

ABSTRAK

DETEKSI OBJEK BERGERAK MENGGUNAKAN METODE *YOLO* DAN *DEPTH ESTIMATION* BERBASIS VIDEO

Nama : Dwi Bagus Prakoso

NPM 2255201207

Pembimbing : Yuza Reswan, S.Kom., M.Kom

Perkembangan teknologi kecerdasan buatan (Artificial Intelligence) khususnya bidang computer vision telah mendorong perkembangan pengolahan citra dan video untuk memahami kondisi lingkungan secara otomatis. Salah satu penerapannya adalah deteksi objek, tetapi metode deteksi umumnya hanya memberikan informasi jenis dan lokasi objek dalam bounding box tanpa informasi jarak terhadap kamera. Keterbatasan tersebut menyebabkan informasi yang dihasilkan belum cukup untuk mendukung sistem yang membutuhkan pemahaman lingkungan secara lebih lengkap. Oleh karena itu, diperlukan pendekatan yang menggabungkan deteksi objek dan estimasi jarak menggunakan kamera RGB tanpa perangkat keras tambahan. Penelitian ini bertujuan mengembangkan sistem deteksi objek dan estimasi jarak berbasis video menggunakan kombinasi YOLOv10 dan ZoeDepth. YOLOv10 digunakan untuk mendeteksi objek dan menentukan posisi pada setiap frame video, sedangkan ZoeDepth digunakan untuk melakukan estimasi kedalaman (Depth Estimation) guna memperoleh perkiraan jarak objek terhadap kamera. Metode penelitian meliputi pengumpulan data, prapemrosesan, implementasi model, dan evaluasi sistem. Data yang digunakan berupa 250 sampel video CCTV, terdiri dari 162 data sekunder dan 88 data primer. Seluruh data digunakan sebagai data pengujian karena YOLOv10 merupakan model pre-trained yang telah dilatih menggunakan dataset COCO. Evaluasi deteksi objek menggunakan metrik mAP, sedangkan estimasi jarak menggunakan Root Mean Square Error (RMSE) dengan dataset iBims-1 sebagai acuan ground truth Depth. Hasil pengujian menunjukkan sistem mampu mendeteksi beberapa objek, yaitu car, bus, truck, dan person, serta memberikan estimasi jarak pada objek yang terdeteksi. Hasil evaluasi deteksi objek memperoleh nilai mAP sebesar 81,03%, sedangkan evaluasi estimasi jarak menggunakan ZoeDepth menghasilkan nilai RMSE sebesar 4,2248 meter. Berdasarkan hasil tersebut, kombinasi YOLOv10 dan ZoeDepth berhasil menghasilkan sistem yang mampu melakukan deteksi objek sekaligus memberikan informasi estimasi jarak berbasis video menggunakan kamera RGB. Sistem ini dapat menjadi salah satu pendekatan dalam pengembangan aplikasi computer vision yang membutuhkan informasi objek dan jarak secara bersamaan. Pendekatan ini dapat meningkatkan kelengkapan informasi visual untuk mendukung pemantauan lingkungan secara otomatis.

Kata kunci : Deteksi Objek, Estimasi Jarak, *YOLOv10*, *ZoeDepth*

ABSTRACT

MOVING OBJECT DETECTION USING THE YOLO METHOD AND VIDEO-BASED DEPTH ESTIMATION

Name : Dwi Bagas Prakoso

NPM 2255201207

Supervisor : Yuza Reswan, S.Kom., M.Kom.

The development of artificial intelligence (Artificial Intelligence), particularly in the field of computer vision, has driven advances in image and video processing to automatically understand environmental conditions. One of its applications is object detection; however, detection methods generally only provide information about the type and location of objects within bounding boxes without information about their distance from the camera. This limitation means that the information generated is insufficient to support systems requiring a more comprehensive understanding of the environment. Therefore, an approach that combines object detection and distance estimation using an RGB camera without additional hardware is required. This study aims to develop a video-based object detection and distance estimation system using a combination of YOLOv10 and ZoeDepth. YOLOv10 is used to detect objects and determine their positions in each video frame, while ZoeDepth is used to perform depth estimation to obtain the estimated distance of objects from the camera. The research methodology includes data collection, preprocessing, model implementation, and system evaluation. The data consist of 250 CCTV video samples, comprising 162 secondary data and 88 primary data. All data were used as testing data because YOLOv10 is a pre-trained model previously trained using the COCO dataset. Object detection performance was evaluated using the mAP metric, while distance estimation was evaluated using Root Mean Square Error (RMSE), with the iBims-1 dataset serving as the ground-truth depth reference. The test results show that the system is capable of detecting several objects, namely cars, buses, trucks, and persons, while also providing distance estimates for the detected objects. The object detection evaluation achieved an mAP value of 81.03%, while the distance estimation evaluation using ZoeDepth produced an RMSE value of 4.2248 meters. Based on these results, the combination of YOLOv10 and ZoeDepth successfully produces a system capable of performing object detection while simultaneously providing video-based distance estimation using an RGB camera. This system can serve as an approach for developing computer vision applications that require both object and distance information. This approach can improve the completeness of visual information to support automated environmental monitoring.

Keywords : Object Detection, Distance Estimation, YOLOv10, ZoeDepth

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi kecerdasan buatan *AI (Artificial Intelligence)* telah mendorong kemajuan yang signifikan dalam bidang pengolahan citra dan video. Salah satu bidang yang berkembang pesat adalah *computer vision*, yang berfokus pada kemampuan sistem komputer dalam memahami dan menafsirkan kondisi lingkungan melalui data visual (Rohim, 2025). Deteksi objek merupakan salah satu penerapan utama dalam *computer vision* yang bertujuan untuk mengenali dan mengidentifikasi objek bergerak di lingkungan sekitar, seperti manusia, kendaraan, dan objek statis lainnya. Teknologi ini telah banyak diterapkan dalam berbagai aplikasi nyata, termasuk sistem pengawasan, robotika, kendaraan pintar, dan sistem bantuan navigasi. Penelitian terdahulu menunjukkan bahwa pendekatan berbasis *deep learning* mampu menghasilkan kinerja deteksi objek yang lebih akurat dan cepat dibandingkan metode tradisional yang masih bergantung pada fitur manual (Raisha Agma, 2025).

Algoritma *YOLO (You Only Look Once)* dikenal sebagai metode deteksi objek yang memiliki kecepatan tinggi dan mampu bekerja pada video karena proses deteksi dilakukan dalam satu tahap jaringan saraf (Nafis Alfarizi *et al.*, 2023). Karakteristik tersebut menjadikan *YOLO* sesuai untuk diterapkan pada sistem deteksi objek berbasis video karena mampu memberikan keseimbangan antara kecepatan proses deteksi

dan akurasi hasil. Namun, *YOLO* tidak menyediakan informasi kedalaman atau jarak objek. Di sisi lain, *ZoeDepth* merupakan metode estimasi kedalaman berbasis single image *Depth Estimation* yang mampu menghasilkan peta kedalaman dari satu citra tanpa memerlukan sensor tambahan dan tanpa pelatihan ulang khusus. *ZoeDepth* dinilai fleksibel dan mampu bekerja pada berbagai kondisi lingkungan, meskipun tidak memiliki kemampuan untuk mengenali jenis objek secara langsung (Bhat *et al.*, 2023). Kemampuan tersebut menjadikan *ZoeDepth* sesuai untuk diintegrasikan dengan *YOLO* dalam menghasilkan informasi estimasi jarak menggunakan kamera *RGB* tanpa memerlukan perangkat keras tambahan.

Namun demikian, sebagian besar sistem deteksi objek yang dikembangkan saat ini masih terbatas pada kemampuan mengenali jenis objek dan menentukan posisi objek pada citra dua dimensi. Informasi jarak antara kamera dan objek belum menjadi fokus utama dalam banyak penelitian, padahal informasi tersebut memiliki peranan penting dalam sistem yang bekerja di lingkungan nyata. Pada aplikasi seperti navigasi robot, sistem keselamatan, dan penghindaran tabrakan, ketepatan estimasi jarak objek sangat memengaruhi proses pengambilan keputusan sistem. Beberapa penelitian menunjukkan bahwa sistem deteksi objek konvensional masih mengalami kesulitan dalam memperkirakan jarak objek secara akurat (Bachtiyar, 2024)(Ramadhani, Muttakin and Irsyad, 2026).

Berbagai pendekatan telah dikembangkan untuk mengatasi permasalahan estimasi jarak objek dalam sistem deteksi objek bergerak lingkungan sekitar.

Beberapa penelitian memanfaatkan sensor tambahan seperti *LiDAR* atau kamera *stereo* untuk memperoleh informasi kedalaman yang lebih presisi (Pinasthika and Utaminingrum, 2022). Meskipun pendekatan tersebut mampu meningkatkan akurasi, penggunaan sensor tambahan berdampak pada meningkatnya biaya dan kompleksitas sistem, sehingga kurang sesuai untuk penerapan pada sistem berbasis *real-time* dan perangkat berskala kecil. Pendekatan lain menggunakan metode estimasi kedalaman berbasis citra tunggal (*Monocular Depth Estimation*), namun metode ini masih menghadapi tantangan terkait akurasi, kestabilan hasil, serta kebutuhan komputasi yang tinggi (Katiyar *et al.*, 2024).

Berdasarkan permasalahan dan pendekatan yang telah dijelaskan, penelitian ini diharapkan mampu menggabungkan keunggulan metode *YOLO* dan *ZoeDepth* dalam membangun sistem deteksi objek bergerak di lingkungan sekitar dan estimasi jarak berbasis video. *YOLO* digunakan untuk mendeteksi serta menentukan lokasi objek, sedangkan *ZoeDepth* dimanfaatkan untuk memperoleh informasi kedalaman yang digunakan dalam proses perhitungan jarak objek dari kamera. Penelitian ini diharapkan dapat menghasilkan sistem yang lebih efektif, efisien, dan aplikatif, serta memberikan kontribusi dalam pengembangan sistem berbasis *computer vision* yang mampu bekerja pada perangkat dengan spesifikasi terbatas.

Berbeda dengan penelitian sebelumnya, penelitian ini menggabungkan algoritma *YOLO* sebagai metode deteksi objek dengan *ZoeDepth* sebagai metode estimasi jarak. Dengan demikian sistem yang dikembangkan tidak hanya mampu

mendeteksi objek, tetapi juga memberikan informasi jarak objek terhadap kamera, sehingga informasi yang dihasilkan menjadi lebih lengkap.

1.2 Pertanyaan Penelitian

Berdasarkan latar belakang yang telah diuraikan penelitian tersebut, maka penulis merumuskan masalah:

Bagaimana implementasi dan kinerja metode *YOLO* dan *Depth Estimation* dalam mendeteksi objek serta menghitung estimasi jarak berbasis video dengan variasi jarak, dan kondisi pencahayaan yang tidak teratur?

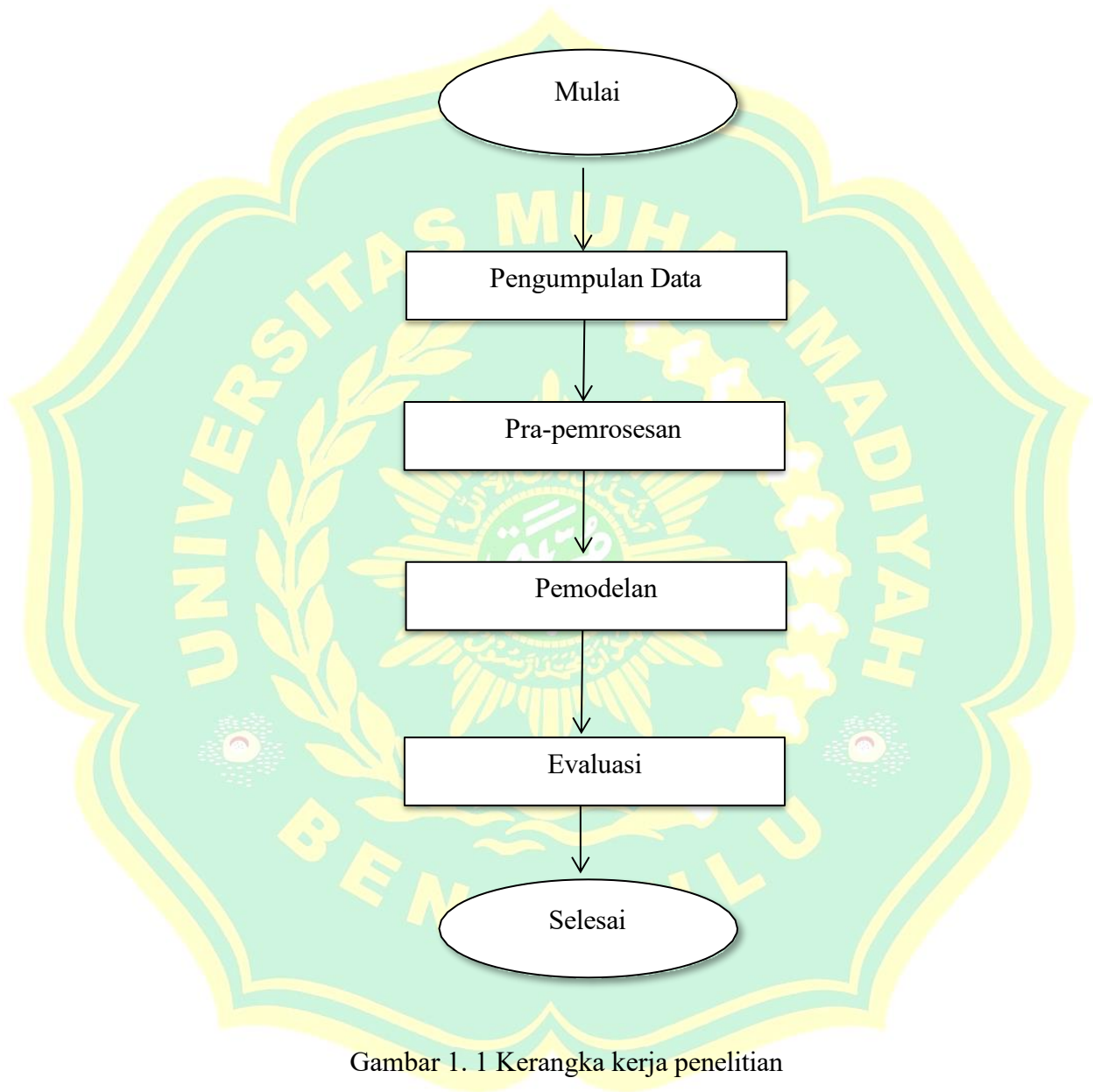
1.3 Tujuan Penelitian

Penelitian ini bertujuan mengembangkan dan mengimplementasikan metode *YOLOv10* dan *ZoeDepth* untuk mendeteksi objek serta menghitung estimasi jarak objek pada video. Hasil penelitian diharapkan dapat menghasilkan sistem yang efektif dan mudah diterapkan dalam memberikan informasi deteksi objek beserta estimasi jaraknya. Pengujiannya menggunakan dengan menghitung *mAP* untuk menilai kinerja deteksi objek, serta *RMSE* untuk mengukur tingkat ketepatan estimasi jarak. Penelitian ini diharapkan dapat menghasilkan sistem yang sederhana, efektif, dan mudah diterapkan.

1.4 Kerangka Kerja Penelitian

Penelitian ini dilakukan melalui beberapa tahapan, yaitu pengumpulan data, pra-pemrosesan, pemodelan menggunakan algoritma *YOLOv10* dan *ZoeDepth*,

serta evaluasi hasil. Setiap tahapan disusun secara sistematis untuk mencapai tujuan penelitian. Adapun kerangka kerja penelitian yang menggambarkan keseluruhan proses tersebut dapat dilihat pada Gambar 1.1.



Gambar 1. 1 Kerangka kerja penelitian