

BAB I

PENDAHULUAN

I.1. Latar Belakang

Badan Pusat Statistik (BPS) melaporkan bahwa sepanjang tahun 2019 telah terjadi 116.411 kasus kecelakaan di jalan raya [1]. Sebanyak 25.671 jiwa dinyatakan meninggal dunia dan total kerugian materi ditaksir mencapai Rp254.779.000 [1]. Sementara itu, dilansir dari Departemen Transportasi Amerika Serikat (AS), lebih dari 90% kecelakaan mobil di AS disebabkan oleh kesalahan pengemudi (*human error*). Teknologi kendaraan otonom (*autonomous vehicle*) digadang-gadang menjadi teknologi yang dapat mengatasi dampak negatif dari kecelakaan di jalan raya yang disebabkan oleh kesalahan pengemudi [2]. Selain itu, teknologi ini juga diharapkan dapat mengurangi kemacetan arus lalu lintas dan meningkatkan efisiensi bahan bakar [2]. Akan tetapi, kehandalan teknologi kendaraan otonom hanya dapat dicapai jika sistem kendaraan dapat secara akurat mengetahui keadaan lingkungan di sekitarnya.

Sistem persepsi kendaraan otonom merupakan sistem yang berfungsi untuk mengetahui keadaan lingkungan sekitar dengan cara mengubah data sensorik menjadi informasi semantik [3]. Informasi semantik yang dihasilkan dapat berupa posisi dan dimensi objek, identifikasi rambu lalu lintas dan pemetaan area yang dapat dilalui [3]. Dari banyaknya fungsi sistem persepsi, dalam tugas akhir ini penulis hanya membahas pada fungsi estimasi posisi, dimensi dan orientasi objek. Dengan demikian, peran sistem persepsi ini menjadi sangat penting, karena kegagalan sistem persepsi untuk mengenali keadaan lingkungan dapat menyebabkan insiden kecelakaan lalu lintas [3].

Salah satu faktor kegagalan dalam sistem persepsi muncul dari keterbatasan sensor dan adanya variasi lingkungan seperti pencahayaan dan kondisi cuaca [4]. Keterbatasan sensor, selain merupakan faktor internal sensor itu sendiri juga dapat disebabkan oleh kegagalan sensor dalam mengidentifikasi objek. Oklusi



merupakan kegagalan sensor membaca keseluruhan objek dikarenakan objek yang dituju terhalang objek lain. Ukuran dan jarak objek juga dapat secara dramatis mempengaruhi pembacaan sensor sehingga dapat menghasilkan representasi yang berbeda untuk objek dari kelas yang sama [4].

Faktor yang disebutkan di atas memberikan tantangan pada bidang persepsi kendaraan otonom. Walaupun begitu, kinerja sistem identifikasi atau deteksi objek secara 2 dimensi (2D) meningkat pesat sepanjang tahunnya. Kinerja tersebut diukur dalam metrik *Average Precision* (AP). Kinerja sistem deteksi 2D dewasa ini dapat mencapai nilai AP lebih dari 90% pada dataset KITTI (*Karlsruhe Institute of Technology and Toyota Technological Institute*) yang dijadikan sebagai dataset tolok ukur (*benchmark*) dalam penelitian ini [5].

Sistem persepsi deteksi objek 2D memberikan informasi posisi objek dalam bingkai (*frame*) kamera, tetapi tidak memberikan informasi posisi objek pada bidang dunia atau bidang 3D. Oleh karena itu, dihadirkan sistem persepsi deteksi objek 3D untuk mengatasi masalah ini. Sistem persepsi 3D dapat memberikan informasi kedalaman (*depth*), sehingga informasi lokasi, ukuran dan orientasi objek dalam bidang 3D dapat diperoleh. Dengan demikian, sistem deteksi objek 3D dapat memberikan informasi yang lebih kaya dibanding deteksi objek 2D.

Manusia menggunakan sistem visual dan pendengaran saat mengemudi kendaraan. Berbeda dengan manusia, sistem mengemudi otonom mengandalkan beberapa modalitas sensor untuk mengatasi kekurangan dari sensor individu. Sensor tersebut meliputi sensor pasif seperti kamera monokular dan kamera stereo dan sensor aktif seperti LiDAR, radar dan sonar. Sistem sensor yang mengandalkan kamera saja atau kombinasi kamera-LiDAR banyak digunakan pada penelitian dewasa ini [6].

Sensor kamera tunggal (monokular) memiliki keunggulan pada harga yang murah, konfigurasi yang mudah, dan resolusi yang tinggi. Berbeda dengan sensor kamera ganda (*stereo*) dan LiDAR, sensor kamera monokular tidak bisa mendapatkan informasi kedalaman (*depth*) yang diperlukan untuk melakukan



deteksi objek 3D. Akan tetapi perkembangan bidang *computer vision* dan *deep learning* telah memungkinkan sensor kamera monokular digunakan untuk deteksi objek 3D [6].

Kamera monokular digunakan dalam beberapa penelitian terakhir yang berkaitan dengan deteksi objek 3D. Mousavian *et al.* [7] menggunakan metode regresi kotak pembatas (*bounding-box*) 2D menjadi kotak pembatas 3D dengan menggunakan *Convolutional Neural Network* (CNN). Sementara itu, Li *et al.* [8] dan Liu *et al.* [9] menggunakan metode regresi titik kunci (*keypoint*) menjadi kotak pembatas 3D menggunakan model CenterNet [10]. Ketiga penelitian terakhir tersebut menjadi *State-of-The-Art* (SOTA) atau yang tercanggih pada sistem deteksi objek 3D.

Selain sensor yang digunakan, perangkat keras (*hardware*) komputasi juga menjadi pertimbangan pada tugas akhir ini. Mikrokontroler Nvidia Jetson Nano merupakan perangkat berdaya rendah yang dilengkapi dengan akselerator GPU (*Graphic Processing Unit*). Akselerator ini sangat cocok digunakan untuk komputasi gambar berbasis *deep learning*. Dengan demikian, pada tugas akhir ini penulis menggunakan mikrokontroler Nvidia Jetson Nano untuk proses pengujian model *deep learning*.

Berdasarkan pemaparan latar belakang di atas, pada tugas akhir kali ini akan digunakan metode berbasis *deep learning* untuk deteksi objek 3D yang akan ditanamkan pada mikrokontroler Nvidia Jetson Nano. Model *deep learning* yang dibuat akan mengacu pada arsitektur Mousavian *et al.* [7] dengan kebaruan pada model detektor berbasis YOLO (*You Only Look Once*) [11] dan model regresi CNN berbasis MobileNet [12]. Kombinasi YOLO-MobileNet penulis yakin menjadi kombinasi yang optimal dikarenakan kedua model dikenal sebagai model yang ringan sekaligus akurat.



I.2. Perumusan Masalah

Berdasarkan latar belakang di atas disusun rumusan masalah sebagai berikut:

1. Bagaimana rancangan arsitektur model *deep learning* untuk deteksi objek 3D menggunakan kamera monokular?
2. Bagaimana metode implementasi model deteksi objek 3D pada mikrokontroler Nvidia Jetson Nano?

I.2.1. Batasan Masalah

Berikut batasan masalah yang digunakan dalam penelitian ini:

1. Model *deep learning* hanya dilatih pada tiga kelas objek yaitu *car*, *pedestrian* dan *cyclist* menggunakan data kamera monokular RGB (berwarna) dari dataset KITTI.
2. Kinerja model hanya diukur menggunakan metrik yang disediakan oleh dataset KITTI (mAP2D, mAP3D, mAPBEV, AOS, dan latensi).

I.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang dipaparkan, dapat disusun tujuan penelitian sebagai berikut:

1. Merancang arsitektur model *deep learning* untuk deteksi objek 3D dari data kamera monokular.
2. Mengimplementasikan model pada mikrokontroler Nvidia Jetson Nano.

I.4. Manfaat Penelitian

Manfaat penelitian yang dapat dicapai sebagai berikut:

1. Memperoleh model *deep learning* yang ringan dan telah dilatih pada dataset KITTI untuk keperluan deteksi objek 3D.
2. Memperoleh metode implementasi model yang optimal pada perangkat berdaya rendah mikrokontroler Nvidia Jetson Nano.

