mathematics

MDPI

Article

Driving Style Recognition Method Based on Risk Field and Masked Learning Techniques

Shengye Jin, Zhengyu Zhu, Junli Liu and Shouqi Cao *

College of Engineering Science and Technology, Shanghai Ocean University, Shanghai 201306, China; m210801333@st.shou.edu.cn (S.J.); m220851459@st.shou.edu.cn (Z.Z.); m210811370@st.shou.edu.cn (J.L.)
* Correspondence: sqcao@shou.edu.cn

Abstract: With the increasing demand for road traffic safety assessment, global concerns about road safety have been rising. This is particularly evident with the widespread adoption of V2X (Vehicle-to-Everything) technology, where people are more intensively focused on how to leverage advanced technological means to effectively address challenges in traffic safety. Through the research of driving style recognition technology, accurate assessment of driving behavior and the provision of personalized safety prompts and warnings have become crucial for preventing traffic accidents. This paper proposes a risk field construction technique based on environmental data collected by in-vehicle sensors. This paper introduces a driving style recognition algorithm utilizing risk field visualization and mask learning technologies. The research results indicate that, compared to traditional classical models, the improved algorithm performs excellently in terms of accuracy, stability, and robustness, enhancing the accuracy of driving style recognition and enabling a more effective evaluation of road safety.

Keywords: driving style recognition; driving risk field; mask learning; environmental data; safety tips and warnings; vehicle-to-everything

MSC: 68T07

Citation: Jin, S.; Zhu, Z.; Liu, J.; Cao, S. Driving Style Recognition Method Based on Risk Field and Masked Learning Techniques. *Mathematics* **2024**, *12*, 1363. https://doi.org/10.3390/math12091363

Academic Editor: Marjan Mernik

Received: 3 April 2024
Revised: 26 April 2024
Accepted: 29 April 2024
Published: 30 April 2024

1. Introduction

With the continuous growth of urban traffic congestion and the increasing number of vehicles, traffic accidents have become a serious societal issue. In this context, V2X (Vehicle-to-Everything) technology, as a key component of intelligent transportation systems, offers new solutions to enhance traffic safety, playing a crucial role in improving road safety [1]. V2X technology enables vehicles to communicate wirelessly, sharing information such as location, speed, and direction, fostering real-time connectivity among vehicles, and facilitating information exchange with infrastructure, pedestrians, and other traffic participants. Despite the potential of V2X technology to enhance communication and coordination among vehicles, preventing traffic accidents still poses challenges. One reason is the behavioral differences among drivers, and existing technologies have not fully leveraged V2X data to address this issue effectively. Research on driving style recognition plays a vital role in improving road safety assessment and preventing traffic accidents. Through the study of driving style recognition, a more accurate assessment of a driver's driving habits and style can be achieved, leading to personalized safety prompts and warnings and ultimately reducing the occurrence of traffic accidents.

Driver style recognition methods are primarily categorized into unsupervised learning, semi-supervised learning, and supervised learning. Unsupervised learning [2–6] and semi-supervised learning [7–9] methods require a smaller amount of data but face challenges in obtaining reliable sample features within limited data. In situations where data are sufficiently abundant, researchers opt for supervised learning for driver style

recognition [10–16]. This approach achieves high accuracy but demands high requirements for both the quantity and quality of training data.

Researchers have seldom considered the variations in environmental data in the algorithms for discriminating driving styles. However, it is evident that driving styles that involve the same operations differ across different environments. With advancements in sensor technology and the gradual proliferation of V2X (Vehicle-to-Everything) technology, smart vehicles now acquire a more diverse and extensive set of driving data. Consequently, there is a growing body of research related to assessing environmental conditions. The concept of a driving risk field serves as a model for evaluating driving risks on roads. By modeling a driving risk field, one can assess environmental risks and gather relevant variable information about the current environmental conditions. Through real-time monitoring and analysis of driving risks, this model allows for the assessment of real-time risks [17–25] in driving environments and the planning of feasible paths [21]. It also facilitates the prediction of potential risks under different road and traffic conditions.

In comparison to traditional machine learning methods, contrastive learning, as a form of self-supervised learning, has gained widespread research and application in fields such as computer vision and natural language processing. It is characterized by high data efficiency, strong generalization capabilities, and robust resistance to interference, achieving results that approach or even surpass the performance of supervised learning [26–28]. Mask learning, as a branch of contrastive learning, can handle more complex data features and exhibits superior performance compared to traditional contrastive learning methods. Currently, mask learning has demonstrated outstanding performance in the fields of image recognition and video recognition [29,30].

This study aims to explore a new approach for comprehensively assessing driving styles through changes in driving risk. We believe that the evaluation of driving styles should not be solely based on characteristics observed at a single moment but should instead delve into the trends of driving risk variations over a period of time. For instance, a driver who transitions suddenly from a prolonged period of low-risk driving to a high-risk state may indicate a temporary lapse of attention, reflecting a more aggressive driving tendency. Similarly, a consistent high-risk driving state may reveal a lower sensitivity to risk perception, manifesting as an impulsive driving pattern.

Furthermore, we recognize the pivotal role of environmental factors in driving risk variations and have thus introduced the concept of driving risk fields to address the oversights in earlier research. By further refining the driving risk field model, we strive to comprehensively consider various factors that influence driving risk, including vehicle dynamics, road environmental factors, and individual driver braking behavior characteristics. This enhancement not only enriches the dataset for prediction models but also significantly enhances the dimensionality and precision of the data.

To address the challenges of processing high-dimensional data, we have adopted a driving style recognition model similar to the MAE architecture, which efficiently extracts features from high-dimensional data, demonstrating significant advantages in handling such data. Compared to traditional methods, this model exhibits more stable performance when dealing with complex data, effectively overcoming the limitations of traditional methods in handling high-dimensional data with decreasing performance.

In summary, this study aims to achieve accurate identification of driving styles by improving the driving risk field model and combining advanced feature extraction and recognition techniques, thereby providing stronger technical support for road traffic safety.

2. Method

2.1. Design of Driving Risk Field Model

The driving risk field is divided into the "Vehicle Driving Risk Field" and the "Road Boundary Risk Field Model". The former assesses the risks generated during the vehicle's travel, while the latter evaluates the risks associated with road boundaries (including solid and broken lines).

In the past, research on driving risk fields has exhibited several notable shortcomings. Firstly, these studies have failed to adequately consider the braking reaction characteristics of drivers when operating a vehicle, which is a crucial factor in real-world driving scenarios. Secondly, some risk value functions exhibit excessively large differences over similar distances, potentially leading to inaccurate assessments of driving risks. Additionally, when road participants change, the computational burden of existing models often becomes significant, compromising their efficiency in practical applications.

To address these issues, we propose utilizing the sigmoid function to optimize the model. By incorporating the sigmoid function, we aim to more accurately capture the braking reaction characteristics of drivers, thereby enhancing the model's precision. Based on this, we introduce a novel vehicle driving risk field model that incorporates the braking reaction time. The braking reaction time, defined as the duration from when a driver detects the need to brake to the point when their foot reaches the brake pedal, serves as a critical metric for assessing driving safety. By comprehensively considering the braking reaction time and other relevant factors, our model offers a more comprehensive assessment of driving risks, providing effective support for road traffic safety.

Some studies have suggested [20] that from the perspective of physical fields, vehicles driving on roads are subject to a virtual "force" due to the presence of driving risks. Under the influence of this force, vehicles adjust their motion state to ensure driving safety, which is very similar to the phenomenon of particles being affected by forces in physical fields. Following this consideration, Tian et al. [20], drawing inspiration from the Yukawa potential [31] for field construction, formulated the driving risk field model in the form of an exponential function, integrating both physical attributes and kinematic states. When the vehicle heading angle is 0°, the driving risk assessment formula constructed by Tian et al. is shown in Formula (1).

$$E_i = \lambda_1 M_{\mathrm{eq}-i} \mathrm{e}^{\frac{\lambda_2 m_j \Delta v}{|k|}} \cdot \mathrm{e}^{-\beta a \cos\theta} \cdot \frac{k}{|k|} \tag{1}$$

In this context, E_i represents the driving risk generated by object i towards its surroundings, v_i denotes the speed magnitude of object i, λ_1, λ_2, and β are coefficients to be determined, θ is the angle between the distance vector from the vehicle's centroid to a certain point around the vehicle and the positive direction of the x-axis, Δv represents the speed difference, a is the acceleration of the vehicle, m_{eq-i} is the equivalent mass of object i, m_j is the mass of object j, k is the distance vector from the vehicle's centroid to a point around the vehicle, and $|k|$ is the scalar distance from the vehicle's centroid to that specific point.

The definition of equivalent mass is crucial for assessing the potential hazards encountered during vehicle operation. It takes into account both the mass and speed of a vehicle to quantify the risk it poses to other road users. Put simply, the greater the mass and speed of a vehicle, the larger its equivalent mass becomes, thus increasing the potential driving risk. This viewpoint is strongly supported by multiple studies, including the findings of the World Bank and the World Health Organization's report on "Road Safety Countermeasures in Developing Countries" published in 2004. The report points out that in developing countries, there is a significant correlation between the number of traffic accidents, the number of injured individuals, and the number of fatalities, and the average road speed, exhibiting a relationship of the second, third, and fourth power, respectively.

Building on this theoretical foundation, Wu and their colleagues [17,32] conducted further research on the impact of speed on driving risk, utilizing a polynomial incorporating speed power function terms to represent this effect. By leveraging accident data from highways, they fitted relevant parameters and derived an empirical formula (Formula (2)) for vehicle equivalent mass. By substituting this empirical formula into the calculation formula for vehicle equivalent mass, the original formula (Formula (1)) was transformed into a new formula (Formula (3)). This transformation not only enhances the accuracy of risk assessments but also provides a more scientific and effective tool for our subsequent research.

$$M_{\mathrm{eq}-i} = 1.566 m_i v_i^{6.687} \times 10^{-14} + 0.3345 \tag{2}$$

$$E_i = \left(1.566 m_i v_i^{6.687} \times 10^{-14} + 0.3345\right) \lambda_1 \mathrm{e}^{\frac{\lambda_2 m_j \Delta v}{|k|}} \cdot \mathrm{e}^{-\beta a \cos\theta} \cdot \frac{k}{|k|} \tag{3}$$

Let $A = e^{\frac{\lambda_2 m_j \Delta v}{|k|}}$ and $B = e^{-\beta a \cos\theta}$. The formula is then transformed into Formula (4).

$$E_i = \lambda_1 m_{\mathrm{eq}-i} A \cdot \mathrm{B} \cdot \frac{k}{|k|} \tag{4}$$

The function A serves to assess the level of risk based on the Time-to-Collision (TTC) model [20], while the function B is designed to evaluate the risk distribution under different angles. According to Formula (1), it is evident that function A is an exponential function, indicating that, when Δv remains constant, the rate of change is also an exponential function. Over the domain [0, +inf], the derivative of this function is consistently less than 0, but the absolute value of its rate of change gradually decreases.

However, scientific studies indicate that drivers have a reaction time when encountering danger. During this braking reaction time, drivers are in an unconscious state and are unable to actively perform braking, steering, or other operations. Therefore, during this period, the risk should be higher, and the rate of change in risk should be smaller. After exceeding the reaction time, it can be assumed that drivers have the ability and initiate a response, at which point the rate of change in risk should reach its maximum and gradually decrease. Therefore, to align with the operational characteristics of drivers, and based on relevant research and experiments, the sigmoid function is chosen as the core function for part A. The formula for function A is then modified to meet driver behavior, as shown in Formula (5).

$$A = c_1 \times \ sigmoid\left(-c_2\left(|k| - c_{reaction}\ v - \frac{v^2}{2a_{max}}\ \right)\right) \tag{5}$$

where c_1, c_2, and $c_{reaction}$ are constants. c_1 is a threshold constant, a positive constant such that the threshold of A is constrained within the range $(0, c_1)$; c_2 is another positive constant controlling the horizontal shape of the sigmoid function, with larger values of c_2 leading to a quicker saturation of A; $c_{reaction}$ is a positive constant representing the driver's braking reaction time. $|k|$ is the distance from the experimental point to the current vehicle's center of mass; v is the speed of the vehicle itself; and a_{max} is the maximum acceleration of the vehicle. Ultimately, the Vehicle Driving Risk Field Model is obtained, as shown in Formula (6).

$$E_i = \lambda_1 m_{eq-i} c_1 \times \ sigmoid\left(-c_2\left(|k| - c_{reaction}\ v - \frac{v^2}{2a_{max}}\ \right)\right) e^{\beta a \cos\theta} \cdot \frac{k}{|k|} \tag{6}$$

The Road Boundary Risk Field refers to a collection of potential hazards or adverse factors associated with the road edge. It describes various risks that may occur along the road boundary, such as traffic accidents, conflicts between pedestrians and vehicles, visibility issues, and more. It assists decision makers in understanding potential risk factors, optimizing road design and traffic planning, and implementing preventive measures to reduce the occurrence of traffic accidents.

Based on research on the Road Boundary Risk Field [17,20], the formula for the road risk field used is depicted in Formula (7).

$$E = c_3 e^{\frac{-s^2}{2\gamma^2}} \tag{7}$$

where c_3 is a positive constant used to control the maximum field strength; s is the perpendicular distance from a point to the lane line; and γ is a positive constant used to control

the decay rate of the field strength from the boundary line to the center, with a larger γ resulting in slower decay.

2.2. Driving Style Recognition Based on Mask Learning Technology

The masked autoencoder (MAE) [29] is a self-supervised learning method used in computer vision. It is based on the Vision Transformer (ViT) architecture, known for its strong scalability and simplicity. The MAE method trains by randomly masking parts of input images and reconstructing the missing pixels. Today, MAE is also applied to process temporal images, with VideoMAE [30] being an extension that views images as individuals within temporal data, making it suitable for handling video data. Similarly, each frame of a risk field can be considered as an image, and the temporal data of the risk field can be likened to video data, enabling relevant feature extraction.

In this research, we have developed a driving style recognition process grounded in the MAE philosophy with the goal of effectively classifying driving risk styles. The classification aims to categorize driving styles into "Aggressive", "Moderate", and "Conservative". This procedure is principally divided into two phases: the pre-training phase and the downstream training phase. Figure 1 illustrates the training process and network architecture.

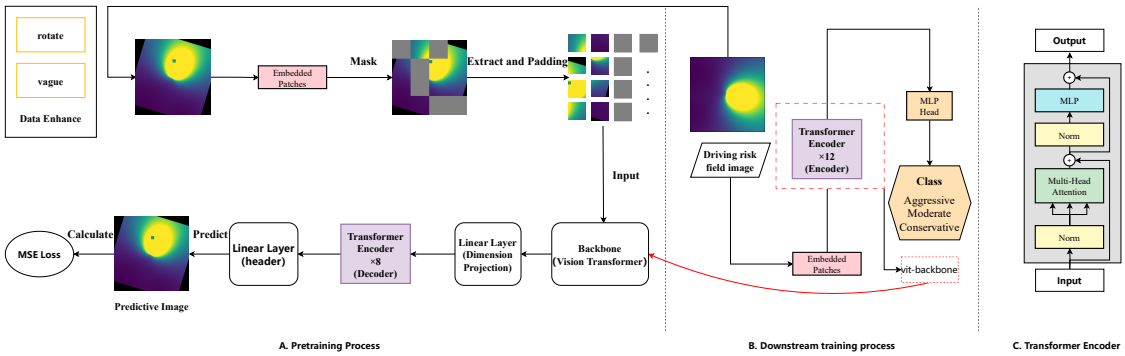


Figure 1. The overall training process and structure of masked autoencoder.

As indicated on the left side of Figure 1, in the pre-training phase, we initially apply data augmentation techniques such as rotation and blurring to the input images to enhance the model's generalization abilities. Once augmented, we mask certain regions of the image to replicate the visual system's handling of incomplete information. Then, segments from the unmasked image are extracted and processed through the backbone network. We chose the Vit-Base as our backbone network, which employs the self-attention mechanism from Transformers, particularly suited for image-centric tasks. The processed image segments are then passed on to the Header network, a densely connected neural network consisting of multiple dense layers. Each dense layer features neurons tightly interconnected, with each neuron connecting to all neurons from the preceding layer. The Header network is tasked with predicting the content of the masked portions of the image. During this process, the mean square error (MSE) loss between the prediction and the actual image is computed to guide the model's self-optimization in future iterations. Additionally, in our research on driving style recognition, we recognize that the data augmentation approach differs from that of traditional image recognition. Traditional image recognition mainly focuses on detecting the presence of relevant semantics within the image and emphasizes the extraction of image contour features; on the other hand, driving style recognition is less sensitive to the extraction of image contour features but more attentive to the variation differences between adjacent risk field intensity images and certain statistical parameters. Therefore, data augmentation strategies from traditional image recognition cannot be entirely applicable, and we opt for only those data augmentation strategies that marginally impact semantics. The final determined data augmentation strategy is the application of random rotation and blurring to temporal images. Rotation itself does not affect semantics,

so each data augmentation instance includes a random rotation, and the range of random rotation is between −10 degrees and 10 degrees; blurring is randomly applied with a minimal probability and range during the data augmentation process.

As shown on the right side of Figure 1, we have selected the ViT-Base architecture as our backbone network due to its exemplary performance in processing visual information. The "Embedded Patch" module divides the input image into 16x16 patches and adds positional encoding to these feature vectors. Afterward, these patches are transformed into a series of lower-dimensional vectors through a linear layer to be fed into the Transformer model. Subsequently, a series of self-attention layers process each vector within the sequence, with each layer capable of establishing intricate dependencies between different patches. The entire network also employs residual connections and layer normalization to refine the process, which aids in preventing gradient vanishing issues in deep networks while also expediting the training procedure. This entire process, through the cohesion of pre-training and downstream training, fortifies the model's capability in identifying driving risks. This comprehensive training methodology not only enhances the model's efficiency in recognizing complex driving risk images but also, through an in-depth combination of self-supervised and supervised learning, is set to improve driving style recognition performance.

3. Experiment

3.1. Experimental Design

The experiment utilized a dataset constructed by X. Liu et al. [13]. This dataset was acquired through Liu's self-built vehicle data collection platform, capturing natural driving trajectories such as straight driving and lane changes on roads. The data were collected on highways and urban roads in Shanghai, covering various regions of the city and amassing over 1000 km of travel data. The dataset includes diverse road conditions and spans different time periods throughout the day.

For the experiment, representative left lane change and right lane change data were selected from the dataset. Figure 2 illustrates the trajectory routes of lane change data. In these sets, each line of a different color represents a different lane change trajectory. It is notable that the starting point of each lane change trajectory is located in the very center of the lane, while the end point is the last trajectory point where the vehicle stops after completing the lane change. In establishing the coordinate system, we used the starting point of each lane change trajectory as the origin of coordinates. The forward direction of the lane is defined as the positive direction of the x-axis, and rotating the positive direction of the x-axis counterclockwise by 90 degrees yields the positive direction of the y-axis. Such a definition constructs a global Cartesian coordinate system, enabling us to describe more accurately the position and movement of the vehicles on the road. Additionally, in the data collection experiment, the width of each lane was 4.4 m.

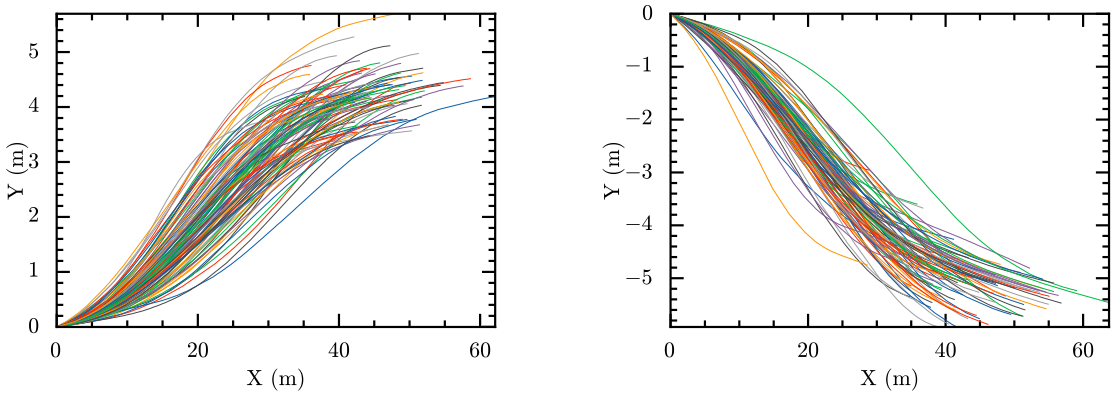


Figure 2. Visualization of lane change data.

Table 1 illustrates the basic attributes of the dataset. In this context, the *lon_speed* field represents longitudinal speed, and the *lat_speed* field represents lateral speed, both measured in meters per second. The *lon_acc* field corresponds to longitudinal acceleration, the *lat_acc* field corresponds to lateral acceleration, and both are measured in meters per second squared. The *Angleheadingrate* field represents the yaw rate.

Table 1. Basic properties of the dataset.

	lon_speed	*lat_speed*	*lon_acc*	*lat_acc*	*Angleheadingrate*
mean	9.610189	1.153941	−0.189604	−0.001717	0.050463
std	1.233309	0.436482	0.535301	0.588758	3.525656
min	6.495732	−0.099031	−4.649437	−2.236996	−9.725
25%	8.669245	0.786684	−0.531937	−0.498131	−3.004
50%	9.476567	1.168756	−0.222171	−0.048021	−0.018
75%	10.347681	1.496506	0.132453	0.543476	3.317
max	13.376584	2.453871	3.579057	1.606629	9.001

The experiment will utilize the aforementioned dataset and, after further preprocessing, generate a new dataset for driver style recognition. We will compare the use of different data extraction modules and prediction networks, examining the differences across various models.

3.1.1. Data Preprocessing

The original dataset has a sampling frequency of 100 Hz, recording data every 0.01 s. Following previous research [13], this experiment uses data collected every 0.3 s as input to discern driving styles. For every 30 raw data entries, 30 images of driving risk fields are generated. The data preprocessing workflow is as shown in Figure 3. Subsequently, the images of driving risk fields are fed into the model for driving style discrimination. Figure 4 illustrates the variation pattern of field strength when both the road boundary risk field and vehicle risk field coexist under different parameters. The field strength generated by both types of fields is vector-based. Directly adding these vectors does not assess the overall level of risk but only the combined field strength of the experimental object in multiple fields. Therefore, a vector modulus superposition method is employed to assess the risk level at a specific point.

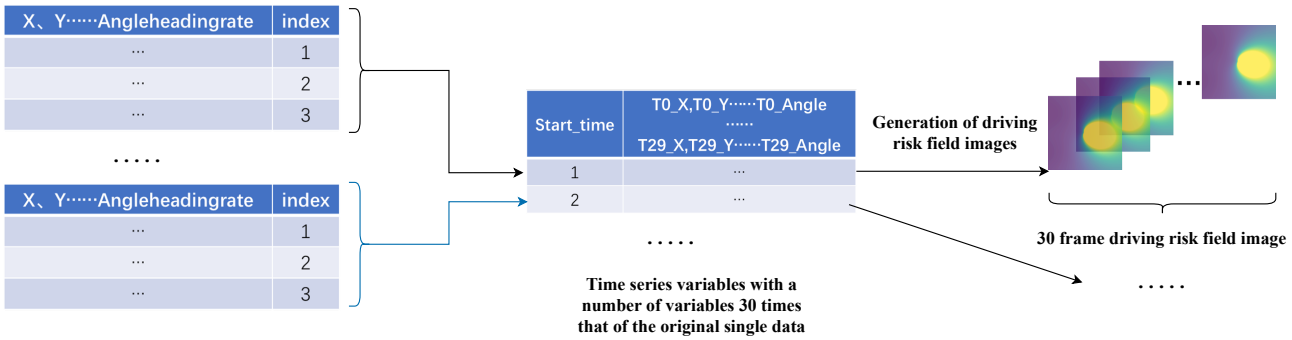
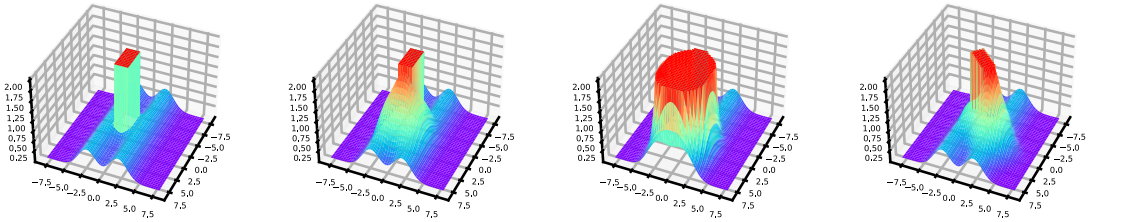


Figure 3. Data preprocessing process.



The speed is 1m/s, the speed direction is the positive x-axis direction, and the acceleration is 0m/s^2

The speed is 5m/s, the speed direction is the positive x-axis direction, and the acceleration is 0m/s^2

The speed is 5m/s, the speed direction is the positive x-axis direction, the acceleration is 3m/s^2, and the direction is the positive x-axis direction

The speed is 5m/s, the speed direction is in the direction of the bisector of the x-axis positive direction and the y-axis positive direction angle, and the acceleration is 0m/s^2

Figure 4. Field strength images of road boundary risk field and vehicle risk field coexisting under different parameters.

After obtaining the driving risk images, the dataset is labeled for driving style. Referring to the labeling method proposed by X. Liu et al. [13], relevant statistical properties are computed. A Gaussian Mixture Model is used for pre-labeling, with a cluster count of 3. The three driving styles are categorized as aggressive, moderate, and conservative. We posit that data points with a confidence level greater than 90% have a distinct driving style classification. Conversely, other data points exhibit some classification ambiguity, yet the labels themselves are accurate. Therefore, data points with confidence levels below 90% can be utilized to assess robustness. The experiment selects data points with confidence levels exceeding 90% for each driving style as their labels. For these data points, 20% are used as the training set, 80% as the validation set, and the remaining points with confidence levels below 90% are designated as the test set. The experiment will conduct relevant tests on this dataset.

3.1.2. Comparative Experiment

We will prove the effectiveness of the models under the MAE framework through comparative experiments, and the model comparison framework is illustrated in Figure 5.

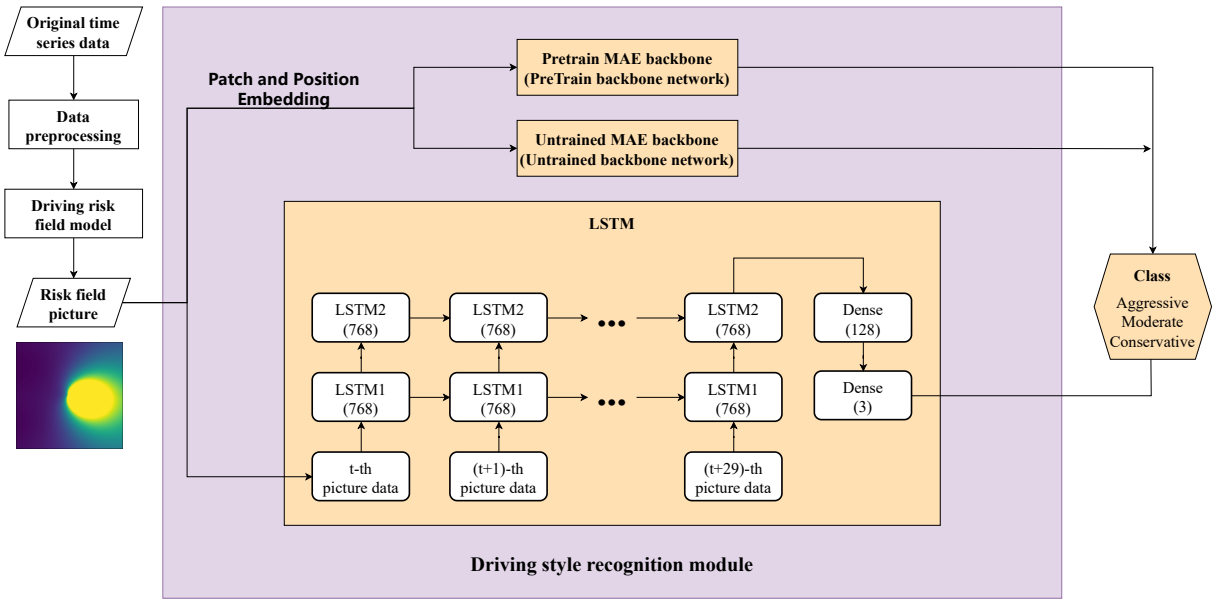


Figure 5. Model comparison framework.

Three models are designed for comparison: MAE-Pretrain, MAE-Untrained (ViT), and LSTM. The ViT model and LSTM model each have advantages in image feature extraction

and temporal feature extraction, making them classic benchmark models. Additionally, since MAE, as a training framework, shares the same backbone network and ViT-base network structure, the MAE-Untrained model is equivalent to the ViT-base model. Comparing MAE-Pretrain and MAE-Untrained is essentially comparing the effectiveness of the pre-trained structure of MAE. Long Short-Term Memory (LSTM) [33] is a variant of recurrent neural networks (RNNs) and is a classic model in extracting temporal data features. Moreover, X. Liu et al. proposed using LSTM for driver style recognition [13], demonstrating the effectiveness of LSTM in extracting features from temporal variables.

3.2. Experimental Parameters

As part of the upstream task training, the MAE-Pretrain model employs a masking rate of 70% to enhance the model's generalization capabilities. The Decoder network architecture utilizes the TransformerEncoder structure, repeated eight times, with an embed_dim of 512. Correspondingly, the Encoder network architecture is also based on the TransformerEncoder, repeated twelve times, and has an embed_dim set to 768. In the upstream task, the header section comprises a single linear layer with 768 neurons, whose primary function is to transform the output of the Decoder network into an image, fulfilling the output requirements of the target task.

When transitioning to the downstream task, the backbone network directly adopts the model obtained from the upstream task training to ensure the effectiveness of knowledge transfer. The head network consists of two LinearLayers, with the number of neurons in the first and second layers being 768 and 3, respectively. This design facilitates the extraction and refinement of feature information. The loss function chosen is the cross-entropy function, suitable for classification tasks.

The MAE-Untrained model, on the other hand, has not undergone the pre-training phase. Therefore, its network structure and loss function remain consistent with the MAE-Pretrain model in downstream tasks, maintaining fairness in comparisons. Additionally, in the context of LSTM models, a two-layer stacked LSTM model with a hidden layer feature count of 768 is employed, and the cross-entropy function is selected as the loss function.

The experiment was conducted using the Ubuntu 20.04 system and the NVIDIA A4000 graphics card for training. During the pre-training phase of MAE-Pretrain, the AdamW optimizer was used, employing a cosine annealing schedule for learning rate adjustment, and a total of 200 epochs were trained. For both the MAE-Pretrain and MAE-Untrained downstream tasks, the AdamW optimizer was utilized, employing a cosine annealing with restarts scheduler for learning rate adjustment. Full fine-tuning was applied as the training method for these downstream tasks, with each task trained for a total of 200 epochs. Additionally, an LSTM model was trained separately using the SGD optimizer, also adopting a cosine annealing with restarts scheduler for learning rate adjustment, and underwent 200 epochs of training.

3.3. Result Analysis

The processing and analysis of experimental data demonstrate the accuracy, stability, and robustness of the MAE-based model. In the experiments assessing generalization under different conditions, both models exhibit excellent performance, accurately discerning driving styles.

Figure 6 illustrates the accuracy and loss variations during the training process of different models under two operating conditions. Figure 6a,c depict the data changes during left lane change conditions, while Figure 6b,d show the data changes during right lane change conditions. Whether in left lane change or right lane change conditions, the accuracy of the MAE-Pretrain model is consistently higher than that of the MAE-Untrained and LSTM models at any given moment. Based on the statistical data presented in Table 2 and an integrated analysis of the statistical characteristics of left-turn and right-turn conditions, it can be observed that the accuracy characteristics are noteworthy. The MAE-Pretrain model has an average accuracy of around 97%, with peak accuracy consistently exceeding

98%. In comparison to the other two classical models, its accuracy is significantly improved. Moreover, the accuracy curves of the MAE-Pretrain model and the LSTM model fluctuate more smoothly in both conditions, with the MAE-Pretrain model exhibiting greater stability than the LSTM model. In contrast, the fluctuation of the MAE-Untrained model is pronounced. Combining the relevant statistical data from Tables 2 and 3, it is evident that the range and variance of the MAE-Pretrain model are significantly smaller than the other two models. Additionally, its average accuracy surpasses the other two models, indicating better training stability.

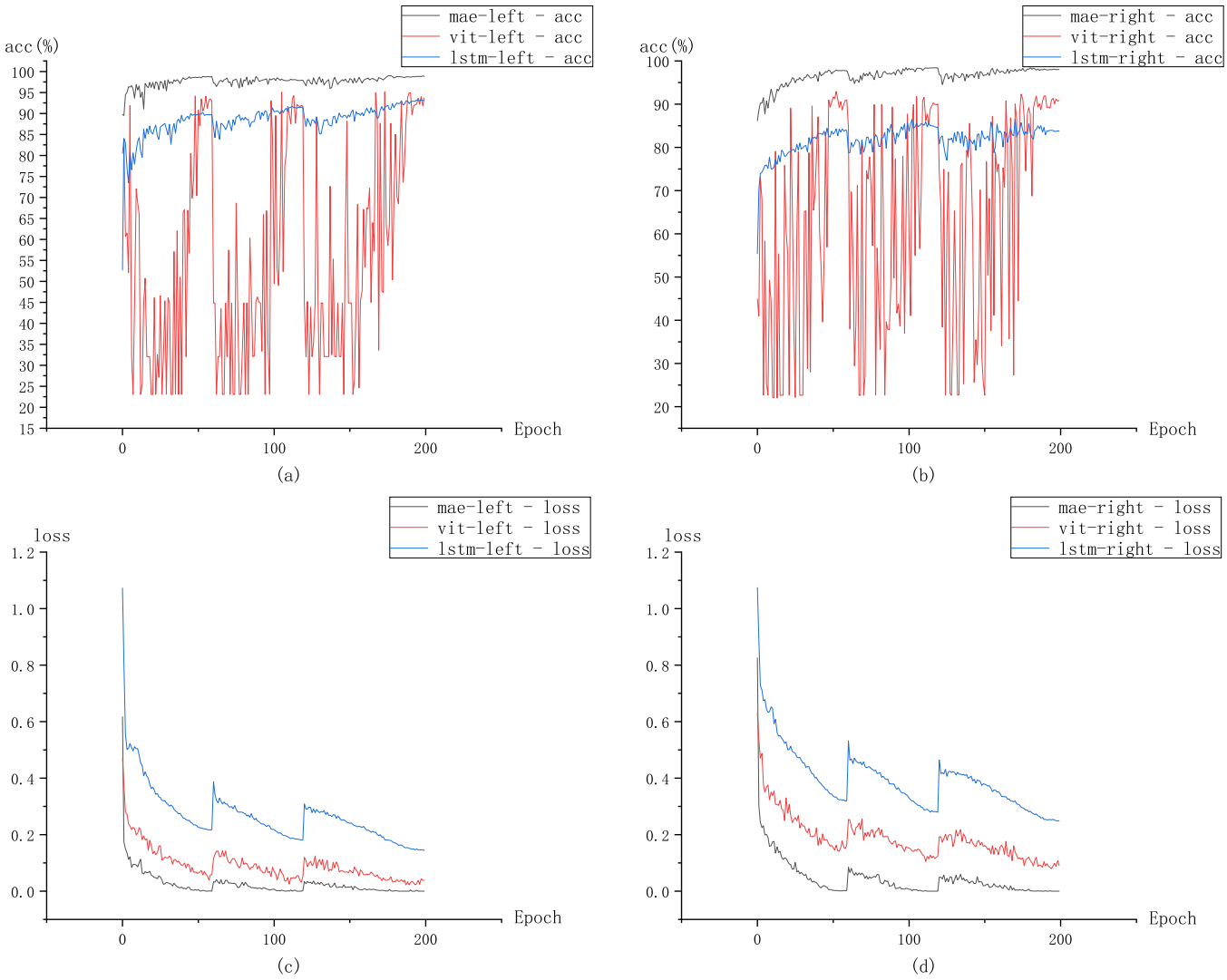


Figure 6. Accuracy and loss changes during downstream task training under left and right lane changing conditions. (**a**) Accuracy variation graphs of the three models under left lane change conditions, (**b**) Accuracy variation graphs of the three models under right lane change conditions, (**c**) Loss variation graphs of the three models under left lane change conditions, (**d**) Loss variation graphs of the three models under right lane change conditions.

Table 2. Statistical data on accuracy during downstream task training process.

	MAE	ViT	LSTM		MAE	ViT	LSTM
count	200	200	200	count	200	200	200
mean	97.64517	56.54148	88.47549	mean	96.58314	63.51776	81.86565
std	1.33474	25.01102	4.08483	std	2.04889	24.80087	3.34323
min	89.54546	23.08712	52.67578	min	86.17064	22.00397	55.38651
25%	97.36269	32.08334	87.33887	25%	96.14584	39.7123	80.29914
50%	97.95455	49.89584	89.24805	50%	97.10318	70.1885	82.56579
75%	98.4375	82.06439	90.60059	75%	97.89683	88.69544	83.8456
max	98.97727	95.17046	93.20313	max	98.47223	92.93651	86.51316
	Left				**Right**		

Table 3. Statistical data on losses during downstream task training.

	MAE	ViT	LSTM		MAE	ViT	LSTM
count	200	200	200	count	200	200	200
mean	0.02443	0.09204	0.26841	mean	0.04367	0.18668	0.39453
std	0.05108	0.05989	0.10822	std	0.07612	0.08009	0.11427
min	1.00E-06	0.0219	0.14414	min	0.00001	0.07958	0.24852
25%	0.00269	0.0219	0.20878	25%	0.00373	0.1404	0.31935
50%	0.01306	0.07677	0.25021	50%	0.02319	0.17134	0.37981
75%	0.02795	0.11113	0.29671	75%	0.0526	0.2116	0.437
max	0.61776	0.471	1.07257	max	0.82601	0.63153	1.07428
	Left				**Right**		

To further evaluate the effectiveness of the experiments, testing was conducted on additional data with confidence levels below 90%, i.e., the trained models were applied to a new dataset for testing. Table 4 illustrates the difference in accuracy between the validation set and the test set under different methods.

Table 4. The difference in accuracy between the validation set and the test set under different methods.

	MAE	ViT	LSTM		MAE	ViT	LSTM
Val-acc	98.97727	95.17046	93.20313	Val-acc	98.47223	92.93651	86.51316
Test-acc	90.2789	84.9561	78.7145	Test-acc	88.9945	81.5197	80.27
	Left				**Right**		

The MAE model achieved the highest accuracy on all three sets, especially on the test set, which had a significantly larger data quantity than the training and validation sets. This implies that, in scenarios with a small amount of high-confidence data as input, the robustness of the MAE-based driving style recognition model is significantly stronger than that of other models.

In conclusion, by taking into account Figure 6 and Tables 2–4, we can deduce the advantages of the MAE model in comparison to other models, as outlined in Table 5.

Table 5. Performance improvement comparison of MAE model relative to ViT and LSTM models.

	ViT	**LSTM**
Accuracy (Training Set)	At least 3% improvement	At least 5% improvement
Stability	Significantly improved	Improved
Accuracy (Test Set)	At least 5% improvement	At least 8% improvement

Regarding the superior accuracy, stability, and robustness of the MAE-Pretrain model in comparison, we have the following hypotheses: The effectiveness of the MAE-Pretrain model in accuracy, stability, and robustness stems from its superior feature extraction performance in the upstream task. The reason for this superiority lies in the high similarity between the pre-training task under the mask mode and the downstream task in driving style recognition.

The MAE-Pretrain and MAE-Untrained models have identical network structures, with the only difference being their involvement in upstream task training, indicating the crucial role of the upstream task in improving accuracy. Generally, in the full fine-tune training mode, when the similarity between the upstream and downstream tasks is high, i.e., when the two tasks require similar features, the effect is better. Based on this experience, it can indirectly suggest that the reconstruction features extracted from the upstream tasks of the two methods in this paper, i.e., driving risk field visualization and MAE temporal image reconstruction, are closely related and effective for driving style discrimination.

From another perspective, both image-based models achieved higher maximum accuracy than the LSTM model, indicating that after visualizing the driving risk field, features can be better extracted through image-based methods.

As the features extracted from the upstream task have high similarity to the target features required by the downstream task, the MAE-Pretrain model experiences minor changes to the backbone network during downstream task training, leading to relatively small variations in loss and accuracy. In contrast, the MAE-Untrained model, without pre-training, needs continuous learning and adjustment to find the extremum of the loss function, resulting in larger fluctuations.

Furthermore, in discussing the architecture of this paper, Figure 1A illustrates the upstream task. Within this task, the Encoder module plays a crucial role in extracting the inherent features from the sequential image series. The quality of this feature extraction directly impacts the capabilities of the downstream task. On the other hand, the Decoder module utilizes a relatively simple feature reconstruction network to reconstruct image information. Figure 1B depicts the downstream task, which relies heavily on the Encoder module from the upstream task. If the Encoder module is able to extract features effectively, it will significantly enhance the training speed, accuracy, and overall performance of the downstream task.

4. Conclusions

In this research, we present a cutting-edge methodology for risk evaluation and driving style identification, designed to navigate the complexities of driving environments. Utilizing a method that constructs driving risk field images based on braking reaction times, and coupled with an autoencoder-based driving style recognition algorithm that leverages masked learning for data feature enhancement, our approach not only offers a fresh perspective in visualizing risk fields but also fully exploits the potential of masked learning for data augmentation, refining the risk discrimination process.

Regarding performance evaluation, the model demonstrates at least a 3% increase in accuracy over the ViT and at least a 5% increase over LSTM networks on the training dataset. In terms of stability, it shows marked improvements compared to both ViT and LSTM. Most notably, on the testing dataset, the model's accuracy outperforms ViT by at least 5% and LSTM by at least 8%.

In summary, the model proposed in this paper not only introduces innovation in the construction of driving risk field imagery but also exhibits significant advantages in data feature extraction. Empirical evidence confirms its superior performance in accuracy, stability, and robustness, promising to offer an effective technological solution for driving safety evaluation and personalized driving style recognition.

The main contributions of this paper are as follows:

1. An innovative method for constructing a driving risk field based on braking reaction time is proposed. This method breaks through the limitations of existing research by more comprehensively considering the actual reaction characteristics of drivers during the driving process, thus improving the accuracy and reliability of driving risk assessment.
2. The concept of converting the driving risk field into image representation is creatively proposed, and the idea of masked autoencoder is utilized for feature extraction. This innovation provides a more effective means of feature extraction for the pre-trainer, thereby contributing to the enhancement of subsequent driving style recognition performance.
3. A multi-stage training approach is adopted, which effectively reduces the influence of subjective factors on transfer tasks while addressing the common clustering bias issues in traditional unsupervised clustering algorithms. The application of this method improves accuracy and stability, providing more reliable technical support for the practical application of driving risk style recognition.
4. In response to the problem of decreasing effectiveness of traditional time-series algorithms in handling high-dimensional data, this paper innovatively adopts algorithms from the field of computer vision to address this issue, providing a new perspective for driving style recognition.

Further research:

We will enrich the data augmentation techniques used in the upstream task stage, considering factors such as the driver's position in the main driving seat within the vehicle. We will explore generating similar paths and evaluate driving style recognition by comparing the driving styles of generated paths with the original paths.

Author Contributions: Conceptualization, S.C.; methodology, S.J.; software, S.J. and J.L.; validation, Z.Z. and J.L.; investigation, Z.Z.; writing—original draft preparation, S.J.; writing—review and editing, S.J. and Z.Z.; visualization, Z.Z. and J.L.; supervision, S.C.; project administration, S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy and confidentiality agreements.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ghosal, A.; Conti, M. Security Issues and Challenges in V2X: A Survey. *Comput. Netw.* **2020**, *169*, 107093. [CrossRef]
2. Xing, Y.; Lv, C.; Wang, H.; Cao, D.; Velenis, E.; Wang, F.Y. Driver Activity Recognition for Intelligent Vehicles: A Deep Learning Approach. *IEEE Trans. Veh. Technol.* **2019**, *68*, 5379–5390. [CrossRef]
3. de Zepeda, M.; Meng, F.; Su, J.; Zeng, X.J.; Wang, Q. Dynamic Clustering Analysis for Driving Styles Identification. *Eng. Appl. Artif. Intell.* **2021**, *97*, 104096. [CrossRef]
4. Ma, Y.; Li, W.; Tang, K.; Zhang, Z.; Chen, S. Driving Style Recognition and Comparisons among Driving Tasks Based on Driver Behavior in the Online Car-Hailing Industry. *Accid. Anal. Prev.* **2021**, *154*, 106096. [CrossRef] [PubMed]

5. Li, X.S.; Cui, X.T.; Ren, Y.Y.; Zheng, X.L. Unsupervised Driving Style Analysis Based on Driving Maneuver Intensity. *IEEE Access* **2022**, *10*, 48160–48178. [CrossRef]
6. Li, Y.; Zhang, H.; Wang, Q.; Wang, Z.; Yao, X. Study on Driver Behavior Pattern in Merging Area under Naturalistic Driving Conditions. *J. Adv. Transp.* **2024**, *2024*, e7766164. [CrossRef]
7. Wang, W.; Xi, J.; Chong, A.; Li, L. Driving Style Classification Using a Semisupervised Support Vector Machine. *IEEE Trans. Hum. Mach. Syst.* **2017**, *47*, 650–660. [CrossRef]
8. Liu, W.; Deng, K.; Zhang, X.; Cheng, Y.; Zheng, Z.; Jiang, F.; Peng, J. A Semi-Supervised Tri-CatBoost Method for Driving Style Recognition. *Symmetry* **2020**, *12*, 336. [CrossRef]
9. Zhang, W. A Semi-Supervised Learning Method for Multi-Condition Driving Style Recognition. Master's Thesis, Jilin University, Changchun, China, 2022. [CrossRef]
10. Silva, I.; Eugenio Naranjo, J. A Systematic Methodology to Evaluate Prediction Models for Driving Style Classification. *Sensors* **2020**, *20*, 1692. [CrossRef]
11. Guo, Y.; Wang, X.; Huang, Y.; Xu, L. Collaborative Driving Style Classification Method Enabled by Majority Voting Ensemble Learning for Enhancing Classification Performance. *PLoS ONE* **2021**, *16*, e0254047. [CrossRef]
12. Kim, D.; Shon, H.; Kweon, N.; Choi, S.; Yang, C.; Huh, K. Driving Style-Based Conditional Variational Autoencoder for Prediction of Ego Vehicle Trajectory. *IEEE Access* **2021**, *9*, 169348–169356. [CrossRef]
13. Liu, X.; Wang, Y.; Zhou, Z.; Nam, K.; Wei, C.; Yin, C. Trajectory Prediction of Preceding Target Vehicles Based on Lane Crossing and Final Points Generation Model Considering Driving Styles. *IEEE Trans. Veh. Technol.* **2021**, *70*, 8720–8730. [CrossRef]
14. Jia, L.; Yang, D.; Ren, Y.; Qian, C.; Feng, Q.; Sun, B. A Dynamic Driving-Style Analysis Method Based on Drivers' Interaction with Surrounding Vehicles. *J. Transp. Saf. Secur.* **2024**, *0*, 1–24. [CrossRef]
15. Zhang, S.; Shao, X.; Wang, J. Research on Lane-Changing Decision Model with Driving Style Based on XGBoost. In Proceedings of the Third International Conference on Intelligent Traffic Systems and Smart City (ITSSC 2023), Xi'an, China, 10–12 November 2023; SPIE: Cergy-Pontoise, France, 2023; Volume 12989, pp. 95–100. [CrossRef]
16. Wang, K.; Qu, D.; Yang, Y.; Dai, S.; Wang, T. Risk-Quantification Method for Car-Following Behavior Considering Driving-Style Propensity. *Appl. Sci.* **2024**, *14*, 1746. [CrossRef]
17. Wang, J.; Wu, J.; Li, Y. Concept, Principles, and Modeling of Driving Risk Field Based on Human-Vehicle-Road Coordination. *China J. Highw. Transp.* **2016**, *29*, 105–114. [CrossRef]
18. Mullakkal-Babu, F.A.; Wang, M.; He, X. Probabilistic Field Approach for Motorway Driving Risk Assessment. *Transp. Res. Part C Emerg. Technol.* **2020,** *118*, 102716. [CrossRef]
19. Xiong, J.; Shi, J.; Wan, H. Construction of Integrated Human-Vehicle-Road Risk Field Model and Driving Style Evaluation. *J. Transp. Syst. Eng. Inf. Technol.* **2021**, *21*, 105–114. [CrossRef]
20. Tian, Y.; Pei, H.; Yan, S.; Zhang, Y. Extension and Application of Driving Risk Field Model under Vehicle-Road Coordination Environment. *J. Tsinghua Univ. (Sci. Technol.)* **2022**, *62*, 447–457. [CrossRef]
21. Luo, J.; Li, S.; Li, H.; Xia, F. Intelligent Network Vehicle Driving Risk Field Modeling and Path Planning for Autonomous Obstacle Avoidance. *J. Mech. Eng. Sci.* **2022**, *236*, 8621–8634. [CrossRef]
22. Tan, H.; Lu, G.; Liu, M. Risk Field Model of Driving and Its Application in Modeling Car-Following Behavior. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 11605–11620. [CrossRef]
23. Chen, C.; Lan, Z.; Zhan, G.; Lyu, Y.; Nie, B.; Li, S.E. Quantifying the Individual Differences of Drivers' Risk Perception via Potential Damage Risk Model. *IEEE Trans. Intell. Transp. Syst.* 2024, *Early Access*. [CrossRef]
24. Zhong, N.; Gupta, M.K.; Kochan, O.; Cheng, X. Evaluating the Efficacy of Real-Time Connected Vehicle Basic Safety Messages in Mitigating Aberrant Driving Behaviour and Risk of Vehicle Crashes: Preliminary Insights from Highway Scenarios. *Elektron. Elektrotechnika* **2024**, *30*, 56–67. [CrossRef]
25. Xiong, X.; Zhang, S.; Chen, Y. Review of Intelligent Vehicle Driving Risk Assessment in Multi-Vehicle Interaction Scenarios. *World Electr. Veh. J.* **2023**, *14*, 348. [CrossRef]
26. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum Contrast for Unsupervised Visual Representation Learning. *arXiv* **2020**, arXiv:1911.05722. [CrossRef]
27. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. *arXiv* **2020**, arXiv:2002.05709. [CrossRef]
28. Chen, X.; He, K. Exploring Simple Siamese Representation Learning. *arXiv* **2020**, arXiv:2011.10566. [CrossRef]
29. He, K.; Chen, X.; Xie, S.; Li, Y.; Dollár, P.; Girshick, R. Masked Autoencoders Are Scalable Vision Learners. *arXiv* **2021**, arXiv:2111.06377. [CrossRef]
30. Tong, Z.; Song, Y.; Wang, J.; Wang, L. VideoMAE: Masked Autoencoders Are Data-Efficient Learners for Self-Supervised Video Pre-Training. *arXiv* **2022**, arXiv:2203.12602. [CrossRef]
31. Khrapak, S.; Ivlev, A.; Morfill, G.; Zhdanov, S.; Thomas, H. Scattering in the Attractive Yukawa Potential: Application to the Ion-Drag Force in Complex Plasmas. *IEEE Trans. Plasma Sci.* **2004**, *32*, 555–560. [CrossRef]

32. Wu, J. Research on Driving Risk Assessment Method Considering Human Vehicle Road Factors. Master's Thesis, Tsinghua University, Beijing, China, 2017.
33. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]