

Fundamental Study on Inferring Aerodynamic Characteristics by Large Language Models Using Latent Representations of 3D Shapes

Satoru Ito ^{1), 2)} Koji Nishiguchi ^{1), 2)} Koichiro Nakaya ^{1), 2)} Yuuri Ozaki ^{1), 2)}
 Hiromune Kanamori ³⁾ Shingo Noritake ³⁾ Hiroshi Tanaka ³⁾

*1) Nagoya University
 Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8602, Japan (E-mail: ito.satoru.s4@s.mail.nagoya-u.ac.jp)
 2) RIKEN
 15F, Nihonbashi 1-chome Mitsui Building, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, 103-0027, Japan
 3) Toyota Systems Corporation
 32F, JP Tower Nagoya, 1-1-1 Meieki, Nakamura-ku, Nagoya, Aichi, 450-6332, Japan*

KEY WORDS: aerodynamic performance/aerodynamic noise, CFD, Large Language Model [D2]

To accelerate the aerodynamic design process, developing a 3D-to-Text model is crucial for providing rapid, natural-language feedback on 3D design proposals. Such models read 3D vehicle geometries and output comprehensive evaluations, acting as interactive design-support assistants. To realize this system, LLMs must recognize 3D shapes while preserving physical and geometric consistency accurately. However, directly generating text from 3D shapes is currently challenging because of the severe shortage of paired 3D-geometry and natural-language datasets. To address this bottleneck, we propose partitioning the pipeline into two successive stages: "3D-to-Parameter" and "Parameter-to-Text." This paper specifically focuses on developing the 3D-to-Parameter model.

To achieve this, we propose an approach that compresses 3D car shapes into latent vectors using DeepSDF and directly projects them into the embedding space of a Large Language Model (LLM) to predict aerodynamic performance, specifically the drag coefficient (C_D). Since aerodynamic characteristics are highly sensitive to subtle, non-linear geometric features, we systematically investigate the effect of expanding the dimensionality of DeepSDF latent vectors on the LLM's physical inference accuracy.

First, we utilized the DrivAerNet++ dataset and trained its neural network to convert each 3D shape into a low-dimensional latent vector. DeepSDF is an auto-decoder that jointly optimizes the decoder parameters and a latent code for each training shape. For each test shape, only the latent code was optimized while the trained decoder was kept fixed. DeepSDF's use of a continuous signed distance function (SDF) representation provides a continuous implicit representation without relying on a fixed voxel grid. To examine how differences in the dimension of the latent vector affect the performance of models, we varied the vector dimension in five steps: 40, 80, 120, 160, and 200.

Next, the extracted latent vectors were projected into the embedding space of an LLM (Llama3.2-3B) using a trainable linear layer. The model was fine-tuned to predict the C_D value from the final hidden state of the sequence. The results, as shown in Fig. 1, showed that a trend of the mean relative error rate on the LLM test data decreasing from 5.18% to 3.33% was observed as the latent dimensionality increased under the present experimental conditions.

In conclusion, this study successfully developed a 3D-to-Parameter model architecture as a foundational step toward a comprehensive 3D-to-Text model. We investigated how the dimensionality of the latent vector affects the predictive performance of the LLM-based model. For future work, we plan to improve the model to predict multiple scalar quantities or vector fields and develop a Parameter-to-Text model.

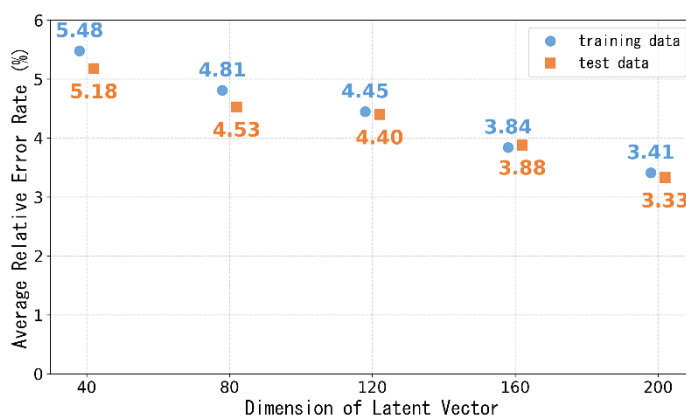


Fig.1 Mean relative errors of C_D predictions for different latent-vector dimensions