A Knowledge Distillation-based Training Framework Improves the Accuracy of Pre-crash Occupant Injury Prediction Models with High Computation Efficiency

Qingfan Wang, Dongyang Xu, Qing Zhou, Bingbing Nie*

I. INTRODUCTION

Data-driven occupant injury prediction has become an essential part of intelligent vehicles because it can provide critical information to trajectory planning systems (during pre-crash), adaption restraint systems (during in-crash), and advanced automatic collision notification systems (during post-crash) [1-2]. Considering the time domain, different available information is used for the relevant injury prediction tasks. For example, with limited pre-crash warning time, concise scalars that describe the collision condition (e.g. delta-*v* and belt usage) are often used as inputs for pre-crash prediction [3]. More information (e.g. crash pulses) is available during post-crash, which is also more relevant to occupant injuries. High-complexity deep learning methods are generally applied to process such time series and have obtained higher accuracy, despite their low computation efficiency [4].

Existing pre-crash and post-crash injury prediction models were developed separately. Whether it is feasible to combine their merits to obtain high prediction accuracy and high computation efficiency remains unknown. Knowledge distillation (KD) is an emerging model compression method that trains a lightweight model ('student') using the supervision signals provided by a larger and better model ('teacher') [5]. Motivated by this teacher-student training mode, we proposed a KD-based training framework to improve a pre-crash occupant injury prediction model by making it learn from a post-crash counterpart.

II. METHODS

Technically, the KD-based training was divided into two steps (Fig. 1). First, we trained a teacher network (T-Net) for post-crash prediction with high-complexity architecture to obtain high accuracy. Then, receiving the extra 'knowledge' provided by its teacher, a high-efficiency lightweight student network (S-Net) was trained by using both the traditional prediction loss L_{pred} and the distillation loss L_{dist}. That is, it learned not only from the ground truth but also from its teacher. Note that the auxiliary T-Net is only used in the offline training process of the S-Net; this means the S-Net can be applied online and provide real-time pre-crash injury prediction independently.

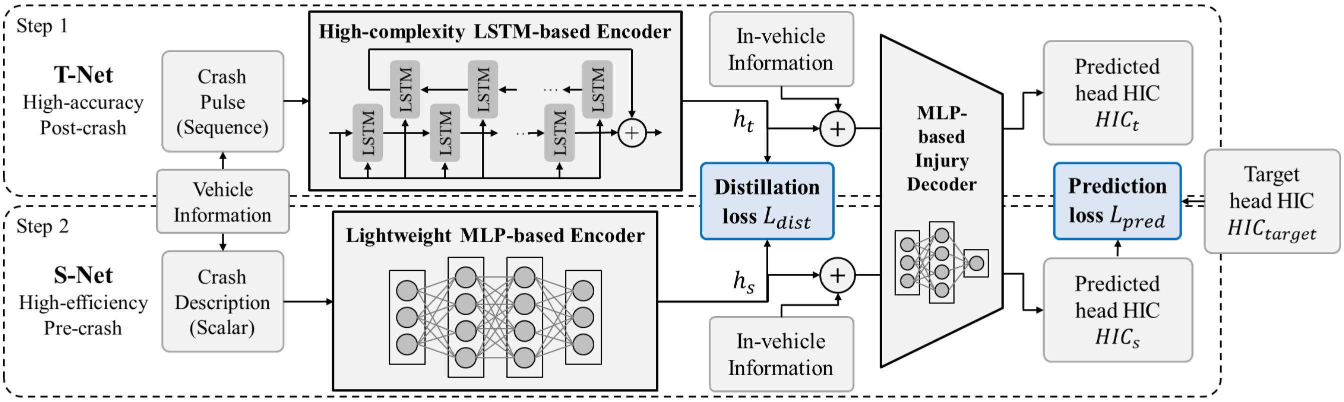

Fig. 1. The proposed knowledge distillation-based training framework of pre-crash injury prediction models.

T-Net. Long- and short-term memory (LSTM), a classic time series processing module, was used as the encoder to mine useful information hidden in crash pulses. Together with the in-vehicle information (e.g. restraint configurations and occupant characteristics), the encoder's output (h_t) was fed into the injury decoder implemented by a multilayer perceptron (MLP). We selected the head injury criterion HIC_t as the final output.

S-Net. Due to the inaccessibility of the informative crash pulses at the pre-crash stage, we had to use crash descriptions in the scalar form as a substitute. Since scalars are easier to process, we replaced the complex LSTM with a simple two-layer MLP, effectively reducing the computational burden. The remaining parts of the S-Net were the same as the T-Net. During the training, the T-Net's parameters were frozen, and the S-Net not only

B. Nie (e-mail: nbb@tsinghua.edu.cn; tel: +86-10-6278-8689) is an Associate Professor, Q. Wang is a PhD student, D. Xu is a MSc student, Q. Zhou is a Professor, all in the School of Vehicle and Mobility at Tsinghua University and State Key Laboratory of Intelligent Green Vehicle and Mobility, China.

managed to make its output HIC_s closer to the target HIC_{target} but also received the supervision signals h_t from its teacher's LSTM-based encoder to learn how to correctly process the vehicle information in the intermediate layers, which can be seen as the 'distilled knowledge'. Thus, the total loss L_{all} is represented as:

$$L_{all} = L_{pred} + \alpha \cdot L_{dist} = MSE\left(HIC_{target}, HIC_s\right) + \alpha \cdot MSE(h_t, h_s)$$

where $MSE(\cdot,\cdot)$ denotes the mean square error, and α is a weight factor optimised by prediction performance.

III. INITIAL FINDINGS

Based on a previously developed 5,000-case numerical crash dataset [6], we compared the performance of the following models to verify the KD-based framework: (1) T-Net; (2) S-Net without KD; and (3) S-Net with KD. Besides regression metrics for HIC prediction, we also translated HIC into head AIS to introduce classification metrics. As a result (Table I), the T-Net outperformed the S-Net without KD in all the accuracy metrics despite its longer prediction time (all tested on Intel i7-9700k 3.60GHz). Specifically, the S-Net without KD performed worse in high-risk cases, e.g. cases with HIC higher than 1400 or AIS in level 5 (Fig. 2). In contrast, KD training improved the S-Net's prediction accuracy from 55.8% to 60.6%, even slightly higher than its teacher's (59.6%). Most predicted cases of the S-Net with KD were distributed on or around the diagonals of both the HIC scatter diagrams and AIS confusion matrices, and it could predict higher HIC for high-risk cases (Fig. 2). More importantly, the high computation efficiency was reserved as KD training did not change the S-Net's architecture.

TABLE I

PREDICTION PERFORMANCE OF THE THREE MODELS

	HIC: RMSE, MAE, R2	AIS: Accuracy, G-mean, F1	Online prediction time
T-Net	245.7 160.7 0.696	59.6% 0.504 0.593	20.0 ± 5.1 ms
S-Net w/o KD	269.7 166.4 0.634	55.8% 0.000 0.551	2.4 ± 0.5 ms
S-Net w/ KD	258.7 158.5 0.663	60.6% 0.522 0.602	2.4 ± 0.6 ms

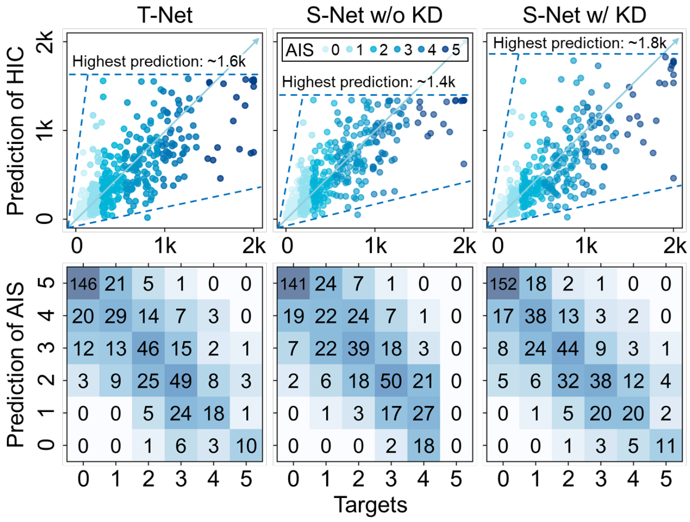

Fig. 2. HIC scatter diagrams and AIS confusion matrices.

IV. DISCUSSION

Prediction accuracy and computation efficiency are two critical but traditionally contradictory factors for injury prediction. This issue is especially challenging during pre-crash with limited information and time. This study proposes to solve it via the KD-based framework. The teacher-student training mode skilfully caters to the characteristics of the two injury prediction tasks: 'experienced' high-complexity T-Net for post-crash prediction and high-efficiency lightweight S-Net for pre-crash prediction. Interestingly, the S-Net with KD outperformed the T-Net in accuracy. Intuitively, it is difficult for a student to surpass its teacher due to the natural information gap, i.e. the S-Net's simple scalar inputs are less informative than the T-Net's sequence inputs. However, by optimising the losses in both intermediate and output layers, KD training divided a complex injury prediction task into smaller pieces and instructed the S-Net step by step. Such multilevel loss function design was tailored to pre-crash injury prediction models and eventually contributed to improved performance. However, this preliminary study only validated this framework on a HIC prediction task with numerical data. More efforts are needed to test it in real-world datasets with multiple injury metrics to predict, eventually taking this technology into real applications.

V. ACKNOWLEDGEMENTS

This study is supported by the National Natural Science Foundation of China (52131201, 52072216).

VI. REFERENCES

[1] Stitzel, J. D., *et al.*, *J Am Coll Surg*, 2016.
[2] Bance, I., *et al.*, *IRCOBI*, 2019.
[3] Wang, Q., *et al.*, *iScience*, 2022.
[4] Wang, Q., *et al.*, *Accid Anal Prev*, 2021.
[5] Gou, J., *et al.*, *Int J Comput Vis*, 2021
[6] Chen, W., *et al.*, *IRCOBI Asia*, 2020.