum 0,626), yang mengindikasikan bahwa model masih melewatkan beberapa objek (*false negative* relatif tinggi). Nilai mAP@0,5 tergolong cukup baik (0,717), tetapi nilai mAP@0,5–0,95 lebih rendah (0,319), yang menunjukkan bahwa model mengalami kesulitan dalam mempertahankan kinerja pada berbagai skala *intersection over union* (IoU). Hal ini juga dapat diinterpretasikan sebagai ketidakkonsistenan akurasi *bounding box*. Nilai *box loss* dan *DFL loss* relatif stabil, sedangkan nilai *validation class loss* tergolong cukup tinggi, yang mengindikasikan bahwa model masih mengalami



Gambar 3. Hasil parameter menggunakan data asli dengan YOLOv8.

kesulitan dalam membedakan kelas-kelas objek. Selain itu, laju belajar mencapai nilai maksimum sebesar 0,0012, kemudian menurun secara konsisten untuk menjaga stabilitas proses pelatihan.

B. DETEKSI OBJEK MENGGUNAKAN ESRGAN

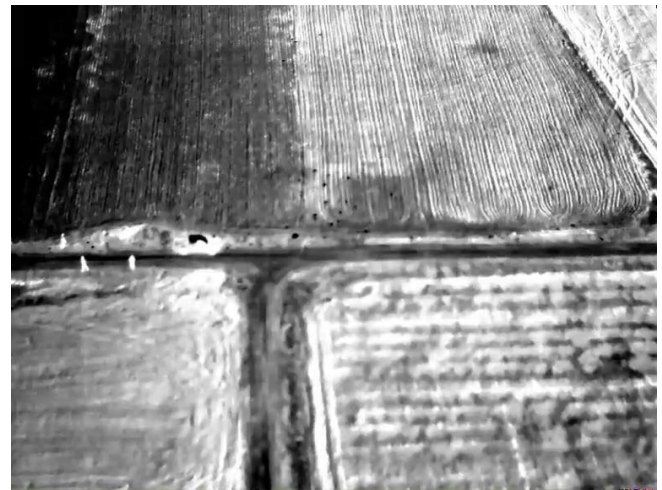
Arsitektur jaringan *generator* yang digunakan dalam ESRGAN adalah *residual-in-residual dense block network* (RRDBNet). Dalam penelitian ini, parameter yang digunakan adalah citra masukan dan keluaran yang memiliki tiga kanal (RGB), dengan jumlah awal peta fitur (*channel width*) sebesar 64. Arsitektur jaringan menunjukkan bahwa terdapat 23 blok RRDB yang tersusun secara berurutan, dengan setiap *dense block* dalam RRDB menambahkan 32 kanal fitur baru. Penelitian ini menggunakan ESRGAN untuk proses SR dengan faktor skala sebesar 4. Hal ini menyebabkan citra hasil ESRGAN memiliki jumlah piksel empat kali lebih besar dibandingkan citra asli. *Dataset* yang digunakan terdiri atas citra dengan resolusi 800×600 piksel dan hasil pemrosesan menggunakan ESRGAN menghasilkan citra dengan resolusi 3.200×2.400 piksel. Gambar 4 dan Gambar 5 menampilkan perbandingan antara citra asli dan citra hasil ESRGAN.

Gambar 5 menunjukkan bahwa pola-pola pada citra hasil ESRGAN tampak jauh lebih tajam dibandingkan dengan citra asli pada Gambar 4. Sebagai contoh, garis-garis pola persawahan terlihat lebih halus dan lebih terdefinisi dengan baik dibandingkan dengan hasil interpolasi sederhana. Selain itu, batas-batas objek seperti tepi lahan dan kendaraan (ditunjukkan sebagai titik-titik putih di jalan) tampak lebih jelas dan terpisah dari latar belakang. Berdasarkan Gambar 5, juga dapat diamati bahwa pada area pertanian, ESRGAN mampu memperjelas pola tanam, yang ditunjukkan oleh garis-garis vertikal dan horizontal yang lebih kontinu dan terstruktur. Detail-detail halus ini mendekati kualitas fotorealistik, sedangkan pada citra asli (Gambar 4) detail tersebut masih tampak lebih buram.

Gambar 6 menunjukkan hasil parameter yang diperoleh menggunakan data pelatihan SR dengan YOLOv8. Berdasarkan hasil penelitian, nilai akurasi (presisi vs *recall*) dan mAP pada citra yang telah melalui proses SR hampir sama dengan citra asli, tetapi nilai *validation loss* secara signifikan lebih tinggi, yang mengindikasikan adanya potensi *overfitting*. Laju belajar yang digunakan juga terlihat tidak stabil, yang ditunjukkan oleh meningkatnya nilai *training loss* ketika laju belajar bertambah. Padahal, pada kondisi laju belajar yang stabil dan sesuai, nilai *training loss* seharusnya menurun atau setidaknya tetap konstan.

C. DETEKSI OBJEK MENGGUNAKAN CBAM

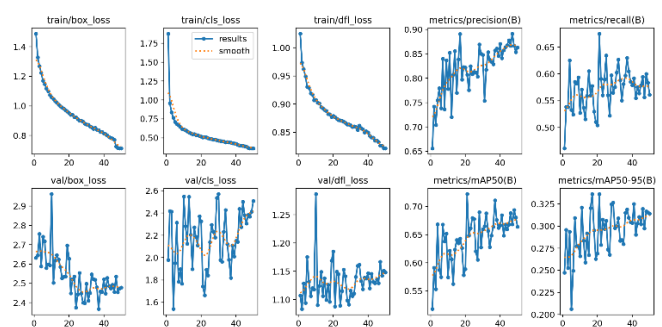
Pelatihan menggunakan YOLOv8 dengan modifikasi CBAM dilakukan pada *dataset* asli. Modifikasi CBAM



Gambar 4. Gambar asli dari basis data.



Gambar 5. Gambar hasil ESRGAN.



Gambar 6. Hasil parameter menggunakan ESRGAN dengan YOLOv8.

diterapkan pada bagian *backbone*, yang kemudian diintegrasikan ke dalam arsitektur YOLOv8. Dalam penelitian ini digunakan dua modul utama, yaitu atensi kanal dan atensi spasial. Atensi kanal berfokus pada penentuan informasi penting (*what*) pada setiap kanal peta fitur, sedangkan atensi spasial berfokus pada penentuan lokasi (*where*) fitur-fitur penting setelah penerapan atensi kanal.

Tahapan dalam atensi kanal meliputi *global pooling*, *multi-layer perceptron* (MLP), dan *combination and sigmoid*. Pada tahap *global pooling*, fitur masukan x dengan dimensi $[B,C,H,W]$ diproses menggunakan *global average pooling* dan *global max pooling* pada setiap kanal. Proses ini dilakukan melalui *AdaptiveAvgPool2d* dan *AdaptiveMaxPool2d*, sehingga dihasilkan dua vektor deskriptor kanal dengan dimensi $[B,C,1,1]$. Pada tahap MLP (bottleneck), setiap

deskriptor kanal diteruskan ke jaringan *multi-layer perceptron* yang diimplementasikan menggunakan dua konvolusi 1×1 , yaitu $fc1$ dan $fc2$, dengan lapisan *ReLU* di antara keduanya. Proses ini bertujuan untuk mereduksi dimensi fitur kanal ke ukuran yang lebih kecil sesuai dengan rasio *bottleneck*, kemudian mengembalikannya ke ukuran semula. Keluaran dari jalur *average pooling* dan *max pooling* dijumlahkan, kemudian diterapkan fungsi *sigmoid* untuk membentuk peta atensi kanal $M_c(F)$ dengan ukuran $[B,C,1,1]$ dan nilai berada pada rentang $[0,1]$. Secara matematis, kombinasi ini dinyatakan dalam (1) [39].

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))). \quad (1)$$

Peta atensi kanal yang dihasilkan kemudian dikalikan secara *element-wise* dengan peta fitur awal x . Dengan demikian, kanal-kanal fitur yang dianggap paling informatif akan diberi bobot yang lebih tinggi, sedangkan kanal yang kurang penting akan ditekan.

Dengan melalui keempat tahapan tersebut, modul atensi kanal menghasilkan peta atensi per kanal yang memperkuat kanal-kanal fitur yang relevan. Tahap selanjutnya adalah penerapan atensi spasial, dengan tahapan sebagai berikut: *channel pooling*, *channel combination and convolution*, serta *sigmoid and combination*.

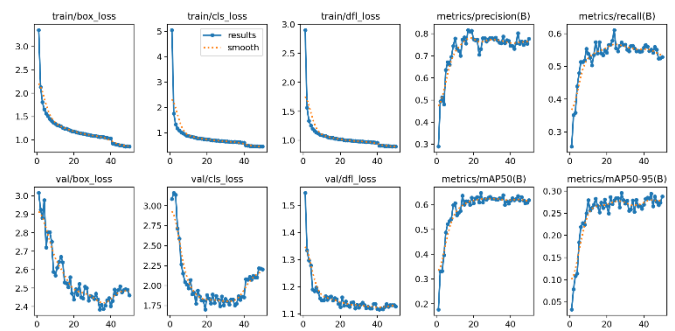
Berdasarkan masukan pada atensi spasial, dilakukan penghitungan nilai rata-rata (*average*) dan nilai maksimum (*maximum*) pada dimensi kanal. Proses ini menghasilkan dua peta 2D berukuran $[B,1,H,W]$ yang merepresentasikan ringkasan informasi spasial, masing-masing berupa peta rata-rata kanal dan peta maksimum kanal. Selanjutnya, pada *channel combination and convolution*, kedua peta 2D yang dihasilkan kemudian digabungkan pada dimensi kanal, sehingga membentuk tensor berukuran $[B,2,H,W]$. Tensor dua kanal ini selanjutnya diproses menggunakan satu lapisan konvolusi 2D berukuran 7×7 , yang menghasilkan peta satu kanal berukuran $[B,1,H,W]$. Proses konvolusi ini berfungsi untuk mengidentifikasi pola spasial pada fitur kanal yang telah digabungkan. Dalam *sigmoid and combination*, keluaran dari proses konvolusi kemudian dilewatkan melalui fungsi *sigmoid* untuk menghasilkan peta atensi spasial $M_s(F)$ dengan ukuran $[B,1,H,W]$ dan nilai berada pada rentang $[0,1]$. Peta yang dihasilkan menunjukkan lokasi piksel yang paling penting. Selanjutnya, peta atensi spasial diterapkan dengan mengalikan secara *element-wise* peta fitur awal dengan masker atensi tersebut, sehingga area spasial yang informatif menjadi lebih diperkuat. Proses ini sesuai dengan formulasi atensi spasial yang dinyatakan pada (2) [39].

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])). \quad (2)$$

Nilai $f^{7 \times 7}$ adalah operasi konvolusi 7×7 pada peta gabungan. Dengan demikian, atensi spasial menyoroti lokasi-lokasi pada peta fitur yang mengandung informasi paling penting. Berdasarkan atensi kanal dan atensi spasial yang telah diterapkan, keduanya dikombinasikan menggunakan (3).

$$F' = M_c(F) \otimes F, F'' = M_s(F') \otimes F'. \quad (3)$$

Gambar 7 menunjukkan hasil parameter yang diperoleh menggunakan CBAM. Berdasarkan hasil penelitian, nilai akurasi (*presisi vs recall*) menunjukkan bahwa nilai presisi tertinggi yang dicapai adalah sebesar 0.81, yang mengindikasikan bahwa model cukup baik dalam membedakan objek yang valid dari *false positive*. Nilai *recall* memiliki rata-



Gambar 7. Hasil parameter menggunakan CBAM dengan YOLOv8.

rata sebesar 0,53, yang menunjukkan bahwa model belum sepenuhnya mampu menangkap seluruh objek berukuran kecil atau bertekstur halus. Nilai mAP memiliki rata-rata sebesar 0,58, yang tergolong cukup baik untuk model dengan penerapan CBAM. Namun, nilai mAP@0,5–0,95 hanya mencapai rata-rata sebesar 0,25, yang menunjukkan adanya kesenjangan yang cukup signifikan dibandingkan dengan nilai mAP@0,5. Hal ini mengindikasikan bahwa model masih mengalami kesulitan dalam mempertahankan kinerja pada berbagai tingkat IoU.

Nilai *validation box loss* masih lebih tinggi dibandingkan dengan *training loss*, tetapi bersifat stabil, yang menunjukkan bahwa tidak terjadi *overfitting* yang signifikan. Nilai *validation class loss* mengalami fluktuasi ringan, tetapi menunjukkan tren penurunan dibandingkan pada tahap awal pelatihan. Sementara itu, nilai *validation DFL loss* sangat stabil. Berdasarkan hasil penelitian, laju belajar yang tinggi pada awal proses pelatihan terbukti efektif dalam menurunkan nilai *loss* secara signifikan pada tahap awal. Selanjutnya, penurunan laju belajar secara bertahap memungkinkan model mencapai kinerja puncak, yang ditunjukkan oleh tercapainya nilai presisi, *recall*, dan mAP tertinggi pada pertengahan proses pelatihan. Pada tahap akhir pelatihan, laju belajar yang sangat rendah membuat model menjadi stabil, dengan nilai *loss* yang mendekati konstan serta metrik kinerja yang relatif terjaga tanpa fluktuasi yang signifikan. Perbandingan ringkasan metrik utama dari ketiga model ditampilkan pada Tabel I.

Tabel I menunjukkan bahwa model YOLOv8 + SR mencapai kinerja keseluruhan terbaik. Model SR memiliki nilai presisi tertinggi ($\approx 86,3\%$) serta nilai mAP@0,5 tertinggi ($\approx 66,4\%$). Hal ini menunjukkan bahwa YOLOv8 + SR sedikit lebih unggul dibandingkan model asli (presisi 85,7% dan mAP@0,5 sebesar 65,2%). Sementara itu, model CBAM menunjukkan kinerja paling rendah dengan nilai presisi sekitar 77,7% dan mAP@0,5 sekitar 61,9%. Dari sisi *recall*, model asli memiliki nilai yang sedikit lebih tinggi (57,0%) dibandingkan SR (56,1%), tetapi perbedaannya sangat kecil.

Ditinjau dari mAP@0,5:0,95, yang merefleksikan kualitas deteksi secara lebih ketat, model ESRGAN memperoleh nilai tertinggi sebesar 31,4%, diikuti oleh model asli sebesar 30,4%, dan model CBAM dengan nilai terendah sebesar 28,8%. Perbedaan nilai metrik pada berbagai parameter tersebut menunjukkan bahwa penambahan modul SR memberikan peningkatan kecil pada akurasi deteksi dibandingkan YOLOv8 standar, sedangkan penambahan modul CBAM pada YOLOv8 justru menghasilkan kinerja yang lebih rendah dibandingkan model asli untuk data deteksi objek termal.

Jika dianalisis dari aspek konvergensi *loss* selama proses pelatihan, masing-masing model menunjukkan karakteristik yang berbeda. Pada awal pelatihan (*early epochs*), model

TABEL I
PERBANDINGAN RINGKASAN METRIK UTAMA UNTUK TIGA MODEL

No	Model	Presisi	Recall	mAP@0,5	mAP@0,5:0,95
1	YOLOv8 asli	85,7	57,0	65,2	30,4
2	YOLOv8 + CBAM	77,7	52,9	61,9	28,8
3	YOLOv8 + ESRGAN	86,3	56,1	66,4	31,4

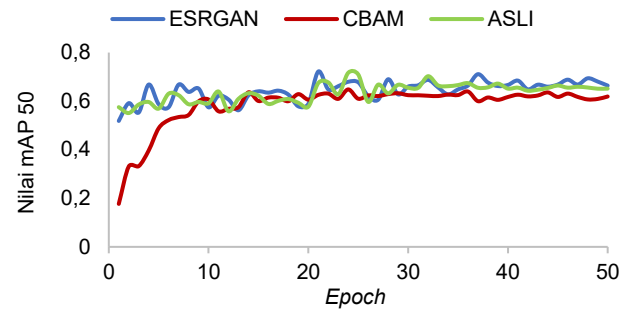
CBAM mengalami *training loss* dengan nilai yang jauh lebih tinggi dibandingkan dua model lainnya, berbeda dengan model asli dan SR yang memulai pelatihan dengan nilai *loss* awal lebih rendah. Nilai *loss* awal yang tinggi pada CBAM kemungkinan disebabkan oleh kompleksitas modul atensi CBAM yang membuat proses optimisasi menjadi lebih sulit pada tahap awal atau karena bobot awal yang kurang optimal dibandingkan model lainnya.

Penurunan *loss* tercepat terjadi pada model CBAM, yang disebabkan oleh titik awal *loss* yang sangat tinggi. Pada lima *epoch* pertama, *training loss* CBAM turun lebih dari 60%, sedangkan model asli dan SR hanya mengalami penurunan sekitar 36%. Meskipun CBAM menunjukkan penurunan *loss* yang paling cepat pada tahap awal, nilai awalnya jauh lebih buruk. Setelah sepuluh *epoch*, *training loss* CBAM masih sedikit lebih tinggi dibandingkan model asli dan ESRGAN. Selain itu, seluruh model menunjukkan tren penurunan *loss* yang melambat pada tahap akhir pelatihan, yang mengindikasikan bahwa ketiga model telah mencapai kondisi stabil. Nilai *training loss* pada model ESRGAN secara konsisten sedikit lebih rendah dibandingkan model asli pada pertengahan hingga akhir pelatihan, yang menunjukkan bahwa model ESRGAN memiliki proses pembelajaran yang sedikit lebih efektif. Sementara itu, model CBAM tetap memiliki nilai *loss* tertinggi hingga akhir pelatihan, meskipun perbedaannya tidak terlalu signifikan.

Dari sisi *validation loss*, model CBAM memiliki nilai yang sedikit lebih rendah dibandingkan model asli maupun SR. Namun hal ini tidak merefleksikan kinerja deteksi yang lebih baik. Faktanya, nilai mAP pada model CBAM lebih rendah. Temuan ini menunjukkan bahwa nilai *loss* tidak selalu berbanding lurus dengan nilai mAP, terutama ketika model cenderung menghasilkan prediksi dengan tingkat kepercayaan yang bervariasi.

Gambar 8 menunjukkan bahwa pada *epoch* ke-1, model asli dan model ESRGAN dengan YOLOv8 langsung mencapai kinerja awal yang cukup tinggi dalam hal mAP@0,5, masing-masing sebesar 57,5% dan 51,8%. Hal ini menunjukkan bahwa dengan bobot *pre-trained* yang digunakan, kedua model tersebut telah mampu mendeteksi objek termal dengan tingkat akurasi yang cukup baik sejak awal proses pelatihan. Sebaliknya, model CBAM dengan YOLOv8 memulai pelatihan dari tingkat kinerja yang jauh lebih rendah, yaitu hanya sebesar 17,7% pada *epoch* ke-1. Kondisi ini menunjukkan bahwa model CBAM pada tahap awal mengalami kesulitan dalam melakukan deteksi, yang kemungkinan disebabkan oleh penambahan modul CBAM yang mengubah distribusi fitur, sehingga membutuhkan penyesuaian yang lebih besar dibandingkan bobot awal yang digunakan.

Meskipun demikian, model CBAM menunjukkan peningkatan kinerja yang sangat cepat pada beberapa *epoch* awal, dengan nilai mAP@0,5 meningkat tajam hingga 48,8%



Gambar 8. Perbandingan nilai mAP@0,5 (mAP 50) pada gambar asli, CBAM, dan SR (ESRGAN).

pada *epoch* ke-5, yang mendekati capaian model asli maupun model ESRGAN pada *epoch* yang sama. Dengan kata lain, model CBAM mampu mengejar ketertinggalan dalam rentang 5–10 *epoch* pertama. Bahkan, pada *epoch* ke-10, kinerja model CBAM telah menyamai dan sedikit melampaui kinerja model asli dan model ESRGAN. Hal ini menunjukkan bahwa dari sisi kecepatan pembelajaran, setelah melewati fase awal yang sulit, model CBAM mampu belajar dengan cepat.

Berdasarkan perbandingan hasil penelitian, dapat disimpulkan bahwa nilai *training loss* yang lebih rendah tidak selalu berbanding lurus dengan peningkatan akurasi deteksi. Sebagai contoh, model yang dimodifikasi dengan CBAM menghasilkan nilai *validation loss* yang relatif stabil, tetapi menghasilkan nilai mAP yang lebih rendah (0,58) dibandingkan dengan model yang menggunakan ESRGAN (0,72). Temuan ini menunjukkan bahwa keberadaan modul atensi membantu proses pembelajaran fitur, tetapi tidak serta-merta meningkatkan kemampuan diskriminatif model terhadap data citra termal. Sebaliknya, model yang menggunakan ESRGAN menunjukkan nilai *training loss* yang lebih rendah dan nilai presisi yang lebih tinggi, yang mengindikasikan bahwa peningkatan kualitas resolusi citra secara langsung berkontribusi terhadap representasi fitur yang lebih baik.

Untuk mengevaluasi kecepatan dan stabilitas proses pelatihan, tidak hanya dari sisi kinerja akhir, tetapi juga dari proses konvergensinya, diperlukan penyajian tabel konvergensi yang membandingkan pendekatan berbasis SR dan pendekatan berbasis mekanisme atensi.

Tabel II menunjukkan bahwa model mempelajari keberadaan objek lebih cepat dibandingkan dengan penentuan lokasi yang presisi. Hal ini terlihat dari konvergensi mAP@0,5 yang lebih cepat dibandingkan mAP@0,5–0,95. Selain itu, dapat diamati bahwa pendekatan SR memfasilitasi pembelajaran presisi pada tahap awal, sedangkan mekanisme atensi memerlukan fase adaptasi terlebih dahulu agar dapat memberikan bobot fitur secara efektif.

Ditinjau dari sisi efisiensi model, penggunaan ESRGAN menambah beban pada tahap prapemrosesan, tetapi tidak memengaruhi kompleksitas inferensi YOLOv8. Hal ini terjadi karena ESRGAN hanya memodifikasi resolusi masukan. Sebaliknya, integrasi CBAM pada bagian *backbone* meningkatkan jumlah parameter dan komputasi pada setiap lapisan, sehingga konvergensinya lebih lambat serta akurasi keseluruhannya lebih rendah. Temuan ini menunjukkan bahwa meskipun mekanisme atensi sangat kuat untuk tugas visual, penerapannya secara langsung pada YOLOv8 berbasis citra termal dapat menyebabkan fokus berlebihan pada area dengan tekstur rendah, sehingga menurunkan efektivitas deteksi. Nilai presisi pada seluruh model secara konsisten lebih tinggi

TABEL II
TABEL KONVERGENSI

No	Model	Matriks	mAP Final	90% Final	Epoch pada 90%
1	ESRGAN+YOLOv8	mAP50	0,663940	0,314320	4
2	ESRGAN+YOLOv8	mAP50 - 95	0,597546	0,282888	2
3	CBAM+YOLOv8	mAP50	0,619020	0,287530	9
4	CBAM+YOLOv8	mAP50 - 95	0,557118	0,258777	10

dibandingkan *recall*, yang mengindikasikan bahwa model detektor cenderung bersifat konservatif dalam mengidentifikasi objek, yaitu lebih memprioritaskan ketepatan dibandingkan kelengkapan deteksi.

Ketika ketiga model dianalisis secara berdampingan, YOLOv8 yang dikombinasikan dengan ESRGAN menunjukkan keseimbangan terbaik di antara metrik-metrik akurasi. Peningkatan yang relatif kecil dibandingkan YOLOv8 asli mengindikasikan bahwa SR memberikan kontribusi paling efektif ketika menangani objek berukuran kecil atau bertekstur rendah. Pada hasil penelitian, diketahui bahwa modifikasi menggunakan CBAM tampak kurang sesuai dengan karakteristik citra termal yang memiliki gradien spasial yang lemah. Berdasarkan hasil ini, disimpulkan bahwa peningkatan resolusi citra lebih bermanfaat dibandingkan modifikasi arsitektur jaringan dalam konteks deteksi objek termal.

D. ANALISIS GALAT

Hasil penelitian menunjukkan bahwa model YOLOv8 yang dikombinasikan dengan ESRGAN mencapai tingkat akurasi keseluruhan tertinggi, diikuti oleh YOLOv8 standar dan modifikasi CBAM. Hal ini sejalan dengan penelitian-penelitian sebelumnya [14], [16], [17], [40]. Penelitian lain menunjukkan bahwa CBAM berfokus pada bagian-bagian tertentu, sehingga dapat meningkatkan hasil deteksi pada sebuah objek [18], [19]. Akan tetapi, pada penelitian yang dilakukan didapatkan hasil bahwa modifikasi CBAM kurang meningkatkan nilai pada parameter deteksi objek jika dibandingkan dengan model asli dan model yang dikombinasikan dengan ESRGAN. Hal ini terjadi karena perbedaan objek yang diteliti. Pada penelitian sebelumnya, penggunaan CBAM dilakukan bukan pada citra UAV, sedangkan pada penelitian yang dilakukan ini digunakan citra termal dari UAV [18]–[20]. Hal ini menunjukkan bahwa mekanisme atensi tidak dapat sepenuhnya menggantikan kebutuhan peningkatan kualitas citra awal.

Penggunaan ESRGAN yang berfungsi memperbesar objek juga berpotensi untuk memunculkan artefak tekstur sintetis. Namun, pada konteks deteksi manusia dengan menggunakan citra termal dari hasil kamera UAV dengan target objek kecil, penggunaan ESRGAN yang dilakukan sebelum proses deteksi dilakukan menjadi strategi yang lebih efektif dibandingkan dengan modifikasi arsitektur jaringan deteksi.

Dalam konteks implikasi dunia nyata, temuan penelitian ini memiliki relevansi langsung terhadap implementasi sistem SAR berbasis *drone*. Berdasarkan hasil yang diperoleh, model YOLOv8 berbasis ESRGAN mampu mendeteksi keberadaan manusia secara lebih andal pada lingkungan dengan pencahayaan rendah atau wilayah pegunungan, sehingga dapat mempercepat proses identifikasi korban secara signifikan. Pada model yang meningkatkan resolusi citra menggunakan

ESRGAN, meskipun terdapat tahap prapemrosesan tambahan, proses pelatihan tetap berlangsung stabil dengan konvergensi yang mulus pada setiap *epoch*. Stabilitas ini menunjukkan bahwa peningkatan resolusi citra masukan membantu model mempelajari representasi fitur secara lebih efisien sekaligus mengurangi fluktuasi selama proses *backpropagation*.

Sebaliknya, model yang dimodifikasi dengan CBAM memerlukan waktu pelatihan yang lebih lama serta sumber daya komputasi yang lebih besar akibat penambahan lapisan konvolusi pada bagian *backbone*. Peningkatan jumlah parameter yang diperkenalkan oleh mekanisme atensi menyebabkan proses pembelajaran menjadi lebih lambat, khususnya pada tahap awal pelatihan.

Dalam aplikasi dunia nyata, menjaga keseimbangan antara akurasi dan beban komputasi merupakan hal yang krusial. Meskipun ESRGAN meningkatkan kejelasan fitur dan keandalan deteksi, metode ini menambahkan beban komputasi pada tahap prapemrosesan. Namun, karena prapemrosesan dapat dilakukan sebelum tahap inferensi, ESRGAN tetap menjadi pilihan yang praktis untuk deteksi termal berbasis *drone*, terutama ketika menggunakan perangkat keras *edge* yang mendukung pemrosesan paralel. Hal ini menjadikan model ESRGAN–YOLOv8 layak untuk diimplementasikan secara *onboard* menggunakan GPU tertanam seperti NVIDIA Jetson atau perangkat *edge computing* sejenis lainnya.

V. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa model ESRGAN merupakan model yang paling unggul dalam hal akurasi deteksi dibandingkan dengan model asli dan model CBAM. Model ESRGAN mampu mendeteksi objek termal dengan akurasi yang sedikit lebih tinggi dibandingkan dengan model YOLOv8 asli, yang ditunjukkan oleh nilai presisi, mAP@0,5, dan mAP@0,5–0,95, yaitu model dengan ESRGAN masing-masing mencapai nilai tertinggi sebesar 86,3%, 66,4%, dan 31,4%. Meskipun demikian, model YOLOv8 asli juga menunjukkan kinerja yang sangat kuat dan mendekati kinerja ESRGAN, khususnya pada parameter *recall*, yaitu model asli memperoleh nilai tertinggi sebesar 57,0%. Oleh karena itu, model YOLOv8 tanpa modifikasi tetap dapat menjadi pilihan yang layak apabila efisiensi waktu pelatihan menjadi pertimbangan utama. Pada penelitian selanjutnya dapat dilakukan modifikasi pada modul SR. Selain itu, bagian pengujian dapat diperkaya dengan melakukan metode *ablation* atau juga dengan menggunakan *5 fold-cross-validation*

KONFLIK KEPENTINGAN

Para penulis menyatakan bahwa selama proses pelaksanaan dan penyusunan penelitian ini tidak terdapat konflik kepentingan, baik yang bersifat finansial, personal, maupun profesional, yang dapat memengaruhi objektivitas dan integritas hasil penelitian. Seluruh proses penelitian dilakukan secara independen tanpa adanya pengaruh dari pihak mana pun yang memiliki kepentingan terhadap hasil penelitian ini. Dengan demikian, hasil penelitian disajikan secara objektif.

KONTRIBUSI PENULIS

Konseptualisasi, Helfy Susilawati dan Akhmad Fauzi Ikhsan; metodologi, Helfy Susilawati; validasi, Arief Suryadi Satyawan; analisis, Helfy Susilawati dan Firman; investigasi, Chandra; sumber daya, Helfy Susilawati; kurasi data, Helfy Susilawati; penulisan draf awal, Helfy Susilawati; peninjauan dan penyuntingan naskah, Helfy Susilawati; visualisasi, Akhmad Fauzi Ikhsan.

REFERENCES

- [1] Krisdianto, "Pencarian pendaki yang hilang di Gunung Binaya dilanjutkan." Kementerian Kehutanan RI. Tanggal akses: 29-Sep-2025. [Online]. Tersedia: <https://www.kehutan.gov.id/news/pencarian-pendaki-yang-hilang-di-gunung-binaya-dilanjutkan>
- [2] M.B. Ibrahim, "Pemuda Sidoarjo hilang saat mendaki Gunung Buthak." Detikjatim. Tanggal akses: 29-Sep-2025. [Online]. Tersedia: <https://www.detik.com/jatim/berita/d-8068473/pemuda-sidoarjo-hilang-saat-mendaki-gunung-buthak>
- [3] M. Erdelj, E. Natalizio, K.R. Chowdhury, dan I.F. Akyildiz, "Help from the sky: Leveraging UAVs for disaster management," *IEEE Pervasive Comput.*, vol. 16, no. 1, hal. 24–32, Jan.-Mar. 2017, doi: 10.1109/MPRV.2017.11.
- [4] Meenakshi dkk., "AI-driven autonomous robots for search and rescue operations in disaster zones," dalam *2025 Int. Conf. Data Sci. Agents Artif. Intell. (ICDSAAI)*, 2025, hal. 1–6, doi: 10.1109/ICDSAAI165575.2025.11011570.
- [5] F. Mercogliano dkk., "ITINERIS geophysical technologies @CNR-IREA: Drone-based tests at Altopiano di Verteglia, Avellino (Southern Italy)," *Comput. Sci. Appl. – ICCSA 2025 Workshops*, 2025, hal. 373–383, doi: 10.1007/978-3-031-97663-6_33.
- [6] H. Shi dkk., "Improving surface soil moisture content estimation of winter oilseed rape by integrating multimodal remote sensing information and structural characteristics," *Comput. Electron. Agric.*, vol. 237, hal. 1–18, Okt. 2025, doi: 10.1016/j.compag.2025.110779.
- [7] T. Sam-Odusina dkk., "Detection and geolocation of peat fires using thermal infrared cameras on drones," *Drones*, vol. 9, no. 7, hal. 1–19, Jul. 2025, doi: 10.3390/drones9070459.
- [8] B. Chang dkk., "Enhancing wildlife detection using thermal imaging drones: Designing the flight path," *Drones*, vol. 9, no. 1, hal. 1–14, Jan. 2025, doi: 10.3390/drones9010052.
- [9] S. Yeom, "Thermal image tracking for search and rescue missions with a drone," *Drones*, vol. 8, no. 2, hal. 1–15, Feb. 2024, doi: 10.3390/drones8020053.
- [10] S. Yeom, "Multi-rotor drone-based thermal target tracking with track segment association for search and rescue missions," *Drones*, vol. 8, no. 11, hal. 1–18, Nov. 2024, doi: 10.3390/drones8110689.
- [11] H. Farman dkk., "Enhancing UAV-based human detection in thermal imaging with custom dataset," dalam *Proc. ICSDI 2024 Vol. 3*, 2024, hal. 437–444, doi: 10.1007/978-981-97-8345-8_53.
- [12] J. McGee, S.J. Mathew, dan F. Gonzalez, "Unmanned aerial vehicle and artificial intelligence for thermal target detection in search and rescue applications," dalam *2020 Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, 2020, hal. 883–891, doi: 10.1109/ICUAS48674.2020.9213849.
- [13] V. Teju dan D. Bhavana, "An efficient object tracking in thermal imaging using optimal Kalman filter," *Int. J. Eng. Trends Technol.*, vol. 69, no. 12, hal. 197–202, Des. 2021, doi: 10.14445/22315381/IJETT-V69I12P223.
- [14] L. Courtrai, M.T. Pham, dan S. Lefèvre, "Small object detection in remote sensing images based on super-resolution with auxiliary generative adversarial networks," *Remote Sens.*, vol. 12, no. 19, hal. 1–19, Sep. 2020, doi: 10.3390/rs12193152.
- [15] I. García-Aguilar, J. García-González, R.M. Luque-Baena, dan E. López-Rubio, "Object detection in traffic videos: An optimized approach using super-resolution and maximal clique algorithm," *Neural Comput. Appl.*, vol. 35, no. 26, hal. 18999–19013, Sep. 2023, doi: 10.1007/s00521-023-08741-4.
- [16] Y.R. Musunuri, O.-S. Kwon, dan S.-Y. Kung, "SRODNet: Object detection network based on super resolution for autonomous vehicles," *Remote Sens.*, vol. 14, no. 24, hal. 1–19, Des. 2022, doi: 10.3390/rs14246270.
- [17] S. Ren dkk., "Towards efficient video detection object super-resolution with deep fusion network for public safety," *Secur. Commun. Networks*, vol. 2021, hal. 1–14, Mei 2021, doi: 10.1155/2021/9999398.
- [18] Y. Zhang dkk., "Research of maritime object detection method in foggy environment based on improved model SRC-YOLO," *Sensors*, vol. 22, no. 20, hal. 1–14, Okt. 2022, doi: 10.3390/s22207786.
- [19] W. Wang dkk., "Improved detection and reidentification algorithm for natural gas pipeline excavator dynamic tracking based on deep learning," *Measurement*, vol. 257, hal. 1–17, Jan. 2026, doi: 10.1016/j.measurement.2025.118567.
- [20] Z. Lv, S. Zhu, D. Wang, dan Z. Liang, "Infrared-visible person re-identification via dual-channel attention mechanism," *Multimed. Tools Appl.*, vol. 82, no. 15, hal. 22631–22649, Jun. 2023, doi: 10.1007/s11042-023-14486-y.
- [21] C. Song dkk., "Smoke video detection based on double spectrum," *Int. J. Robot. Autom.*, vol. 39, no. 2, hal. 161–169, Jan. 2024, doi: 10.2316/J.2024.206-1020.
- [22] M. Rizk dan I. Bayad, "Bringing intelligence to SAR missions: A comprehensive dataset and evaluation of YOLO for human detection in TIR images," *IEEE Access*, vol. 13, hal. 17208–17235, Jan. 2025, doi: 10.1109/ACCESS.2025.3529484.
- [23] J. Dong, K. Ota, dan M. Dong, "UAV-based real-time survivor detection system in post-disaster search and rescue operations," *IEEE J. Miniaturization Air Space Syst.*, vol. 2, no. 4, hal. 209–219, Des. 2021, doi: 10.1109/JMASS.2021.3083659.
- [24] Z. Song dkk., "An infrared dataset for partially occluded person detection in complex environment for search and rescue," *Sci. Data*, vol. 12, hal. 1–14, Feb. 2025, doi: 10.1038/s41597-025-04600-0.
- [25] G. Batchuluun dkk., "Deep learning-based thermal image reconstruction and object detection," *IEEE Access*, vol. 9, hal. 5951–5971, Des. 2020, doi: 10.1109/ACCESS.2020.3048437.
- [26] F. Ponzini, D. van Hamme, dan M. Martelli, "Human detection in marine disaster search and rescue scenario: A multi-modal early fusion approach," *Ocean Eng.*, vol. 340, hal. 1–15, Nov. 2025, doi: 10.1016/j.oceaneng.2025.122341.
- [27] A. Vaswani dkk., "Attention is all you need," 2023, *arXiv:1706.03762*.
- [28] P. Shetty, D.K.R. dan D.S. Padre, "Deepfake detection using hybrid deep learning: Enhancing accuracy with ResNet, GANs, and attention mechanisms," dalam *2025 3rd Int. Conf. Augment. Intell. Sustain. Syst. (ICAISS)*, 2025, hal. 95–99, doi: 10.1109/ICAISS61471.2025.11041976.
- [29] F. Ying dkk., "Improvements in automatic diagnosis methods for knee osteoarthritis based on deep learning," *Chinese J. Tissue Eng. Res.*, vol. 29, no. 35, hal. 7511–7518, 2025, doi: 10.12307/2026.533.
- [30] J. Wang, J. Gao, dan B. Zhang, "A small object detection model in aerial images based on CPDD-YOLOv8," *Sci. Rep.*, vol. 15, hal. 1–16, Jan. 2025, doi: 10.1038/s41598-024-84938-4.
- [31] Y. Gao dan B. Lin, "Research and design of a small object detection and tracking system based on YOLOv8 and ByteTrack algorithms," dalam *2024 IEEE 4th Int. Conf. Data Sci. Comput. Appl. (ICDSCA)*, 2024, hal. 109–113, doi: 10.1109/ICDSCA63855.2024.10859594.
- [32] H. Zhang, M. Liang, dan Y. Wang, "YOLO-BS: A traffic sign detection algorithm based on YOLOv8," *Sci. Rep.*, vol. 15, hal. 1–11, Mar. 2025, doi: 10.1038/s41598-025-88184-0.
- [33] R. Ma dkk., "InSAR-YOLOv8 for wide-area landslide detection in InSAR measurements," *Sci. Rep.*, vol. 15, hal. 1–22, Jan. 2025, doi: 10.1038/s41598-024-84626-3.
- [34] J. Li dkk., "A small-scale target enhancement framework for aerial pineapple images on accurate agricultural information," *Comput. Electron. Agric.*, vol. 239, hal. 1–14, Des. 2025, doi: 10.1016/j.compag.2025.110874.
- [35] Y. Chen, J. Li, Y. Niu, dan J. He, "Small object detection networks based on classification-oriented super-resolution GAN for UAV aerial imagery," dalam *2019 Chin. Control Decis. Conf. (CCDC)*, 2019, hal. 4610–4615, doi: 10.1109/CCDC.2019.8832735.
- [36] C. Ledig dkk., "Photo-realistic single image super-resolution using a generative adversarial network," 2016, *arXiv:1609.04802*.
- [37] L. Riz dkk., "The MONET dataset: Multimodal drone thermal dataset recorded in rural scenarios," 2023, *arXiv:2304.05417*.
- [38] X. Wang dkk., "ESRGAN: Enhanced super-resolution generative adversarial networks," 2019, *arXiv:1809.00219*.
- [39] S. Woo, J. Park, J. Lee, dan I.S. Kweon, "CBAM: Convolutional block attention module," 2018, *arXiv:1807.06521*.
- [40] H.G. Correa *et al.*, "Thermal image super-resolution using real-ESRGAN for human detection," dalam *Proc. 20th Int. Jt. Conf. Comput. Vis. Imaging Comput. Graph. Theory Appl.*, 2025, hal. 247–254, doi: 10.5220/0013078800003912.