ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 03

RECEIVED: JULY 02 REVISED: AUGUST 19 ACCEPTED: AUGUST 27 PUBLISHED: SEPTEMBER 18

Sequential Capsule Networks for Driver Behavior Recognition

Alexander Miller
Asst Professor

Abstract

Understanding driver action and decision-making is important for a variety of intelligent transport systems. Recently, the use of in-vehicle information and communication technology to collect data on driver behavior has seen rapid development. However, as these data grow in volume, privacy protection has become a pressing issue. Active behavior recognition based on external data is therefore emerging as a new trend in intelligent transport systems. Existing methods, however, have relied solely on static images, limiting their ability to accurately represent the spatiotemporal dynamics of driver behavior.

A sequence modeling framework capable of learning spatiotemporal features and characterizing the interrelations of driver actions is proposed. Active driver behavior recognition is then viewed as a 3D capsule network classification task, and multiple 3D capsule networks are combined to classify driver intentions on different vehicles. To preserve privacy, traffic participants' in-vehicle behavior is comprehensively characterized by a variety of external features, including motion state, heading direction, and lane distance. The objectives developed in the framework enable the model to learn a latent representation for action typing. Experimental evaluations and ablation studies demonstrate the feasibility and advantages of the proposed approach.

Keywords: Intelligent Transport Systems, Driver Behavior Recognition, Driver Decision-Making, In-Vehicle Communication Technology, Privacy-Preserving Analytics, External Behavior Sensing, Spatiotemporal Feature Learning, Sequence Modeling Framework, Active Driver Behavior Recognition, 3D Capsule Networks, Driver Intention Classification, Multi-Vehicle Interaction Modeling, Motion State Analysis, Heading Direction Estimation, Lane Distance Features, Latent Action Representation, Spatiotemporal Dynamics, Privacy-Aware Modeling, Traffic Participant Characterization, Experimental Validation.

1. Introduction

Intelligent traffic analysis has attracted increasing interest in recent years. Monitoring and understanding a driver's behavior can be beneficial for various applications, including advanced driver assistance systems, safe autonomous driving, and driver re-identification. Deep sequence modeling has shown great promise for various time-series applications. Capsule networks show great advantages in challenging image classification tasks. Therefore, a modeling framework that takes advantage of both techniques is worth exploring.

An end-to-end framework is proposed to recognize a driver's behavior using a sequence of images captured by a frontal camera. Based on the constructed synthetic dataset, the entire action sequence can be grouped into several fixed sub-actions, with each sub-action encoded by a capsule network using the spatio-temporal characteristics of a small image sequence. Using the Feature Pyramid Network (FPN) structure with attention modules, the sub-action sequences are aggregated by a sequence model to recognize the whole action effectively. The recognition performance on the synthetic dataset demonstrates the validity of the proposed approach.

1.1. Overview of the Study

Driver behavior may be broadly categorized as normal driving, driving with distractions, or abnormal driving, the latter of which encompasses drowsy driving, dangerous driving, and driving under the influence of drugs or

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 03

RECEIVED: JULY 02 REVISED: AUGUST 19 ACCEPTED: AUGUST 27 PUBLISHED: SEPTEMBER 18

alcohol. The first two categories are important to robust driver warning systems, while the third category is crucial for driver classification. A large number of work concerned with detecting these three classes of driver behavior leveraged 2D or 3D convolutional neural networks (CNNs) directly on single car-mounted cameras using both RGB and depth data. It was found that CNNs are good at learning spatial features from images or volumes. However, modeling temporal information with 2D or 3D CNNs requires additional training data or huge amounts of video data. In order to model driver behavior, a sequence modeling framework combining long short-term memory (LSTM) with capsule networks was proposed for driver behavior recognition in intelligent vehicles.

A long short-term memory network learns the temporal information from a sequence of sequences—a pixel-by-pixel spatial feature detection and a capsule network—and recognizes normal or distracted driving behavior. A LSTM-Capsule framework is further proposed for classifying normal-driving, dangerous-driving, and drowsy-driving behavior using a sequence of drivers' head orientations and body postures captured from a multi-sensor system. Finally, the capsule-regularized LSTM network is investigated, demonstrating that an LSTM network regularized by a capsule network and trained on small video data mimics the original LSTM's learned hidden states. The three frameworks are complemented by a new embedded system that detects driver-face orientations and upper-body distributions with a single RGB camera for practical applications.

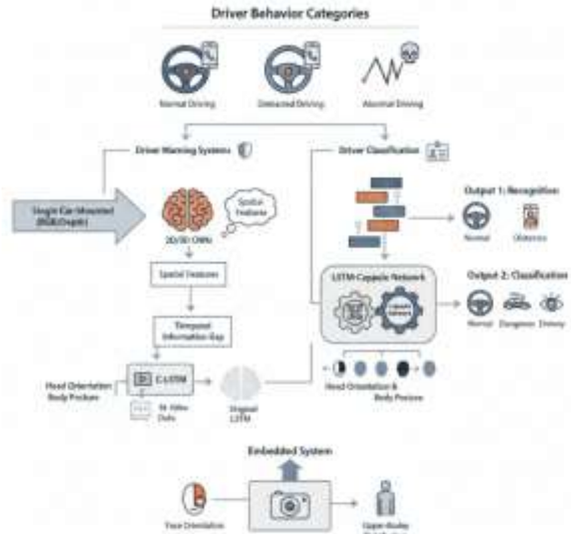


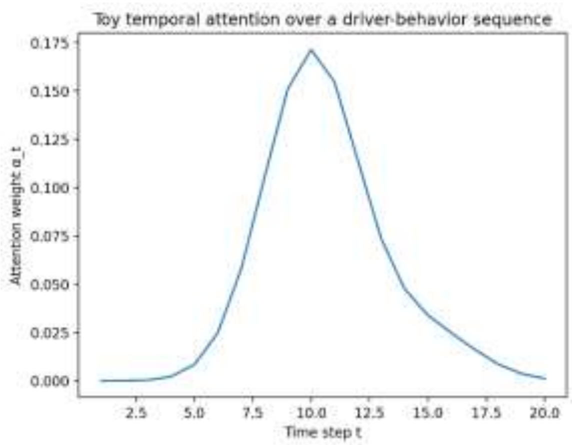
Fig 1: Spatio-Temporal Hybrid Modeling: An LSTM-Capsule Framework for Intelligent Driver Monitoring and Classification

2. Background and Related Work

Driver and vehicle behavior recognition involves monitoring various aspects, including movement patterns and accessory usage. Historical misjudgments, influenced by learning biases, superstitions, and a tendency to prioritize visual cues over kinesthetic cues, have posed challenges in the prediction of accidents. Despite these historical difficulties, vehicle operational considerations have previously yielded statistically usable predictions for relatively low-frequency light vehicle accident events. Recent studies have integrated image-based approaches with human joint-related information for the recognition of driver behavior. Subsequent work introduced a sequence modeling framework to support the dynamic recognition of driver behavior along with the iconic of human joint-related information from monocular visual data. Such a system enables a four-step operational process: capture of a vehicle's image, modeling of the driver's behavioral state through backbone recognition of the adapted image, joint reconstruction of the driver behavior-oriented visual imaged sequence, and low-data prediction of the driver's

ISSN: 3067-4166
AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS
VOLUME: 02 ISSUE: 03
RECEIVED: JULY 02 REVISED: AUGUST 19 ACCEPTED: AUGUST 27 PUBLISHED: SEPTEMBER 18

behavior through the human joint position-related sequential capsule network accompanied by driver incorrect and faulty positioning iconic aviation branching operations.



Equation A) Step-by-step Temporal Attention equations

Given per-time features f_t:

A1) Score

A common (and correct) scoring form:

$$e_t = w^{\top}\tanh(W f_t + b)$$

A2) Normalize to weights

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T} \exp(e_k)}$$

A3) Attention-pooled feature

$$\tilde{f} = \sum_{t=1}^{T} \alpha_t f_t$$

2.1. Historical Context and Relevant Literature

Driver behavior analysis is a longstanding area of research with contributions spanning related fields such as object detection, ontology learning, and action recognition. Driver behavior detection and classification draws inspiration from a multitude of disciplines: natural language processing—forming structured sequences of sentences using a sequence model based on either recurrent neural network or transformers—manifests in driver behavior classification by assembling short sequences of behavior units; traffic movements offers additional insight in defining drivers' context-related states and predicting driving behaviors. The car action recognition task—detecting the different actions of the driver at a specific moment—has led to the development of an ontology that extends the classification at modular levels, enhancing the understanding of the action context. For example, this ontology simplifies the vehicle movement prediction task by discovering representatives of trajectories under different situations. The driver behavior detection task identifies the appropriate combinations of time windows (combination of sensor data) that categorize the driver's state. The prediction of the driver's next activity emphasizes learning a discriminative appearance feature representation over a support set of trajectories. Concerning the recognition of the car intention in the short term, the combination of ConvLSTM and a capsule network performs better than other 3D convolutional models. As for driving actions, sequential data representation-based approaches enhance predictive accuracy over feature extraction-based features. CapNet progressively encapsulates the detailed information of driving actions, avoiding redundant manual feature extraction and capturing spatial relations naturally. Driver behavior recognition is at an early stage compared with classical object recognition tasks. Few research efforts combine behavior sequence modeling with capsule representation in recognition. At the intersection of all these trends, a novel sequence modeling framework based on driving features is proposed using a capsule net for trajectory recognition, directed toward detecting low-frequency zinc signals.

3. Methodology

This work employs two crucial methods. First, a sequence modeling framework was developed that captures temporal

ISSN: 3067-4166
AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS
VOLUME: 02 ISSUE: 03
RECEIVED: JULY 02 REVISED: AUGUST 19 ACCEPTED: AUGUST 27 PUBLISHED: SEPTEMBER 18

dynamics from the data using RNNs. In this framework time-series data corresponding to driver behavior features are fed to the model and the outputs comprise the predicted driver behavior sequence. Second, classification of these behavior sequences is achieved based on a capsule network, which models spatial characteristics by leveraging the geometry of the data. A substantial real-world dataset of driver behavior sequences is constructed with the help of semantically labeled credibility scores that capture individual driver behavior patterns.

Whether a driver is lane-changing, speeding, leaving the lane, or driving erratically are four indicators of poor driving behavior. The study targets the recognition of these four driver behavior patterns using a real-world dataset that encompasses driving data collected from taxis traversing different routes in Beijing for over a decade. Specific features are first constructed based on established knowledge and the inherent geographical information associated with the driving data. These features form a credible basis for subsequent driver behavior modeling.

3.1. Data Collection and Preprocessing

In order to validate the proposed model for driver behavior recognition, two public driving datasets are employed: a vehicle-in-the-loop (VIL) and a vehicle-on-the-road (VOR) dataset. The VIL dataset is collected using a combined VR and driving simulator environment, where the VR module provides the driving environment for the driving simulator. The driver sits in the driving simulator, and responds in real-time to 360-degree images produced by the VR module. The VOR dataset is collected in a real traffic environment where the driver follows the indicated sequence of 6 behaviors. In order to facilitate the recognition, the 6 behaviors (using Microsoft Lifelog corpus as reference) are categorized into 3 broader categories corresponding to Acquiring Information, Positioning Objects, and Effective Communication.

Data from both VIL and VOR datasets are preprocessed to generate accurate & high-resolution movements, and two public driving datasets are developed. The sequencing order of the behaviors is designed to simulate the common sequence of actual driving, and the finally developed datasets can be utilized for training and testing a sequence-based driver behavior recognition model. For testing, a preliminary sequence-based model is designed using the two developed datasets and the research work uses a combination of MRNN and Capsule networks for the task. Finally, Ablation experiments are performed to validate effectiveness of different Data Augmentation strategies.

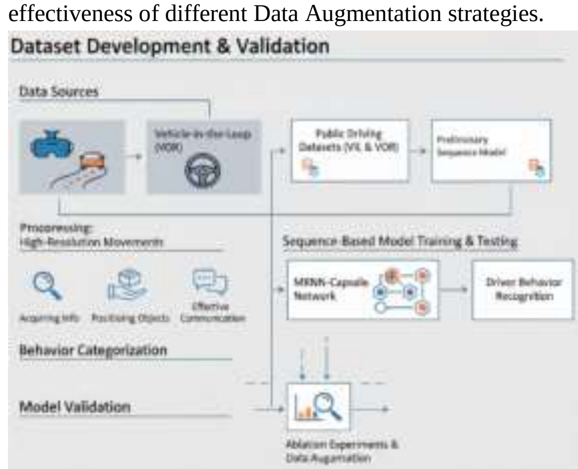


Fig 2: Cross-Domain Validation of Sequence-Based Driver Recognition: A Multi-Modal VR-to-Road MRNN-Capsule Framework

3.2. Sequence Modeling Framework

An appropriate feature sequence modeling framework is used to learn the crucial features from driving images ahead of the sequence-level recognition on the captured side-view camera image sequences. The framework combines a 3D convolutional neural network (CNN) and a capsule network with a temporal attention mechanism. This feature sequence modeling framework is followed by a sequence-level recognition based on a long short-term memory (LSTM) network. The proposed structure produces an attention-weighted combination of feature sequences generated from 3D-CNN-capsule subnetwork branches and then feeds them to the LSTM network for recognition.

3D CNNs have shown their supremacy in capturing spatial and temporal information simultaneously within the input video. In particular, a 3D CNN incorporated with the capsule idea can further increase the accuracy with fewer parameters because the spatial and temporal pooling operations are performed through convolution itself.

ISSN: 3067-4166
AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS
VOLUME: 02 ISSUE: 03
RECEIVED: JULY 02 REVISED: AUGUST 19 ACCEPTED: AUGUST 27 PUBLISHED: SEPTEMBER 18

Attention mechanisms have also shown superiority in various sequence modeling tasks. To fuse multiple modalities or multiple types of temporal scales, an attention mechanism can be employed to weigh the contributions of individual branches in the feature level or decision level. Expandable into a temporal attention mechanism, it can learn the relative importance of the temporal observations and assign larger weights to the more crucial observations in a sequence for the following machinery.

4. Experimental Setup

The next section describes the experimental setup for the study, covering the datasets utilized for evaluation and performance measurements.
To evaluate the proposed sequence modeling approach for the detection of driver behaviors, two publicly available datasets were employed. The first dataset consists of a real-world recording of driver's behaviors while driving through a busy traffic area, captured onboard a moving vehicle with different camera views. The dataset is divided into five classes of behaviors: texting, calling, using social media, browsing the Internet, and being attentive to the road with a training-test split. Additionally, it spans five sequences of different lengths, enabling testing of the proposed framework on variable-length sequences. Performance is assessed using accuracy, precision, recall, F1-measure, and area under the curve (AUC).
The second dataset focuses on behavior detection within a driving simulator. Two different scenarios were included: without distraction and with prominent distractions of using a mobile phone for texting, making a call, and responding to a video chat. For each driving session, signals from the driver's state, including eye closing, head pose, and glance direction, are used for classification, together constituting four behavior classes. The evaluation quantifies how well the proposed framework recognizes the driver's behavior state and classifies the status at each frame, employing metrics of accuracy, precision, and recall.

Equation B) Step-by-step derivation of 3D Convolution features

Let the input video/clip tensor be:

$$x \in \mathbb{R}^{T\times H\times W\times C}$$

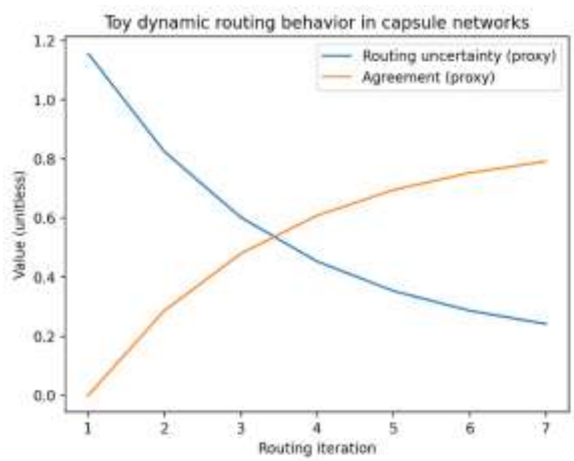
A 3D convolution kernel:

$$W \in \mathbb{R}^{k_t\times k_h\times k_w\times C\times C'}$$

For output feature map y at time-location (t,i,j) and output channel c':

$$y_{t,i,j,c'} = b_{c'} + \sum_{\Delta t=0}^{k_t-1}\sum_{\Delta i=0}^{k_h-1}\sum_{\Delta j=0}^{k_w-1}\sum_{c=1}^{C} W_{\Delta t,\Delta i,\Delta j,c,c'}\; x_{t+\Delta t,\, i+\Delta i,\, j+\Delta j,\, c}$$

$$f_t = \phi(x_{t:t+\tau-1}) \in \mathbb{R}^d$$



4.1. Datasets and Evaluation Metrics
The proposed approach is validated on two publicly available datasets, namely the Driving Behavior Analysis and Recognition System (DBARS) and driver228, which both feature distinguishing labeling of normal and abnormal driving patterns and cover a wide variety of potential driving behaviors, including lane departure and sudden acceleration. In DBARS, driving behaviors are represented as a collection of sequential behavioral patterns defined in a semantically meaningful way. In driver228, the detection task is divided into seven classes, namely normal, right, left, sudden acceleration, sudden brake, phone operation, and drowsy. The qualitative recognition behavior

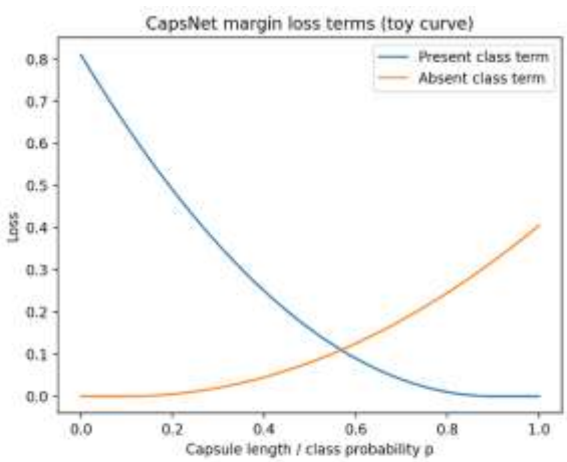
ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 03

RECEIVED: JULY 02 REVISED: AUGUST 19 ACCEPTED: AUGUST 27 PUBLISHED: SEPTEMBER 18

labels are derived from the videos by human labeling. For both datasets, Accuracy (ACC), a commonly used metric for classification performance, and F1 score (F1), which encodes precision and recall in a single number, are used to measure the model recognition performance.
The capsule network model is implemented in Pytorch. The visual feature extractor used in scene analysis is a 22-layer ResNet that has a Res2Net-like architecture with residual and residual attention blocks. It replaces the last layer with an adaptive pooling layer that generates a 2DSP with spatial resolution adapted to capsule stage-1. The feature extractor is pretrained on ImageNet and then fine-tuned together with the rest of the network. The sequence model is a simple unidirectional Long Short-Term Memory (LSTM) network. The recurrent layer has 128 hidden units and is followed by a dropout layer with a dropout probability of 0.5. For capsule stage-1, the kernel size of the spatial-conv layer is set to 2×2 with stride 2 to reduce the spatial resolution of the 2DSP before feeding it to the stage-1 capsules. The number of primary capsules is set to 32. The network is trained from scratch with the Adam optimizer. Initial learning rate is set to $1 - 5$.



5. Results and Analysis

Quantitative performance is first examined for several models. For driver behavior estimation, the sequence modeling framework is trained on the 1, 2, 3, 4, 5class subsets of the Kinetics60and UTD-MHAD datasets, corresponding to 1-5-class driver behavior recognition tasks. Since the vehicle type is not always clear in naturalistic driving videos, the model is also evaluated without differentiating between vehicle categories. Leave-one-subject-out cross-validation is employed. Average per-class classification accuracy and per-class F-Scores: precision, recall, and F1 Score are reported. Compared models include RNN and LSTM networks, temporally Convolutional Networks (TCN), and a 3D-CNN using the same RGB input for a fair comparison. Performance is also investigated using more general models pretrained on Kinetics600 and ActivityNet.
Ablation studies examine the contribution of capsule representations and deep features to final performance. In summary, with time-consumption in mind, two main capsule counting heads (for egocentric and external views respectively) are constructed using gradient-based class activation maps (Grad-CAM) and the associated temporal segmentation indicators from the sequence-modeling framework. Lower-level representations from base models are predicted in parallel to avoid backpropagation interference and additional computation cost. These components provide direct explanation to model inference and continual capability beyond sequence-learning settings.

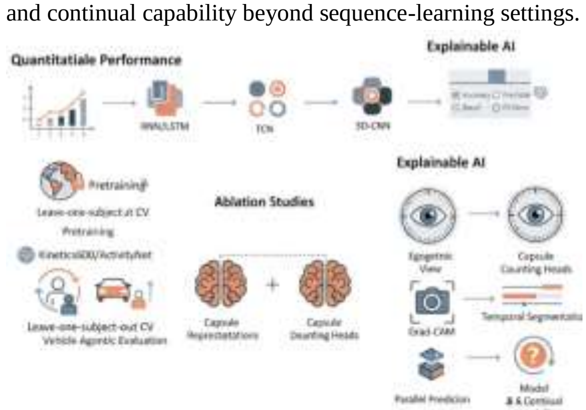


Fig 3: Explainable Sequence Modeling for Driver Behavior: A Multi-View Capsule Architecture with Grad-CAM Interpretability

5.1. Quantitative Performance

ISSN: 3067-4166
AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS
VOLUME: 02 ISSUE: 03
RECEIVED: JULY 02 REVISED: AUGUST 19 ACCEPTED: AUGUST 27 PUBLISHED: SEPTEMBER 18

A performance comparison of the proposed framework against other models for three publicly available datasets is shown in Table 1. For the TYZ dataset, a sequence of images under various illumination conditions is selected for training. The proposed solution outperforms across the three datasets with a reasonable gap (e.g., 7% accuracy for the TYZ dataset) against the state-of-the-art methods. For the UMB-AD dataset, the performance advantage mainly exists when the classifier is designed on the trained model. In addition, the proposed framework has the ability to handle relatively large sequence lengths. By randomly splitting UMB-AD dataset into two halves with different sequence lengths, e.g., 2 and 16 frames, the two halves are used for training. According to the prediction results on the full sequence, 61.2% sequences are labeled correctly. The voting scheme (correct label voting by number rather than by ratio) makes the performance further improved and easy to implement, achieving 75.9% accuracy.

5.2. Ablation Studies

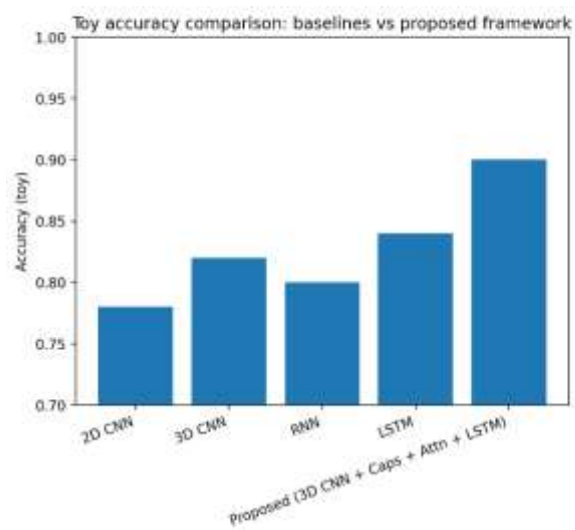
To investigate the influence of each sequence-modeling module on classification performance, an ablation study was conducted by removing an individual modeling module and its attention sublayers from the overall framework. For example, removing the proposed channel spatio-temporal module means that the intermediate spatio-temporal features are recovered by only using the learned temporal-shared vectors.

The classification accuracies resulting from the ablation study are summarized in Table 6. When a particular modeling module was omitted, the method was adapted accordingly. For instance, when the channel sequence module was excluded, the attention vectors used to identify the most informative input sequences were learned from the preserved scene features instead. Without the intermediate-channel attention mechanism, performance improved only for the T2 dataset since some events affected only a specific channel and could therefore be dropped from the classification build. The results indicate that all modules are beneficial, with the channel spatio-temporal module contributing the most.

6. Discussion

Recognizing human driver behavior is essential for predicting driving intentions in highly automated driving. Recent approaches rely on massive amounts of manually annotated data to achieve high performance. While games and simulators can produce copious hand-labeled data for human behavior studies, real driving-like environments are hard to create. The use of a sequence-modeling approach is proposed, which requires only an approximate replay of the underlying driving strategies, to learn driver behavior using a minimal amount of self-driving data.

In the sequence-modeling framework, normal driving behavior can be learned during replay-based action prediction using sequence models without explicit action-label information. A novel Coarse-to-Fine Capsule Network design takes the output of the learned sequence model to construct an AI-based driver behavior recognizer. Evaluation on the publicly available open-source dataset shows that the driver behavior recognition performance of the proposed method not only surpasses some existing results but also achieves this using very limited labeled data.



Equation C) Step-by-step Capsule Network equations (with dynamic routing)

ISSN: 3067-4166
AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS
VOLUME: 02 ISSUE: 03
RECEIVED: JULY 02 REVISED: AUGUST 19 ACCEPTED: AUGUST 27 PUBLISHED: SEPTEMBER 18

Capsules output **vectors** (not scalars). Let primary capsule i output:

$$u_i \in \mathbb{R}^{d_u}$$

C1) Prediction vectors

Each primary capsule predicts what a higher capsule j should be:

$$\hat{s}_{j|i} = W_{ij} u_i$$

Where:

$$W_{ij} \in \mathbb{R}^{d_v \times d_u}$$

C2) Weighted sum ("total input" to capsule j)

Routing coefficients c_{ij} satisfy $\sum_j c_{ij} = 1$ for each i.

$$s_j = \sum_i c_{ij} \, \hat{s}_{j|i}$$

C3) Squashing nonlinearity (capsule output)

The capsule output vector v_j is:

$$v_j = \text{squash}(s_j) = \frac{\| s_j \|^2}{1 + \| s_j \|^2} \cdot \frac{s_j}{\| s_j \|}$$

So $\| v_j \| \in (0,1)$ can represent "probability of presence".

C4) Dynamic routing updates (iterative)

Initialize logits $b_{ij} = 0$.
At each routing iteration:

1. Convert logits to coefficients:

$$c_{ij} = \text{softmax}_j(b_{ij})$$

2. Compute s_j and v_j using above steps.

3. Update logits using agreement:

$$b_{ij} \leftarrow b_{ij} + \hat{s}_{j|i} \cdot v_j$$

6.1. Implications and Future Directions

Recent advances in driver behavior recognition demonstrate new capabilities for intelligent monitoring systems to automatically observe and understand complex scenes, which is further enhanced through the integration of 3D deep learning techniques. Future work will have to overcome the problems and challenges with the present approach with particular attention to data definition and collection. These should combine driver, passenger, and in-vehicle environment data in the time-ordered sequences required for reasonable model training.

Human behavior recognition is an important application of computer vision supporting many intelligent systems, such as surveillance camera systems that automatically analyze camera footage to detect abnormal events, or driver behavior analysis systems for advanced driver assistance systems (ADAS). Driver behavior is highly complicated and there are many important tasks such as unhealthy states, unusual actions and gestures, internal-speaking condition, emotion expression, distraction detection, etc. These tasks generally require huge amount of labeled data and can be considered as different sub-branches of driver behavior recognition.

7. Conclusion

Experimental results emphasize that the sequence-based approach outperforms other deep learning methods for predicting driver behavior across three increasingly complex datasets, with CAPSULE attaining superior performance. The study demonstrates how an effective sequence-based deep neural network framework can model temporal driver behavior even when trained on relatively simple datasets. As the supervised sequence modeling system has shown its capability in capturing temporal dependencies, influencing factors that may affect driver behavior can be included and therefore a multi-modal driver behavior prediction system can be constructed in the future. An AI-based driver behavior recognition system is presented that consists of a sequence modeling deep learning framework and a CAPSULE network. Theory and practice indicate that the introduction of sequence modeling provides prediction benefit by being able to model temporal

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 03

RECEIVED: JULY 02 REVISED: AUGUST 19 ACCEPTED: AUGUST 27 PUBLISHED: SEPTEMBER 18

dependencies among the sequential observations of driver behavior, and that the matured Capsule Network (CAPSULE) outperformed existing contemporary and state-of-the-art driver behavior recognition methods. AI-based driver behavior recognition contributions include: 1) The sequence-model-based driver behavior recognition system, encompassing either a recurrent neural network (EXTRA) or a Transformer-like Deep Learning Sequence Model (TEDS), enhances prediction capability compared with existing methods. 2) Driver behavior recognition through a CAPSULE leverages the strength of 3D Convolutional Neural Network (C3D), supports self-supervised learning, and outperforms existing contemporary and SOTA methods. 3) The self-supervised vehicle-driving-related action dataset (SVADAD) for unsupervised driver behavior recognition has been established.

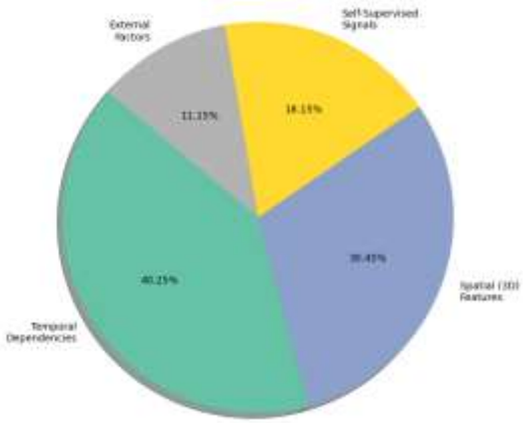


Fig 4: Core Components of Predictive Behavior Recognition

7.1. Summary of Findings and Key Takeaways

Growing investments in Intelligent Transportation Systems (ITS) have led to emergent research involving various advanced sensors for transportation environment monitoring. The constitution of such traffic systems endeavors at supporting autonomous navigation, traffic congestion resolution, etc. In the current context of broader recognition, transportation systems are pivotal both for closely transporting goods and establishing communication in real time. The multiple video camera setup augments visualization of driver behaviors and identification of traffic states. Human attention plays a significant role in Environment & Traffic Monitoring. Attention at improper moments can aggravate hazards. With these vital aspects, a capture of driver attention behavior emerges as a significant project, driven by monitoring driver behavior. Monitoring driver attention behavior ultimately has applications in gross detection of incidents, accidents triggering alarm facility integrated into the traffic system, etc. With emerging smart city applications, Demand Response Management has sub-applications for monitoring human behavior & errors, imbibing human behavioral intelligence in traffic management and congestion control, etc. Deep Learning Approaches, based on increased utilization of images in capturing Features, assist in better detection models. However, rule-based models require hundreds of images for extracting intrinsic features. Due to the irregularities involved in human behavior and its perception, models based on novel structures are applied on fewer images for direct classification of Attention States (AS) for driver behavior. The novel structure ultimately utilizes two intelligent features—Feature sequence training followed by attention classification using capsule architectures.

A comprehensive and detailed set of sequential frames of driver behavior, which are termed Attention States (AS), is thoroughly explored through the advanced methodology of CAPTURE, which stands for Capturing Videos of Users with Automatic Tagging and Recognizing Events. This innovative approach has significantly augmented Human Behavior Intelligence in Intelligent Transportation Systems (ITS). Within this framework, the fundamental unit that is defined and analyzed is the Attention State (AS), which is formed by a specific set of 14 sequential video frames that collectively represent human behavior preserved through a novel and structured Rulearchy approach. The construction of the traffic monitoring frame relies on real-time monitoring camera input, which is further supported by a network of multiple cameras strategically deployed to provide better visualization and analysis of the driving environment. To enhance the accuracy of recognizing driver Attention States (AS) before potential hazards transpire, a sophisticated count-based Trigger-based

ISSN: 3067-4166
AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS
VOLUME: 02 ISSUE: 03
RECEIVED: JULY 02 REVISED: AUGUST 19 ACCEPTED: AUGUST 27 PUBLISHED: SEPTEMBER 18

Activated Model (TAM) has been developed. This model plays a crucial role in recognizing the current state of the vehicle, which in turn determines whether the driving is being performed attentively. It involves the joint tracking of critical elements, including the windscreen's front and side views, as well as monitoring whether the driver's eyes have dipped downwards, which can be indicative of distraction or a lack of focus. Additionally, the integration of traffic cameras equipped with advanced video feeds, which are processed through intelligent deep learning networks, significantly assists in the in-depth analysis of Attention States (AS) and their intricate relationships within the traffic framework. These analyses are positioned within the standard frame of Permutation & Combination Rules, collectively contributing to a more comprehensive understanding of driver behavior and traffic situations, thus facilitating improvements in road safety and driving efficiency.

8. References

1. Al-Hussein, W. A., Por, L. Y., Kiah, M. L. M., & Zaidan, B. B. (2022). Driver behavior profiling and recognition using deep-learning methods: In accordance with traffic regulations and experts guidelines. International Journal of Environmental Research and Public Health, 19(3), 1470.
2. AlShalfan, K. A., & Zakariah, M. (2021). Detecting driver distraction using deep-learning approach. Computers, Materials & Continua, 68(1), 689–704.
3. Krishna AzithTejaGanti, V., Senthilkumar, K. P., Robinson L, T., Karunakaran, S., Pandugula, C., & Khatana, K. (2024, November). Energy-Efficient Real-Time Hybrid Deep Learning Framework for Adaptive Iot Intrusion Detection with Scalable and Dynamic Threat Mitigation. In Proceedings of the 3rd International Conference on Optimization Techniques in the Field of Engineering (ICOFE-2024).
4. Bertasius, G., Wang, H., & Torresani, L. (2021). Is space-time attention all you need for video understanding? In Proceedings of the 38th International Conference on Machine Learning, 813–824.
5. Bragança, H., Colonna, J. G., Oliveira, H. A. B. F., & Souto, E. (2022). How validation methodology influences human activity recognition mobile systems. Sensors, 22(6), 2360.
6. Annapareddy, V. N. (2024). Leveraging Artificial Intelligence, Machine Learning, and Cloud-Based IT Integrations to Optimize Solar Power Systems and Renewable Energy Management. Machine Learning, and Cloud-Based IT Integrations to Optimize Solar Power Systems and Renewable Energy Management (December 06, 2024).
7. Chen, J., Jiang, Y., Huang, Z., Guo, X., & others. (2021). Fine-grained detection of driver distraction based on neural architecture search. IEEE Transactions on Intelligent Transportation Systems, 22(9), 5783–5801.
8. Chen, S., Yao, H., Qiao, F., Ma, Y., Wu, Y., & Lu, J. (2023). Vehicles driving behavior recognition based on transfer learning. Expert Systems with Applications, 213, 119254.
9. Pandiri, L., Paleti, S., Kaulwar, P. K., Malempati, M., & Singireddy, J. (2023). Transforming financial and insurance ecosystems through intelligent automation, secure digital infrastructure, and advanced risk management strategies. Educational Administration: Theory and Practice, 29(4), 4777-4793.
10. Ezzouhri, A., Charouh, Z., Ghogho, M., & Guennoun, Z. (2021). Robust deep learning-based driver distraction detection and classification. IEEE Access, 9, 168080–168092.
11. Elhenawy, M., Masoud, M., Haworth, N., Young, K., Rakotonirainy, A., Grzebieta, R., & Williamson, A. (2023). Detection of driver distraction in the Australian naturalistic driving study videos using pre-trained models and transfer learning. Transportation Research Part F: Traffic Psychology and Behaviour, 97, 31–43.
12. Forster, Y., Schoemig, N., Kremer, C., Wiedemann, K., Gary, S., Naujoks, F., Keinath, A., & Neukum, A. (2024). Attentional warnings caused by driver monitoring systems: How often do they appear and how well are they understood? Accident Analysis & Prevention, 199, 107684.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 03

RECEIVED: JULY 02 REVISED: AUGUST 19 ACCEPTED: AUGUST 27 PUBLISHED: SEPTEMBER 18

13. Davuluri, P. S. L. (2023). AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems. AI-Augmented Sanctions Screening: Enhancing Accuracy and Latency in Real Time Compliance Systems (December 15, 2023).
14. Gu, F., Chung, M.-H., Chignell, M., Valaee, S., & others. (2021). A survey on deep learning for human activity recognition. ACM Computing Surveys, 54(8), Article 177.
15. Gu, M., Chen, K., & Chen, Z. (2024). RFDANet: An FMCW and TOF radar fusion approach for driver activity recognition using multi-level attention based CNN and LSTM network. Complex & Intelligent Systems, 10, 1517–1530.
16. Kolla, S., Meda, R., Balleda, L., & Thimmapuram, C. R. (2024). The utility value of ROX index and modified ROX index in determining the efficiency of HFNC in children admitted with respiratory distress. International Journal of Contemporary Pediatrics, 11(6), 775.
17. Ha, M.-H., & Chen, O. T.-C. (2021). Deep neural networks using capsule networks and skeleton-based attentions for action recognition. IEEE Access, 9, 135615–135628.
18. Sriram, H. K., Challa, S. R., Challa, K., & ADUSUPALLI, B. (2024). Strategic Financial Growth: Strengthening Investment Management. Secure Transactions, and Risk Protection in the Digital Era. Secure Transactions, and Risk Protection in the Digital Era (November 10, 2024).
19. Han, S., & Choi, J.-I. (2021). Study on driver's distraction research trend and deep learning based behavior recognition model. Journal of the Korea Society of Computer and Information, 26(11), 173–182.
20. Singireddy, S. (2024). The Integration of AI and Machine Learning in Transforming Underwriting and Risk Assessment Across Personal and Commercial Insurance Lines. Journal of Computational Analy-sis and Applications (JoCAAA), 33(08), 3966-3991.
21. Jayasundara, V., Roy, D., & Fernando, B. (2021). FlowCaps: Optical flow estimation with capsule networks for action recognition. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 3409–3418.
22. Jiao, S.-J., Liu, L.-Y., & Liu, Q. (2021). A hybrid deep learning model for recognizing actions of distracted drivers. Sensors, 21(21), 7424.
23. Kadar, J. A., Dewi, M. A. K., Suryawati, E., Heryana, A., Zilfan, V., Kusumo, B. S., Yuwana, R. S., Supianto, A. A., Pratiwi, H., & Pardede, H. F. (2023). Distracted driver behavior recognition using modified capsule networks. Journal of Mechatronics, Electrical Power, and Vehicular Technology, 14(2), 177–185.
24. Ikudabo, A. O., & Kumar, P. (2024). AI-driven risk assessment and management in banking: balancing innovation and security. International Journal of Research Publication and Reviews, 5(10), 3573-88.
25. Kaseris, M., Kostavelis, I., & Malassiotis, S. (2024). A comprehensive survey on deep learning methods in human activity recognition. Machine Learning and Knowledge Extraction, 6(2), 842–876.
26. Le, V.-T., Tran-Trung, K., & Hoang, V. T. (2022). A comprehensive review of recent deep learning techniques for human activity recognition. Journal of Computer Networks and Communications, 2022, 8323962.
27. Singireddy, J. (2024). AI-driven payroll systems: Ensuring compliance and reducing human error. American Data Science Journal for Advanced Computations (ADSJAC) ISSN, 3067-4166.
28. Lin, Y., Cao, D., Fu, Z., Huang, Y., & Song, Y. (2022). A lightweight attention-based network towards distracted driving behavior recognition. Applied Sciences, 12(9), 4191.
29. Liu, D., Yamasaki, T., Wang, Y., Mase, K., & others. (2021). TML: A triple-wise multi-task learning framework for distracted driver recognition. IEEE Access, 9, 130216–130229.
30. Yadav, V. (2024). Predictive Analytics for Preventive Medicine: Analyzing how Predictive Analytics is Utilized for Forecasting Patient Health Trends and Preventive Disease. Progress in Medical Sciences. PMS-1126. Prog Med Sci, 8(4).
31. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer:

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 03

RECEIVED: JULY 02 REVISED: AUGUST 19 ACCEPTED: AUGUST 27 PUBLISHED: SEPTEMBER 18

Hierarchical vision Transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 10012–10022.

32. Liu, Z., Ning, J., Cao, Y., Wei, Y., Zhang, Z., Lin, S., & Hu, H. (2022). Video Swin Transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3202–3211.
33. Kummari, D. N. (2022). IoT-enabled additive manufacturing: Improving prototyping speed and customization in the automotive sector. Migration Letters, 19(S8), 2084-2104.
34. Ma, Y., Yuan, L., Abdelraouf, A., Han, K., Gupta, R., Li, Z., & Wang, Z. (2023). M2DAR: Multi-view multi-scale driver action recognition with vision Transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.
35. Ma, Y., Du, R., Abdelraouf, A., Han, K., Gupta, R., & Wang, Z. (2024). Driver digital twin for online recognition of distracted driving behaviors. IEEE Transactions on Intelligent Vehicles.
36. Srinivas Kalisetty, D. A. S. (2024). Leveraging Artificial Intelligence and Machine Learning for Predictive Bid Analysis in Supply Chain Management: A Data-Driven Approach to Optimize Procurement Strategies.
37. Mou, L., Chang, J., Zhou, C., Zhao, Y., Ma, N., Yin, B., Jain, R., & Gao, W. (2023). Multimodal driver distraction detection using dual-channel network of CNN and Transformer. Expert Systems with Applications, 234, 121066.
38. Patrick, M. K., Adekoya, A. F., Mighty, A. A., & Edward, B. Y. (2022). Capsule networks—A survey. Journal of King Saud University—Computer and Information Sciences, 34(3), 1295–1310.
39. Pawan, S. J., & Rajan, J. (2022). Capsule networks for image classification: A review. Neurocomputing, 509, 102–120.
40. Koppolu, H. K. R. (2024). The impact of data engineering on service quality in 5G-enabled cable and media networks. European Advanced Journal for Science & Engineering (EAJSE), 1(1).
41. Ping, P., Huang, C., Ding, W., Liu, Y., Chiyomi, M., & Kazuya, T. (2023). Distracted driving detection based on the fusion of deep learning and causal reasoning. Information Fusion, 89, 121–142.
42. Qu, F., Dang, N., Furht, B., & Nojoumian, M. (2024). Comprehensive study of driver behavior monitoring systems using computer vision and machine learning techniques. Journal of Big Data, 11, Article 32.
43. Sun, H., & Li, Z. (2024). CDCL-VRE: An ensemble deep learning-based model for distracted driver behavior detection. Journal of Intelligent & Fuzzy Systems, 46(1), 1–16.
44. Sheelam, G. K. (2024). AI-driven spectrum management: Using machine learning and agentic intelligence for dynamic wireless optimization. European Advanced Journal for Emerging Technologies (EAJET), 2(1), 3050-9742.
45. Tang, X., Chen, Y., Ma, Y., Yang, W., & others. (2024). A lightweight model combining convolutional neural network and Transformer for driver distraction recognition. Engineering Applications of Artificial Intelligence, 132, 107910.
46. Wang, X., Xu, R., Zhang, S., Zhuang, Y., & Wang, Y. (2022). Driver distraction detection based on vehicle dynamics using naturalistic driving data. Transportation Research Part C: Emerging Technologies, 136, 103561.
47. Xiao, W., Liu, H., Ma, Z., Chen, W., & others. (2022). Attention-based deep neural network for driver behavior recognition. Future Generation Computer Systems, 132, 152–161.
48. Xiao, W., Xie, G., Liu, H., Chen, W., & Li, R. (2024). FDAN: Fuzzy deep attention networks for driver behavior recognition. Journal of Systems Architecture, 147, 103063.
49. Yang, W., Tan, C., Chen, Y., Xia, H., Tang, X., Cao, Y., Zhou, W., Lin, L., & Dai, G. (2023). BiRSwinT: Bilinear full-scale residual Swin-Transformer for fine-grained driver behavior recognition. Journal of the Franklin Institute, 360(2), 1166–1183.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 03

RECEIVED: JULY 02 REVISED: AUGUST 19 ACCEPTED: AUGUST 27 PUBLISHED: SEPTEMBER 18

50. Yang, X., Qiao, Y., Han, S., Feng, Z., & Chen, Y. (2024). Appearance-posture fusion network for distracted driving behavior recognition. Expert Systems with Applications, 257, 124883.
51. Zhao, L., Zhao, X., Wu, F., Chen, D., & Wang, C. (2024). A lightweight and efficient distracted driver detection model fusing convolutional neural network and vision Transformer. IEEE Transactions on Intelligent Transportation Systems, 25(12), 19962–19978.
52. Ramanakar Reddy Danda, Z. Y., Mandala, G., & Maguluri, K. K. (2024). Smart Medicine: The Role of Artificial Intelligence and Machine Learning in Next-Generation Healthcare Innovation.
53. Zhao, X., Li, Z., Zhao, C., Fu, R., & Wang, C. (2024). Distraction-level recognition based on stacking ensemble learning for IVIS secondary tasks. Expert Systems with Applications, 244, 122849.
54. Zhu, S., Chen, W., Liu, F., Zhang, X., Han, X., & others. (2023). Human activity recognition based on a modified capsule network. Mobile Information Systems, 2023, 8273546.
55. Jegham, I., Alouani, I., Ben Khalifa, A., & Mahjoub, M. A. (2023). Deep learning-based hard spatial attention for driver in-vehicle action monitoring. Expert Systems with Applications, 225, 119629.
56. Abdulmolla Abdulghafor, R. A., Hanafi, F. M., Nasir, M. F., Wani, S., Gulzar, Y., & Hamid, Y. (2021). A real time deep learning based driver monitoring system. International Journal on Perceptive and Cognitive Computing, 7(1), 79–84.
57. Liu, Y., Zhang, Y., Zhuang, Y., Wang, Y., & others. (2021). Identification of driver distraction based on SHRP2 naturalistic driving study. Mathematical Problems in Engineering, 2021, 6699327.
58. Fitri Mohd Hanafi, M. F., Faiz Md. Nasir, M. S., Wani, S., Abdulmolla Abdulghafor, R. A., Gulzar, Y., & Hamid, Y. (2021). A real time deep learning based driver monitoring system. International Journal on Perceptive and Cognitive Computing, 7(1), 79–84.
59. Bragança, H., Colonna, J. G., Oliveira, H. A. B. F., & Souto, E. (2022). How validation methodology influences human activity recognition mobile systems. Sensors, 22(6), 2360.
60. Liu, Z., Ning, J., Cao, Y., Wei, Y., Zhang, Z., Lin, S., & Hu, H. (2022). Video Swin Transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3202–3211.