別紙４－１（課程博士（英文））

Date of Submission: July 3rd, 2023

Department of Computer Science and Engineering	Student ID Number	209302
Applicant's name	Oskar Natan	

Supervisor	Prof. Jun Miura Prof. Shigeru Kuriyama

Abstract (Doctor)

Title of Thesis	Deep Perception-Action Coupling and Sensor Fusion for End-to-end Autonomous Driving

Approx. 800 words

In autonomous driving, how should we integrate perception and action components properly? What do we need to process and fuse multi-modal data from various sensors? In this study, we aim to answer these questions by conducting some experiments on end-to-end autonomous driving to achieve excellent drivability in complex environments under diverse conditions and scenarios.

End-to-end autonomous driving is a method that allows an autopilot model to directly use raw sensor data as the inputs and outputs the low-level control such as steering angle and throttle level. Since manual integration for joining perception and action parts is no longer needed, this method has become a preferred approach as the model can examine the necessity of information all by itself. Moreover, this method can be combined with imitation learning or behavior cloning which can be done easily with the supervised learning technique. By using the end-to-end behavior cloning strategy, we can create a single deep learning model to mimic an expert driver manipulation on vehicle control in handling complicated situations, which can be simulated using a simulator or derived from publicly available datasets to enrich the driving experiences. Although this method has plenty of benefits in performing human-like autonomous driving, there are two fundamental issues that remain and need to be addressed carefully. First, since the model has multiple outputs as it deals with perception and control tasks simultaneously at the same time, we must ensure that the solution for each task can be learned at an equal pace. This is necessary to prevent the model from tending to focus only on one specific task during the training process. Second, we need to design a complex network architecture using the correct layers and tune some hyperparameters accordingly, so that the model can handle multiple input data with different modalities. This challenge arises as we must create a compact unified model to avoid the burden of linking some independent task-specific modules, which may lead to information loss during the process of forming a connected modular system.

To tackle those issues, we propose an end-to-end multi-task model that can extract meaningful information from a set of observations retrieved by vehicle sensors and solve multiple tasks from the perception stage to the action stage in one forward pass. We also propose an adaptive loss weighting algorithm to balance the learning signal for each task equally. Then, we evaluate our proposed model by doing ablation and comparative studies with other models for clearer performance justification. The experimental results show that our model achieves superior performance in many criteria and aspects of driving. To better clarify the findings, this study is conducted gradually and step-by-step from the perception parts to the integration of perception and action parts, from predicting some driving records to performing automated driving, and from the utilization of simulation programs to real-world implementation on a robotic vehicle. Furthermore, as part of this thesis and to support future studies, we share the codes and data used in our publications publicly at https://github.com/oskarnatan. See the publication page for more details.