

BAB I

PENDAHULUAN

1.1 Latar Belakang

Orang dengan keterbatasan penglihatan seperti penyandang tunanetra perlu melakukan aktivitas sehari-sehari dengan mengandalkan indera lain selain mata. Menurut *World Health Organization* (WHO) sekitar 2.2 milyar di seluruh dunia mengalami gangguan penglihatan pada tahun 2020 (Bourne dkk., 2021). Pernyataan tersebut membuktikan bahwa banyak penderita gangguan penglihatan dan penyandang tunanetra yang perlu penanganan khusus untuk membantu mereka dalam melakukan aktivitas sehari-sehari. Kemajuan teknologi saat ini telah menghadirkan tongkat pintar yang dapat membantu penggunanya dalam melakukan berbagai aktivitas dengan mengandalkan sensor-sensor yang dapat mengidentifikasi objek di depannya. Meskipun tongkat pintar memiliki kemampuan identifikasi objek yang baik, tetapi alat tersebut tidak dapat melakukan pemindaian dan pemahaman terhadap lingkungan dengan baik. (Hidayat & Supriadi, 2019; Supriyadi, 2018; Tullah dkk., 2020). Dalam mengatasi permasalahan pada tongkat pintar tersebut, saat ini telah banyak penelitian yang menggunakan model deteksi objek sebagai solusi dari permasalahan yang ada (Tapu dkk., 2020).

Secara signifikan model deteksi objek dapat melakukan pengenalan lingkungan sekitar dengan baik. Pada umumnya, model diimplementasikan pada sebuah *device* dalam bentuk yang mudah dipakai oleh penyandang tunanetra.

Smart glasses adalah salah satu bentuk implementasi dari model deteksi objek yang digunakan untuk membantu orang dengan gangguan mata dalam beraktivitas sehari-hari (Tapu dkk., 2020). Proses deteksi objek pada beberapa *Smart Glasses* dilengkapi dengan keluaran *text-to-speech* (TTS) dari hasil pendeteksian objek. Namun, yang menjadi masalah adalah bagaimana model deteksi objek dapat melakukan deteksi *multi-class* pada gambar yang ditangkap baik secara *real-time* maupun tidak (Mahendru & Dubey, 2021). Deteksi *multi-class* merupakan metode yang dapat melakukan deteksi objek pada berbagai kelas pada satu tangkapan gambar (J. R. Terven & Cordova-esparza, 2024).

Permasalahan mengenai *multi-class detection* sebenarnya telah berhasil diselesaikan oleh YOLO (J. R. Terven & Cordova-esparza, 2024). Dalam banyak penelitian, pendekatan YOLO banyak digunakan sebagai model deteksi objek pada *smart glasses* atau alat bantu lainnya (Tapu dkk., 2020). Hal tersebut karena model YOLO memiliki performa yang mumpuni baik dari segi akurasi maupun dari kecepatan deteksi. Selain itu, perkembangan YOLO saat ini telah mencapai versi 8 dengan kemampuan yang lebih baik dari versi-versi sebelumnya (J. R. Terven & Cordova-esparza, 2024). YOLOv8 merupakan pengembangan dari YOLOv5 dengan mengganti *CSPLayer* dengan *C2f module* (Jocher dkk., 2023). Penggantian *layer* tersebut memberikan peningkatan akurasi yang lebih baik dibandingkan model tersebut dibandingkan model YOLO versi sebelumnya (J. R. Terven & Cordova-esparza, 2024). Selain itu, YOLOv8 memberikan kebaruan pada perhitungan *loss function* dengan menggunakan *Complete Intersection of Union* (CIoU) (Zheng dkk., 2020) untuk menghitung *bounding box loss* dan perhitungan

binary cross-entropy dengan menggunakan *Distribution Focal Loss* (DFL) (X. Li, Wang, dkk., 2020) untuk menghitung *classification loss* (J. Terven & Cordova-Esparza, 2023). *Bounding box loss* berfokus pada pengukuran seberapa baik model dapat memprediksi kotak pembatas (*bounding box*) yang tepat untuk objek yang terdeteksi. *Binary Cross Entropy* (BCE) adalah sebuah fungsi kerugian (*loss function*) yang sering digunakan dalam tugas klasifikasi biner. Ini digunakan untuk mengukur perbedaan antara distribusi probabilitas prediksi model dan label kelas yang sebenarnya (*ground truth*). Meskipun begitu, YOLOv8 perlu dioptimasi kembali untuk menghasilkan akurasi dan performa model yang lebih baik lagi.

Berdasarkan deskripsi persoalan yang telah diuraikan pada paragraf sebelumnya, masih terdapat kesenjangan penelitian yang menjadi perhatian utama dalam penelitian ini, yaitu performa model YOLOv8 yang harus dioptimalkan kembali agar dapat menghasilkan model dengan akurasi deteksi dan pemahaman lingkungan yang lebih baik. Sehingga dalam penelitian ini berfokus pada pengembangan model deteksi objek yang digunakan untuk membantu penyandang tunanetra. Model dibangun dengan YOLOv8 yang dioptimasi dengan *Adaptive Moment Estimation* (Adam) *optimizer*. Karena pemilihan *learning rate* pada proses optimasi sangat berpengaruh pada hasil dari proses pelatihan (S. Jiang dkk., 2018), maka pemilihan *learning rate* dipilih secara otomatis dengan metode *Particle Swarm Optimization* (PSO) yang dilengkapi *Adaptive Inertia Weight* (Kessentini & Barchiesi, 2015). Selanjutnya agar menjadi lebih representatif, model dibuat agar dapat melakukan estimasi perkiraan jarak terhadap objek yang dideteksi dengan menggunakan teknik *monocular camera*. Teknik tersebut menggunakan kamera

berlensa satu yang dipadukan dengan proses kalibrasi dan perhitungan proporsi. Selain itu, model dilengkapi dengan *speech recognition* untuk membantu para penyandang tunanetra dalam menggunakan alat yang diterapkan model tersebut. *Output* yang dihasilkan dari model ini berupa suara dari hasil pendeteksian objek yang dikonversi menggunakan TTS.

Penelitian ini memberikan peluang dalam mengembangkan sistem model dari sistem deteksi objek berbasis *speech recognition* yang dapat dijadikan landasan dalam pengembangan lebih lanjut, salah satunya untuk membantu para penyandang tunanetra dalam melakukan aktivitasnya. Selain itu, penelitian ini dapat menciptakan sistem deteksi yang dapat menghitung jarak objek dengan kamera menggunakan konsep *monocular camera*. Konsep tersebut mempertimbangkan penggunaan lensa tunggal dalam menghitung jarak objek yang dipengaruhi oleh proses kalibrasi setiap objek. Pengembangan pendekatan ini diharapkan dapat meningkatkan kualitas dari sistem deteksi objek sehingga performanya lebih baik dari model lain.

1.2 Rumusan Masalah

Berdasarkan penyusunan latar belakang di atas, dihasilkan rumusan masalah yang diangkat pada penelitian ini sebagai berikut.

1. Bagaimana melakukan penerapan konsep *monocular camera*, *speech recognition* dan TTS pada model deteksi objek yang dikembangkan?
2. Bagaimana performa model deteksi yang dibangun dengan penerapan YOLOv8 yang dioptimasi oleh PSO dan *Adam Optimizer*?

1.3 Tujuan Penelitian

Tujuan penelitian ini dispesifikasikan berdasarkan rumusan masalah yang telah dibuat sebagai berikut.

1. Menerapkan konsep *Monocular Camera*, *speech recognition*, *Text-to-Speech* (TTS) pada model deteksi objek yang dikembangkan.
2. Melakukan optimasi pada model deteksi yang dibangun oleh YOLOv8 menggunakan PSO dan *Adam Optimizer*.

1.4 Manfaat Penelitian

Penelitian ini dapat memberikan manfaat kepada pengguna khususnya bagi para penyandang tunanetra. Manfaat tersebut di antaranya sebagai berikut.

1. Memberikan kemudahan kepada para penyandang tunanetra dalam beraktivitas di dalam ruangan.
2. Model dilengkapi dengan *speech recognition* sehingga memberikan pengalaman kepada pengguna seperti memiliki asisten.
3. Hasil model ini dapat diterapkan pada alat bantu bagi tunanetra yang menerapkan sistem deteksi objek.
4. Memberikan kebaruan dalam melakukan pemilihan *learning rate* yang paling optimal dalam proses optimalisasi menggunakan PSO dan *Adam Optimizer* pada YOLOv8.
5. Penambahan perhitungan jarak dengan menggunakan konsep *monocular camera* membuat fitur dari model YOLOv8 lebih lengkap.

1.5 Batasan Masalah

Penelitian ini memiliki batasan-batasan masalah yang digunakan untuk membuat proses penelitian menjadi lebih spesifik. Batasan-batasan tersebut adalah sebagai berikut.

1. Terdapat dua *dataset* yang digunakan, yaitu *Common Objects in Context (COCO) dataset* sebagai tolok ukur perbandingan dengan model penelitian lain dan *custom dataset* sebagai percobaan dalam skema *real-time*.
2. Menggunakan bahasa pemrograman *python* untuk pengembangan model pendeteksian objek.
3. Model *speech recognition* dan model TTS yang digunakan terbatas pada model yang sudah ada dan tersedia secara *online* sehingga memungkinkan terjadi respon yang kurang cepat.
4. Percobaan berfokus pada pendeteksian objek di dalam ruangan (*indoor*).

1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini disesuaikan dengan panduan penulisan TA yang diterbitkan oleh Universitas Siliwangi dengan terdiri dari lima BAB sebagai berikut.

BAB I PENDAHULUAN

BAB I membahas mengenai latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, dan sistematika penulisan penelitian yang dilakukan.

BAB II LANDASAN TEORI

BAB II berisi tentang teori-teori pendukung beserta studi-studi penelitian terkait yang berhubungan dengan permasalahan yang akan diteliti sehingga dihasilkan sebuah *state-of-the art* dan menjadi acuan konseptual terhadap penelitian yang dilakukan

BAB III METODOLOGI

BAB III memaparkan tentang metodologi dalam melakukan penelitian beserta penjelasannya secara lengkap, seperti penjelasan tentang studi literatur, pengumpulan data, kalibrasi, praproses data, optimasi parameter, pelatihan, dan evaluasi model.

BAB IV HASIL DAN PEMBAHASAN

BAB IV berisi tentang hasil yang didapat dari proses penelitian yang dilakukan serta pembahasan terkait apa yang telah dicapai dengan merujuk pada poin-poin penting seperti, hasil rancangan arsitektur, hasil pengumpulan data, hasil dari proses optimasi parameter menggunakan PSO, perbandingan performa model deteksi YOLOv8 setelah dioptimasi dengan performa model pada penelitian lain, hasil pengujian terhadap prediksi estimasi jarak yang dihitung oleh sistem, dan hasil pengukuran kinerja sistem pada model yang menggunakan *custom dataset*.

BAB V KESIMPULAN DAN SARAN

BAB V berisi tentang kesimpulan atas semua rangkaian penelitian yang telah dilakukan dengan memaparkan keunggulan dan kekurangan penelitian

serta saran peluang penelitian selanjutnya untuk memperbaiki kekurangan pada penelitian yang telah dilakukan