

Master's Thesis Project

Tokenization as a Neural Compression Strategy in Automotive Embedded Systems

Tim Boleslawsky, gusbolesti@student.gu.se

Emrik Dunvald, gusdunvem@student.gu.se

December 2025

Suggested Supervisor at CSE: Yinan Yu

Suggested Supervisor at Company: Dhasarathy Parthasarathy

Relevant completed courses student 1:

- DAT410 / DIT728 Design of AI systems
- DAT341 / DIT867 Applied machine learning
- DAT246 / DIT246 Empirical software engineering
- DAT450 / DIT247 Machine learning for natural language processing

Relevant completed courses student 2:

- DAT410 / DIT728 Design of AI systems
- DAT341 / DIT867 Applied machine learning
- 12054 Advanced Course on Natural Language Processing
- 05630 Ad Hoc and Sensor Networks

1 Introduction

Modern in-vehicle networks struggle to keep up with the demand of transmitting ever growing amounts of data. This is especially apparent as the importance of machine learning (ML) tasks steadily increases. Traditional data handling techniques like event-triggered logging were originally introduced to reduce the network load, but are increasingly insufficient at providing context-rich data for downstream ML tasks.

Neural compression is an established research field that combines ideas from information theory and machine learning techniques to produce learned compression models that adaptively reduce the data strain on systems. In recent years this has been especially relevant for autonomous driving research, where the multitude of used sensors accentuate the shortcomings of traditional in-vehicle networks. But, while there exists various research into compressing video and image data for exactly these autonomous driving challenges, there is a noticeable lack of neural compression for time-series data in modern research, especially, when also considering computational constraints imposed by embedded systems.

We propose the use of a tokenization-based neural compression framework to address these research gaps. Tokenization is an established practice in natural language processing and has also shown promise in audio and speech processing. In the context of neural compression, tokenization serves as an alternative way to create efficient discrete latent representations, which enable a more lightweight compression approach. By emphasizing computational efficiency, we specifically target the shortcomings of existing neural compression techniques, which are computationally infeasible in constrained in-vehicle embedded systems.

Within the proposed project, the goal is to investigate this tokenization-based neural compression framework by evaluating it against more heavy-weight traditionally used neural compression frameworks. To achieve this, the performance of a subset of industry-relevant downstream ML tasks when trained on uncompresssed data, data compressed using traditional neural compression methods, and data compressed using the tokenization-based neural compression framework, will be evaluated.

The proposal is organized as follows. Section ?? places the discussed topics into the broader context. Section ?? formalizes the problem, reviews related work, and identifies research gaps. Section ?? summarizes open challenges with emphasis on time-series neural compression under embedded constraints and their relevance to industry. Section ?? discusses the motivation behind using tokenization as a potential alternative. Lastly, Section ?? states objectives and constraints, as well as a concrete approach to achieve these goals.

References