ABSTRACT

This research proposes the development of a helmet pose estimation model for detecting distracted driving behavior in motorcyclists, a crucial issue contributing to the high global rate of traffic accidents. Unlike traditional head pose estimation approaches hindered by helmet use, this study addresses the research gap by focusing on helmet orientation detection as an indicator of visual distraction. The CRISP-DM methodology was systematically applied, from problem identification to system implementation and evaluation. The developed solution adopts a two-stage approach: helmet object detection using the YOLOv8n model and helmet pose estimation using a modified CNN architecture with a ResNet18 backbone. The data collection process involved annotating helmet images and labeling pose (yaw, pitch, roll) using a combination of IMU sensors to obtain ground truth. Evaluation results demonstrate strong performance from both models. The YOLOv8n object detection model achieved an mAP50 of 0.9800 and a GPU inference speed of 0.99 ms. Nevertheless, further comparisons indicate YOLOv11n as a more optimal option with an mAP50 of 0.9909, mAP50-95 of 0.9184, CPU speed of 56.1 ms, and a model size of 5,343 KB. For pose estimation, the ResNet18 model achieved low Mean Absolute Error (MAE) for the roll (1.06°), pitch (1.13°), and yaw (1.47°) axes, with a GPU inference speed of 3.67 ms. User validation confirmed the system's effectiveness in problem-solving (score 8) and ease of interface use (score 9), though there was a note regarding the need for response speed optimization (score 7). Overall, this research successfully created a foundation for an accurate and efficient helmet pose estimation system for distracted driving detection, paving the way for further development towards more implementable road safety applications. Future suggestions include improving sensor data label accuracy, exploring one-stage detection models, and designing systems for edge devices.

*Keywords—**helmet pose estimation**, **distracted driving detection, deep learning, computer vision***