1. Pendahuluan

Latar Belakang

Deteksi objek merupakan salah satu tugas fundamental dalam visi komputer [1], yang menjadi tulang punggung berbagai aplikasi modern, mulai dari kendaraan otonom [2] hingga sistem pengawasan keamanan cerdas. Perkembangan pesat dalam bidang *deep learning*, khususnya *Convolutional Neural Networks* (CNN) [1], telah melahirkan berbagai arsitektur deteksi objek yang canggih. Di antara arsitektur tersebut, keluarga model YOLO (*You Only Look Once*) [3] telah mendapatkan popularitas yang signifikan karena kemampuannya menyeimbangkan antara akurasi deteksi dan kecepatan inferensi *real-time*. YOLOv8 [4], sebagai salah satu iterasi terbarunya, menawarkan peningkatan kinerja dan fleksibilitas konfigurasi yang lebih baik.

Meskipun kemajuan ini sangat pesat, deteksi objek di dunia nyata masih menghadapi tantangan signifikan, terutama ketika objek target mengalami oklusi, yaitu kondisi di mana objek terhalang sebagian atau seluruhnya oleh objek lain [5]. Oklusi menyebabkan hilangnya informasi visual yang krusial, seperti bentuk dan tekstur, yang dapat menurunkan akurasi detektor secara drastis [6], [7]. Penurunan kinerja ini menjadi isu kritis dalam aplikasi yang sensitif terhadap keselamatan, di mana kegagalan mendeteksi pejalan kaki atau kendaraan yang terhalang dapat berakibat fatal [2], [8].

Untuk mengatasi keterbatasan ini, mekanisme perhatian (*attention mechanism*) [9], yang terinspirasi dari kemampuan sistem visual manusia untuk fokus pada informasi yang relevan, telah banyak diintegrasikan ke dalam arsitektur CNN. Mekanisme ini memungkinkan model untuk secara adaptif memfokuskan pemrosesan pada fitur-fitur penting, sehingga berpotensi meningkatkan kemampuan diskriminatifnya bahkan dalam kondisi oklusi. Dua modul perhatian yang populer dan terbukti efektif adalah CBAM (*Convolutional Block Attention Module*) [10], yang menggabungkan perhatian spasial dan kanal, serta CoordAtt (*Coordinate Attention*) [11], yang mampu menangkap dependensi jarak jauh dengan mempertahankan informasi posisi yang presisi.

Walaupun integrasi modul atensi telah menunjukkan hasil yang menjanjikan dalam beberapa penelitian yang meningkatkan deteksi pejalan kaki padat [12] atau deteksi kendaraan umum [13], pertanyaan mendasar mengenai lokasi penyisipan yang optimal di dalam arsitektur YOLOv8 untuk penanganan oklusi masih belum terjawab secara komprehensif. Apakah modul atensi lebih bermanfaat jika ditempatkan di tahap awal ekstraksi fitur (*Backbone*), atau di tahap akhir saat fusi fitur multi-skala (*Neck*) seperti FPN [14] dan PANet [15]? Jawaban atas pertanyaan ini sangat penting, karena lokasi penyisipan yang berbeda dapat secara signifikan mempengaruhi aliran informasi dan, akibatnya, kinerja model secara keseluruhan. Penelitian ini dirancang untuk menjawab pertanyaan tersebut melalui investigasi yang sistematis pada dataset KITTI [16].

Topik dan Batasannya

Penelitian ini berfokus pada analisis kuantitatif dan kualitatif mengenai pengaruh lokasi penyisipan (*Neck*, *Backbone*, atau Keduanya) dan jenis modul perhatian (CBAM [10] versus CoordAtt [11]) pada kinerja arsitektur YOLOv8n [4] dalam mendeteksi objek teroklusi. Batasan penelitian ini meliputi penggunaan arsitektur dasar yang terbatas pada YOLOv8n [4] dan modul perhatian yang dievaluasi hanya CBAM [10] serta CoordAtt [11] dengan implementasi standar. Selanjutnya, lokasi penyisipan juga dibatasi pada titik-titik spesifik setelah *layer* C2f [17] tertentu di dalam *Backbone* dan *Neck* arsitektur YOLOv8n [4] standar, serta kombinasi keduanya. Selain itu, evaluasi kinerja hanya dilakukan pada dataset KITTI [16] yang telah dimodifikasi secara khusus untuk tujuan penelitian ini. Terakhir, parameter pelatihan seperti ukuran *batch*, *epoch*, dan laju pembelajaran (*learning rate*) telah ditentukan sebelumnya dan dijaga konsisten di semua eksperimen untuk memastikan perbandingan yang adil.

Tujuan

Penelitian ini memiliki tiga tujuan utama yang saling terkait untuk mencapai sasaran analisis yang komprehensif. Pertama, penelitian ini bertujuan untuk mengimplementasikan dan melatih model YOLOv8n [4] yang dimodifikasi dengan modul perhatian CBAM [10] pada tiga lokasi strategis yang berbeda, yaitu *Neck, Backbone,* dan keduanya. Kedua, tujuan serupa juga berlaku untuk modul perhatian CoordAtt [11], yaitu mengimplementasikan dan melatihnya pada lokasi-lokasi yang sama untuk memungkinkan perbandingan langsung. Ketiga, tujuan puncaknya adalah untuk mengevaluasi dan membandingkan secara sistematis kinerja deteksi objek antara model YOLOv8n [4] dasar dengan keenam varian yang telah dimodifikasi, menggunakan metrik standar seperti mAP@0.5 [18] dan

mAP@0.5:0.95 [19], pada dataset KITTI [16] yang dimodifikasi, dengan fokus utama pada kemampuan model dalam menangani berbagai tingkat oklusi objek. Keterkaitan antara tujuan penelitian, metode pengujian yang dilakukan, dan kesimpulan yang diharapkan dapat diilustrasikan dalam Tabel 1.1 di bawah ini.

Tabel 1.1 Tabel Keterkaitan antara Tujuan, Pengujian, dan Kesimpulan.

No.	Tujuan	Pengujian	Kesimpulan
1	Implementasi & Latih CBAM (*Neck*/BB/*Both*)	*Log* pelatihan berhasil, model tersimpan.	Konfirmasi keberhasilan implementasi dan pelatihan model-model berbasis CBAM.
2	Implementasi & Latih CoordAtt (*Neck*/BB/*Both*)	*Log* pelatihan berhasil, model tersimpan.	Konfirmasi keberhasilan implementasi dan pelatihan model-model berbasis CoordAtt.
3	Evaluasi & Bandingkan Kinerja Dasar vs Modifikasi (mAP, P, R)	Perhitungan metrik mAP [18], [19], P, R pada set validasi untuk semua 7 model. Visualisasi perbandingan metrik dan hasil deteksi.	Identifikasi model dan lokasi penyisipan mana yang berkinerja lebih baik secara umum dan pada metrik spesifik, serta menentukan strategi paling efektif untuk menangani oklusi.

Organisasi Tulisan

Struktur penulisan jurnal ini disusun sebagai berikut. Bagian I adalah Pendahuluan, yang menguraikan latar belakang masalah, topik dan batasan penelitian, tujuan yang ingin dicapai, serta organisasi penulisan. Bagian II, Studi Terkait, akan membahas landasan teori yang relevan dan tinjauan penelitian sebelumnya. Bagian III, Sistem yang Dibangun, akan menjelaskan secara rinci metodologi yang digunakan, termasuk arsitektur model, dataset, modifikasi, dan protokol pelatihan. Bagian IV, Hasil dan Evaluasi, akan menyajikan hasil eksperimen, baik kuantitatif maupun kualitatif, diikuti dengan analisis mendalam. Terakhir, Bagian V, Kesimpulan, akan merangkum temuan utama penelitian dan memberikan saran untuk pengembangan selanjutnya.