CNN-SVM-based Human-Computer Interaction Model for Automotive Systems in Complex Driving Environments

Mei Gao[1,*], Dan Ye[1], Junjie Zhang[2]
[1]School of Electronic and Electrical Engineering, Anhui Wenda University of lnformation Engineering, Hefei 230012, China
[2]China Mobile Communications Group Anhui Co., Ltd. Hefei 230012, China
E-mail: letter_gao@163.com
[*]Corresponding author

Keywords: complex driving environment, multimedia information collection, man-machine interaction, emotional perception

Received: February 19, 2025

In the complex driving environment, with the increase of task difficulty, the change of diversity and relevance, the phenomenon of perceptual mode conflict, strong cognition or increased difficulty of operation appears when drivers deal with tasks, which affects the execution of primary and secondary tasks. The information expressed and transmitted by multimedia technology is real-time, and only with real-time can we interact and transmit information with users. This paper discusses the design of automobile man-machine interaction based on multimedia information acquisition technology in complex driving environment. Based on users' situational awareness, this paper studies users' interactive needs and experiences in different driving situations, and proposes CNN-SVM (Convolutional Neural Network-support vector machine) emotional perception model. After automatically extracting spectral features through CNN, support vector machine (SVM) is used instead of traditional Softmax classifier to achieve accurate classification of multi class emotions. The experiment focuses on identifying the following core emotion categories: Anger, Neutral, Joy, and Anxiety, and verifies the model's generalization ability in driving scenarios through cross validation. In the spectrum segmentation experiment, the same network structure as CNN Net was used, and SVM was used instead of softmax classifier. CNN Net was used for training each time. After training, use the test sample set to calculate the input features of the softmax classifier, and input the new features into SVM to calculate the classification result of CNN-SVM. Select 500 test set images from the CityScapes dataset for testing, with MIoU (Mean Intersection Point on Consortium) used as the testing metric. The experimental results show that the model has improved the segmentation accuracy of roads, sidewalks, buildings, etc. Compared with several mainstream segmentation algorithms currently available, this algorithm has relatively small improvements for smaller target objects such as lights, signs, vegetables, and riders, with improvements of 1%, 0.5%, and 0.9%, respectively. In driving scenarios, users' judgment of "safety" mainly depends on whether secondary tasks occupy the cognitive and interactive channels of the primary driving task. This article explores how avoiding these two factors can provide users with a sense of security and improve their interaction experience.

Povzetek: Predstavljen je model zaznave čustev CNN-SVM za avtomobilske sisteme, ki v kompleksnih voznih okoljih izboljša interakcijo človek-stroj z natančnejšim prepoznavanjem govornih signalov in uporabniških stanj.

1 Introduction

With the wide application of computer technology and network technology in the field of transportation and the continuous development of on-board technology, revolutionary changes are taking place in the internal space of cars, human-machine interface operation and interaction process. At present, the vehicle interior information model is evolving from a single vehicle state information model to a complex information system. Improving the corporate image of its own brand by providing high-quality products has always been the direction of efforts and exploration. Improving the user experience during the use of automobile products is an important aspect of the project. In order to achieve effective control of vehicles, people and machines must cooperate in depth in perception, decision-making and execution, share vehicle control and decision-making power, and cooperate to complete driving tasks. The core problem of human-computer co driving is still the coordination of human-computer interaction. Due to the non-repeatability of autonomous driving testing scenarios and the limitations of safety, cycle, and cost of testing, it is difficult to rely solely on real vehicle road testing experiments for related technical testing. With the progress of simulation technology, the relevant tests of autonomous vehicle can be carried out in a safer, more

comfortable and more economical environment. As a vehicle test tool, the driving simulation system takes into account the data acquisition functions of the vehicle end and the human factor end. Using simulation technology to realize high-precision restoration of multiple scenes greatly improves the efficiency of human-computer interaction test and evaluation of intelligent vehicles. It has important practical value and significance for the research of intelligent vehicle human-machine co driving technology.

Traditional human-computer interaction research emphasizes "task ability matching", but intelligent vehicles need to further optimize the dynamic balance of "cognitive load attention resources". Predicting driver distraction risk through voice emotion perception and automatically reducing information density on the central control screen. Gesture recognition (such as steering wheel touch/head up display gestures) needs to be coordinated with voice interaction to avoid "modal conflicts" (such as conflicts between voice commands and gesture intentions). The accuracy of camera gesture recognition decreases by more than 30% in strong light/rainy and snowy weather [citation]; Complex gestures (such as 3D spatial gestures) have high learning costs and can easily cause driver anxiety. Voice emotion perception can predict users' acceptance of gesture interaction (such as automatically switching to voice priority mode in an "angry" state). Focus more on task efficiency in static scenarios (such as laboratory environments) and ignore emotional interference in dynamic driving scenarios. In recent years, with the wide application of information technology, intelligent system and network technology in the field of vehicles and the development of on-board information technology, the interior space, man-machine interface, operation and interaction process of automobiles are undergoing revolutionary changes, and the control interface is undergoing essential changes. This change improves the mobility and comfort of drivers, and also increases the risk and danger level of drivers' poor performance of vehicle control tasks [1]. The biggest feature of multimedia information collection technology is that non-intuitive multi-source data information can be intuitively expressed and transmitted to users, and information transmission and communication can be realized by stimulating users' sensory awareness. In fact, multimedia information collection technology uses computer to realize information exchange and transmission with users, which is essentially a man-machine interaction technology. With the improvement of drivers' requirements for car comfort and safety, the traditional in-vehicle network technology is no longer fully competent. Therefore, the core of car man-machine interaction design is based on in-vehicle technical equipment, which provides drivers with a good interactive way to cope with complex driving environment and improve driving experience.

2 Related work

At present, the information model inside the automobile has gradually developed from a single driving and vehicle condition information model to a complex information system including vehicle information, inter-vehicle information, and information of interaction between vehicles and other information carriers [2, 3]. Research on automotive ergonomics started in 1950 s, and a lot of surveying and mapping analysis and basic research work have been done. Especially in developed countries, they attach great importance to the ergonomics factors in automobile development. Krishan et al. put forward the modeling and optimization research of human upper limb posture of automobile man-machine interface, aiming at the matching analysis needs of automobile man-machine interface design [4]. As autonomous driving technology evolves towards L4/L5 levels, the limitations of single sensors or deterministic algorithms in complex dynamic scenarios are becoming increasingly prominent. Gao et al. [5] provided a systematic review of intelligent vehicle autonomous driving tracking technology based on fuzzy information processing and multi-sensor fusion. Focus on analyzing its collaborative mechanism in modeling environmental perception uncertainty, spatiotemporal alignment of multimodal data, and dynamic decision optimization. Yang et al. extracted the rotation invariance feature of hand, and used Fourier operator, edge histogram and boundary moment invariants as the basis of pattern recognition [6]. Ji et al. have implemented a virtual environment system, which is mainly used in large-scale biological modeling. Gestures are used as inputs to control devices in the virtual environment, which provides an interactive interface for biological modeling [7]. Chen and Jia think that the virtual assistant of information system in human-computer interaction of electric steam is helpful to improve users' emotional experience, establish trust mechanism and emotional communication between users and the system, and contribute to the establishment of information system model [8]. Fang and Wang user-centered user experience design by collecting user information, refining user requirements, simulating user scenarios, real-time user testing and feedback has become the goal and purpose of human-machine interface design [9]. Huang et al. used CNN and short-term and long-term memory architecture to study end-to-end unmanned driving, and directly controlled the direction of the vehicle [10]. Wang et al. used the top-down structure to combine low-resolution features with high-resolution features, and built a feature pyramid with similar semantic level, which improved the effect of multi-scale object detection [11].

In the case of human-computer interaction, besides ensuring driving safety, driving pays more attention to getting a good interactive experience in the process of human-vehicle interaction. Experience is a typical emotional element, and it is a higher level of experience in cognitive level. In complex driving environment, with the increase of task difficulty, the change of diversity and relevance, the phenomenon of perceptual mode conflict, strong cognition or increased difficulty of operation appears when drivers deal with tasks, which affects the execution of primary and secondary tasks [12]. The information expressed and transmitted by multimedia technology is real-time, and only with real-time can we interact and transmit information with users. Real-time is

mainly manifested in the fact that when multimedia information collection technology interacts with users, multiple kinds of information interact synchronously under multiple sensory stimuli. This paper discusses the design of automobile man-machine interaction based on multimedia information acquisition technology in complex driving environment. Based on users' situational awareness, it is of great guiding significance to study users' interactive needs and experiences in different driving situations, to construct the theoretical framework of in-vehicle human-computer interaction design of intelligent cars, and to propose appropriate design methods.

The key factors of interaction design include visual experience. Buttons of different types of functions are designed in different shapes and sizes. And distinguish the main task from the secondary task in a guiding manner, so as to quickly identify and process the main task. Secondly, auditory experience. In the on-board environment, voice control interaction is considered to be the most convenient human-computer interaction mode with the least impact on driving. Tactile experience. In the design of central control information entertainment products, human tactile characteristics should be considered and appropriate vibration tactile reminders should be used. It can provide tactile sensation for blind operators during driving, and minimize operations that must rely on visual assistance. Although current research has made some progress in the field of automotive human-computer interaction, such as basic research in automotive ergonomics, gesture recognition, user experience design, autonomous driving visual processing, and multi-scale object detection, it mostly focuses on optimizing single dimensions or local functions, and there is relatively little systematic research based on multimedia information collection technology in complex driving environments to comprehensively integrate visual, auditory, tactile and other sensory experiences for user situational perception. There are gaps in the existing research on multimodal interaction synergy and interaction design theoretical frameworks that dynamically adapt to different driving scenarios. The solution proposed in this article aims to fill this research gap by constructing a theoretical framework for intelligent car human-machine interaction design based on user context perception, and proposing a multi-sensory interaction design method that adapts to complex driving environments to enhance user interaction experience and driving safety, Table 1 is the summary table of existing methods.

Table 1: Summary table of existing methods

Method Type	Algorithm method	Evaluation indicators	Benchmark test dataset	Performance	Main limitations
Traditional HCI methods	Threshold method, template matching	Accuracy, precision, recall rate	Self-built dataset (small-scale)	Accuracy: 70% -80%	Relying on artificial feature engineering, poor generalization ability; Difficult to handle complex motion patterns such as jumping and going up and down stairs.
Machine learning methods	SVM, random forest	Accuracy, F1 score	UCI HAR, WISDM	Accuracy: 85% -90%	Feature extraction requires domain knowledge and is difficult to capture high-dimensional nonlinear relationships; Sensitive to data noise.
Deep learning methods	CNN, LSTM, CNN-LSTM hybrid model	Accuracy, precision, recall rate MIoU	UCI HAR, WISDM, Self-built large-scale dataset	Accuracy: 90% -95%	High demand for computing resources; Some models have poor classification performance for short-term movements such as jumping.
SOTA method	Attention mechanism CNN, Transformer	Accuracy, precision, recall rate MIoU	PUSH (Large Scale Gym Exercise Dataset)	Accuracy: 92.1% (50 categories)	High model complexity and slow inference speed; The classification performance of small sample motion categories (such as rare motion) decreases.

3 Research method

3.1 Situational awareness and human-computer interaction of intelligent automobile

Automobile human-computer interaction design, such as the central control frame layout of Audi A8. The left shortcut menu can switch applications on any interface, and the right application menu can be switched by sliding left and right. The structure of the application menu is clearer than before. The icon adopts an easy-to-understand pseudo object style and combines Audi's brand elements to increase the overall design sense of the icon. Interactive feedback mode adds design highlights. For example, if you touch without pressing the icon on the screen, the icon will be displayed in the form of micro animation. Which increases driving pleasure. Situational awareness involves the knowledge of cognitive science and psychology, which refers to the user's perception and cognition of the situation. Situational awareness enables users to perceive the current environment, understand its meaning and predict the future situation, and make behavioral decisions based on the cognitive results of the current situation. Situational awareness is an individual's conscious dynamic reflection of the environment, which reflects the past, present, future and potential characteristics of the environment. Individuals form a psychological cognitive model of the external environment [13, 14]. The cognitive result of the current situation is the premise for users to make decisions and perform actions, and the main source of users' situational awareness is their perception and understanding of the elements of the current situation and their prediction of the future situation.

Traditional man-vehicle interaction is mainly based on the interaction of steering wheel, physical buttons and other related devices (such as steering control lever). The driving environment is relatively complicated, especially at the auditory level. Words are easily blurred by other words, special tone signals are easily masked by other similar tone signals, and most words are incompatible with each other. In a complex driving environment, drivers interact with the equipment in the car to complete various tasks. In order to clearly and intuitively present the task flow and grasp the details and key points of the task, it is necessary to conduct a structured analysis of the task [15]. Therefore, in order to achieve a faster and better development of multimedia information collection technology, we should combine the current development trend of software technology, cloud technology and big data to realize intelligent functions such as cloud storage and data mining of multimedia information collection technology, which can greatly promote the efficiency and effect of multimedia information collection technology in information expression, transmission and human-computer interaction. Situation Awareness (SA), as the core ability of drivers to dynamically perceive the driving environment, directly affects the input data quality, decision logic, and output reliability of human-computer interaction models (such as CNN-SVM emotion perception model). Drivers may lose key data due to distraction or environmental obstruction (such as no GPS signal in the tunnel). By using the CNN-SVM model for multimodal fusion (combining camera and millimeter wave radar data), the impact of single sensor noise can be reduced, resulting in improved accuracy in target recognition.

Because the human body is a unified whole, the depth information in space is continuous. Therefore, the depth histogram always obeys a specific distribution. There is a gap between the depth value of human body in space and the depth value of background. Therefore, the area that constantly searches for the non-zero depth plane to the zero depth plane is the candidate area of human body. The features extracted by static gesture recognition mainly include low-level geometric features (points, lines, faces, angles) and high-level combination features of the hand.

The number of feature descriptors is k, that is, after selecting k -dimensional data samples, Mahalanobis distance is used to classify features. Mahalanobis distance is a measure of the covariance of the data space, which is a method of "twisting and stretching" the data space and then measuring it [16]. Mahalanobis Distance can adaptively adjust the importance of different dimensions by considering the covariance relationship between features, and in most cases, it is superior to Euclidean Distance and Cosine Distance, as shown in (1).

$$D=(x,y)=\sqrt{(x,y)^{T}\sum^{-1}(x,y)} \tag{1}$$

Where $\sum^{-1}$ is the inverse matrix of covariance matrix.

User's behavioral psychology reflects the process in which users make decisions on current tasks by using existing experience and external rules. Users use their left and right brains to make a decision together, in which the left brain is responsible for logic and rationality, while the right brain is responsible for creativity and emotion [17]. Judging users need to collect information and all possibilities when facing tasks, so they often take a long time to make decisions. Generally, perceptual users will make actions immediately according to the status of the current task, which won't delay too much time.

The purpose of endpoint detection is to detect the actual and effective speech part from the speech signal input to the computer, which plays a very important role in speech recognition. It can not only reduce the calculation of acoustic model training in the recognition process, but also have a positive impact on the recognition accuracy of the speech recognition system.

Let the short-term energy of the n th frame speech signal $x_n(m)$ be represented by E_n, and the calculation formula is (2):

$$E_n=\sum_{m=0}^{N-1}\left|x_n(m)\right|^{2} \tag{2}$$

The reason why short-term energy can be used to judge actual speech and noise is that the energy of noise is less than that of actual speech.

Image feature extraction will be affected by quantum noise and image stains, so the algorithm takes into account the elimination of noise and image stains when extracting contour features. The origin $O(x_0, y_0)$ and the point set $P_l(x_l, y_l)$ in the obtained polar coordinate (r, θ) are respectively converted into points of rectangular coordinate system (x, y), as shown in (3) and (4).

$$x_l = r_l^n \cdot \cos\theta_l + x_0 \tag{3}$$

$$y_l = r_l^n \cdot \sin\theta_l + y_0 \tag{4}$$

Where $\theta_l = l \cdot \Phi, r_l^n$ is the value of normalized r_l .

In digital images, the above conditions can undoubtedly be met. Because the linear gray level transformation of the region will affect the moment characteristics; In order to describe the shape characteristics of the target, the influence of linear gray scale transformation can be eliminated by operating on the binary target area.

Let $\{I(x, y), x, y = 0,1,\cdots, N-1\}$ be a regional binary image or a suppressed background image, then its $(p+q)$-order statistical moment is defined as (5):

$$m_{pq} = \sum_{x=1}^{n}\sum_{y=1}^{n} I(x, y)x^p y^q \tag{5}$$

p corresponds to the moment in the x dimension, q corresponds to the moment in the y dimension, and the order represents the index of the corresponding part.

In image processing, the similarity between individuals can be judged by a variety of measurements. The most representative of these measurement methods is distance measurement (6):

$$d_{ij} = \sum_{i=1}^{K}\left|x_i^l - x_j^l\right| \tag{6}$$

In which, the dimension vector of the object i is represented by x_i , and each element of the vector is represented by $x_i^l (l = 1,2,\cdots, K)$.

Select the binary code string uniquely generated by n point pairs (x, y), which are selected according to certain rules in the $S \times S$ area, such as (7):

$$f_n(p) = \sum_{1 \le i \le n} 2^{i-1}\tau(p; x_i, y_i) \tag{7}$$

$f_n(p)$ in the above formula is a descriptor. n is often weighed according to different application scenarios.

Situational awareness is essentially the user's perception, understanding and prediction of the interactive environment and its future development. Different situational factors will affect people's cognitive results and lead to different behaviors [18, 19]. The situation of users includes interpersonal behaviors, and the research process includes users' psychological cognition and interactive behaviors. Environment mainly refers to the current usage scene, and also includes some external objective environmental factors, such as light, weather, etc.

Understanding users' situational awareness and the influencing factors of the formation of situational awareness, based on users' cognitive characteristics, and putting forward reasonable design strategies for the purpose of improving users' cognitive level and interaction efficiency can improve users' interaction experience in automobile man-machine interaction.

3.2 Design of emotion perception model of human-computer interaction in automobile

An important requirement of human-computer interaction technology is that information interaction and feedback must be real-time. Real-time means that when users interact with computers or other information output devices, they should be able to understand users' intentions and give feedback quickly to realize interactive transmission. Multimedia information collection technology is to express, transmit and interact all kinds of data and information resources in a unified way, so all kinds of information data must have good coordination to ensure the smooth implementation of information interaction [20, 21]. Therefore, at present, the sound control of vehicle information system mainly focuses on the sound command based on navigation system. The biggest advantage of navigation lies in its limited geographical location information, and its language information format is relatively simple. To design a complete human-computer interaction system, we must consider three factors: human factors, system equipment factors and interactive environment factors. Such as sound, tactile devices, visual input and output devices, etc. Interactive factors refer to the core of interactive functions, such as system software.

In the interface design of automobile cab, the ergonomic interface elements used in the display device design are the most complex and diverse, and the types of information carried in the display device are diverse and complicated. With the continuous innovation and development of multimedia technology in the car, the information that drivers need to deal with while driving is becoming more and more diversified, which obviously will inevitably increase certain potential safety hazards. In general, when designing the display device, the operator is required to have a good central vision and to be able to achieve normal peripheral vision. For the driver, it is necessary to concentrate his energy in the process of driving. At this time, if there are many kinds of information for the driver to process at the same time, it will bring some potential safety hazards.

The traditional speech emotion perception model is mainly divided into several schools, one is the traditional school of manual features+machine learning classifier, and the other is the neural network school based on manual features+neural network classification. Compared with existing HCI emotion recognition models, the CNN-SVM model has significant advantages in feature extraction ability, generalization performance, multimodal fusion

potential, and real-time optimization. The CNN-SVM model can control the generalization error within 8% through transfer learning (such as pre training based on ImageNet), making it suitable for cross-cultural and cross linguistic emotion recognition tasks. For example, in vehicle localization, the accuracy of the model's recognition of unseen driving scenarios (such as nighttime and rainy days) decreases slightly. The accuracy of traditional HCI models decreases by about 30% in cross dataset testing, and they are sensitive to complex scenes such as lighting changes and occlusions. For example, in vehicle localization, traditional models have poor robustness in recognizing small targets such as traffic signs. In this paper, a CNN-SVM (Convolutional Neural Network-support vector machine) emotion perception model is proposed. The input is the spectrogram of audio, and the feature is extracted automatically by CNN structure, then the extracted features are input into a multivariate SVM for classification, and then the results are output. The detailed structure of the network model is shown in Figure 1.

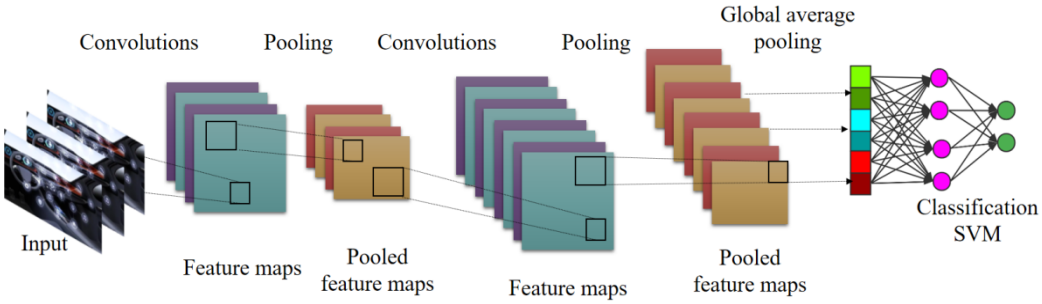

Figure 1: CNN-SVM emotion perception model process

The CNN feature extraction layer adjusts the gradient descent stability by learning rate and batch size to avoid overfitting or slow convergence. The SVM classification layer optimizes the inter class separability of high-dimensional feature space through kernel function selection, improving the recognition accuracy of complex emotional patterns. By introducing attention mechanisms such as CBAM to enhance CNN's ability to focus on time-frequency features of spectrograms. Multi modal inputs, such as combining text transcription with audio features, can be explored to enhance the robustness of emotion recognition. Endpoint detection generally adopts double threshold detection method based on zero-crossing rate and short-time energy. This is because speech signals are generally composed of unvoiced, unvoiced and voiced segments. The silent segment belongs to the background segment, and its average energy is the lowest. Let the voice sampling value at time n be $x(n)$, and the result after pre-emphasis processing is (8):

$$y(n) = x(n) - ax(n-1) \tag{8}$$

Softmax classifier is mainly used in the field of multivariate classification, although traditional binary classification can also be carried out by using multiple binary classifiers to classify and then get the best results. The calculation process of softmax (9):

$$y_i = \frac{e^{x_i}}{\sum_i^m e^{x_i}} \tag{9}$$

For the classification of multi-labels, we should still use binary classifier to carry out multiple groups of experiments, and select the first few with the highest probability.

The speech process of people's speech formation is a double random process, because it is impossible for each speech frame to correspond to a different state, and then the whole speech signal will get a bunch of states. There are as many states as there are speech frames, and whether it is the same person's pronunciation in different time periods or different people's pronunciation, different speech signals will be produced, which are quite different from each other. Therefore, in fact, each state will correspond to multiple adjacent speech frames, which is reasonable because each frame is very short.

If the excitation function is sigmoid nonlinear function, the neuron output is shown in the following (10):

$$f(x) = \frac{1}{1+e^{-x}} \tag{10}$$

Assume that there is a pile of original data, that is, training sample sets, and then build a neural network. Firstly, initialize its weight parameters, and then iteratively train its weight parameters through these training sample sets.

To solve the problems mentioned above, batch gradient descent method is generally adopted. Firstly, the weight parameter matrix w is randomly initialized, and then it is updated by iterative training of the following (11):

$$w_{jk}^{l} := w_{jk}^{l} - \eta \frac{\partial C}{\partial w_{jk}^{l}} \tag{11}$$

$:=$ is the assignment operation, η is the learning rate, and $\frac{\partial C}{\partial w_{jk}^{l}}$ is the partial derivative of the loss function C to the weight coefficient of the k th neuron in the l layer.

3.3 Analysis of interface information in different task situations and user categories

Automobile control equipment refers to the equipment that can change the running state of automobile through functional operation. The rationality of the design of vehicle control equipment directly affects the working efficiency of the whole human-vehicle operating system. In the traditional driving process, the driver's hand can't leave the steering wheel, and it needs to meet other vehicle operating functions and ensure the driver's operating comfort. From the perspective of location and form, the location of the future human-computer interaction graphical interface is no longer limited to the driver's seat, but integrated in the roof, window, etc. The graphical interface is presented through projection, holography, electronic grid, intelligent physical surface, light and other technologies. Future concept vehicles will create interactive fun for users on the basis of emphasizing adaptation to human-machine engineering. At the behavioral level, the current concept car introduces more "black technology", and makes more innovations in human-computer interaction from the perspective of adapting to human-computer engineering and perception. The emotional design of intelligent mobile space in the future needs to integrate the current fragmented functions based on these innovations. From the unified interface such as the on-board steward to the user, the output of the function and interaction layer can improve the efficiency of human-computer cooperation and cultivate the user's emotion.

Central control display is often used for vehicle information display, multimedia display and control, electronic navigation display, etc. Especially, with the increase of all kinds of automobile automation technology, the functions of traditional central control area become more and more complicated, and it becomes more and more difficult to operate by physical keys and the learning cost is higher and higher. Digital display uses digital technology to directly display relevant parameters. The observer's reading of data information is relatively simple and intuitive, and has high accuracy, but it can't give people a vivid image, which makes information reading lack a concept of "degree". At present, two kinds of information display methods are commonly used in automobile instruments.

The driving information area mainly displays vehicle information during driving and parking, such as vehicle road information, vehicle speed, electric quantity information, vehicle body equipment information, etc. The function control area is mainly the detailed control area after the user enters the function of "no words" through function navigation, and the content of this area changes according to the user entering different functions. According to the analysis of users' characteristics and behaviors, based on the functional requirements in the current scene, low-level interaction mode is used in the design of intelligent electric vehicle's on-board central control man-machine interaction information interface to reduce users' learning and use costs, and provide customizable content, so as to reduce the learning cost of primary users, improve the fluency of intermediate users, and increase the available space for expert users.

In the driving situation, the user's judgment of "safety" mainly depends on whether the secondary task occupies the cognitive channel and the interactive channel of the main driving task. By avoiding the two, the user can be given a sense of security and the interactive experience can be improved. In the automatic driving situation, the driver basically does not undertake the driving task and driving load, and the application characteristics of various kinds of interactive control methods are basically consistent with those of manual driving situation and auxiliary driving situation. At this time, the consideration of interaction design should be mainly based on the target customers. Based on the perception and judgment of the current road conditions, the system puts forward the warning of dangerous driving or collision, which can inform users to respond in advance, and even actively avoid it through its lane keeping system, which greatly improves the driving safety.

The previous research work on vehicle location and recognition can be roughly divided into two parts: locating and recognizing vehicles in static images and locating and recognizing vehicles in video sequences. The main difference between them is that video can use dynamic information, while static images have less prior knowledge. Whether a practical vehicle positioning and recognition application system has good positioning and recognition performance is of course mainly determined by the performance of the core algorithm model, but it also depends on the design of the system to a certain extent. The number of car frames is determined by the difference pictures, and the gray values in the row and column directions of the processed difference pictures are statistically analyzed. And the reusable image of the frame boundary and the image pixel coordinate position is obtained according to the judgment result of the number of frames. The actual position and size of the vehicle can be obtained through the coordinate correspondence to realize the recognition and positioning of the vehicle.

Because of the great difference between images, the number, size and position of cars included in each image are different. In order for the system to locate and identify vehicles of different sizes in any image, the image sampling method based on pyramid is used. Window scanning and pyramid image sampling are performed on the whole image to solve the problem of different sizes and positions of vehicles to be located and identified in the image. All the windows identified as vehicles are merged to get the final recognition result, and the vehicle is located according to the position information of relevant windows. The process of pyramid sampling an input image is shown in Figure 2.

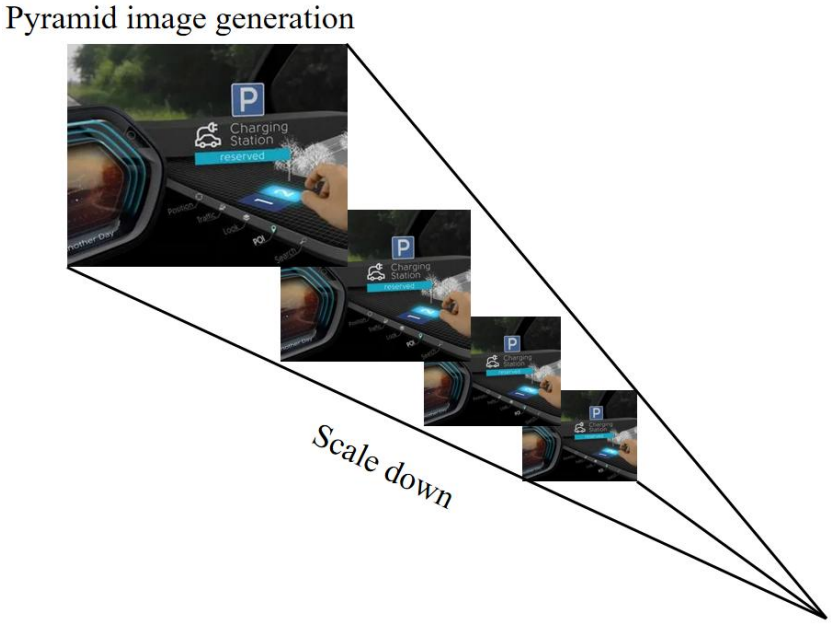

Figure 2: Pyramid sampling

Vehicle positioning depends on the merging of windows. An overlapping system $coef$ is adopted to reflect the overlapping degree of the two windows, so as to judge whether the two windows overlap and merge. Two different windows are represented by $W(x_1, y_1, w_1, h_1), W(x_2, y_2, w_2, h_2)$, then the calculation method of $coef$ is (12)-(14):

$$x = \min(x_1 + w_1, x_2 + w_2) - \max(x_1, x_2) \quad (12)$$

$$y = \min(y_1 + h_1, y_2 + h_2) - \max(y_1, h_2) \quad (13)$$

$$coef = \frac{xy}{s_x s_y} \quad (14)$$

$coef$ represents the coverage degree of $W(x_1, y_1, w_1, h_1), W(x_2, y_2, w_2, h_2)$. If there is more coverage, the coefficient $coef$ is larger. When two different windows do not cross each other, the $coef$ coefficient is defined as zero.

The system scans the windows identified as cars and merges them. In this process, the vehicle is located by the recorded window position information. The two processes of identification and location are combined into one, and the final vehicle location information will be output in the system interface.

In one's own information processing process, only by processing the noticed information can one form a valuable memory, in which the composition of attention has certain selectivity and concentration. Users subjectively evaluate information based on their own preferences and needs, which determines the orientation of attention. When the information demand is a clear desire, it will trigger the motivation to learn information actively, thus generating attention. Therefore, if the stimulation information provided by the interface meets the user's needs under the current task and meets the user's thinking characteristics, it may cater to the user's motivation, generate beneficial attention and effectively help the user achieve the goal.

Touchscreen interaction, as an intuitive and natural way of interaction, has made significant progress and been widely used on handheld mobile devices. However, the application of touch screen interaction in automotive information systems is subject to certain limitations. This is mainly due to the special nature of the internal environment of the car, such as the need for the driver to remain focused during driving, space limitations, and operational safety factors, which make traditional touch screen interaction not always applicable in the car environment. In traditional cars, the interaction between the driver and the vehicle often relies on the movement of the head and hands, such as turning the head to view the dashboard or reaching out to operate the center console. However, with the development of technology, this situation is gradually changing. In order to improve the efficiency of information reception and operational convenience for drivers, we can use multi visual information to optimize interface design. Specifically, by avoiding excessively long visual scanning paths and highlighting key information, the visual burden on drivers during operation can be reduced, thereby improving the speed and accuracy of their information processing. Based on this, we can consider optimizing perceptual modal resources by adjusting the interface design. For example, important information can be placed in a location that is easy for drivers to view, or key information can be highlighted through visual elements such as color and shape. These design strategies aim to enhance drivers' perception and understanding of information, thereby improving their driving experience and safety.

In this paper, a "self-learning" strategy is adopted, which can train the discriminators by using unlabeled target data. The main idea is that the training discriminator can generate a confidence graph, which can find out the

area where the distribution between the prediction result and the source domain label is close enough, and then binarize the segmented prediction confidence graph and the confidence graph corresponding to the source domain label, and the semi-supervised loss constructed is as (15):

$$L = -\sum_{h,w}\sum_{c\in C} J\left(D\left(S\left(I_n\right)\right)^{(h,w)} > T\right) \sum \hat{Y}_n^{(h,w,c)} \log\left(S\left(I_n\right)^{(h,w)}\right) \tag{15}$$

Where $J(\cdot)$ refers to the index function, and T represents the threshold parameter of unlabeled target data. During the training, we assume that the self-study goal $\hat{Y}_n$ and the value of index function are constant.

For the discriminant network in the model, we adopt a structure similar to that of the traditional CNN, but we use the full volume layer instead of the full connection layer to better retain the spatial information. Its definition formula is (16):

$$y_i = \begin{cases} x_i & if\ x_i \geq 0 \\ ax_i & if\ x_i < 0 \end{cases} \tag{16}$$

Where $a \in [0,1]$ is the correction parameter, we set $a = 0.3$ (obtained through many experiments) in this paper.

Drivers have different screens to obtain information in different situations. Usually, the driver's line of sight will not deviate from the road when the vehicle is moving, and the information on the dashboard is the equipment that needs the least head movement to obtain information. In traffic jam, parking, etc., the devices for drivers to obtain information are mostly central control screens. General system design requires users to pay a certain learning cost. This principle aims at thinking about the design of information architecture from the perspective of users, reducing the learning cost of users and improving the usability and usability of the interface.

In terms of hearing, the voice assistant in the interface helps the driver to complete the functional operation without looking away, and improves the driver's driving performance; The interface key sound and reversing prompt sound give the user real-time operation feedback. In terms of touch, the user interface is that the vibration feedback embedded in the screen can stimulate the user's response as quickly as the auditory feedback.

4 Analysis and discussion of results

The collected speech signal is pre-emphasized by mat lab tool, and the speech frame is obtained by windowing and framing. Then through endpoint detection (the highest and lowest energy thresholds are set to 10 and 1, and the zero-crossing rates are 10 and 5; The longest and shortest speech gap time is set to 20 and 15), the valid speech frame is intercepted from the original speech.

In the field of human-computer interaction, acoustic recognition is a key link. As an important medium for human-computer interaction, accurate recognition of speech is the foundation for achieving efficient and natural interaction. In this experiment, research was conducted using HMM isolated word recognition system by preprocessing the training speech and extracting MFCC features. Although the current results show differences in recognition rates between specific and non-specific populations, acoustic recognition research has significant implications for human-computer interaction. Deep learning extracts new features to improve the original MFCC features. If the recognition rate can be improved, it will make the recognition of voice commands in human-computer interaction more accurate, thereby optimizing the interaction experience and promoting the development of human-computer interaction towards a more intelligent and convenient direction. In this experiment, all 260 pieces of training speech were preprocessed, the effective speech frames were intercepted by endpoint detection, and MFCC (Mel-scale frequency cepstral coefficients) features of different dimensions were extracted. The correct rates of 50 words recognized by test sets 1, 2 and 3 are shown in Table 2 and Figure 3.

Table 2: Experimental result

Characteristic	Test set 1	Test set 2	Test set 3
12	0.913	0.819	0.847
24	0.903	0.817	0.871
36	0.863	0.774	0.906
48	0.878	0.754	0.864
60	0.894	0.786	0.929
72	0.84	0.793	0.939
84	0.846	0.814	0.844
96	0.897	0.827	0.831
108	0.864	0.779	0.898
120	0.907	0.745	0.862
132	0.877	0.766	0.852
144	0.831	0.774	0.858

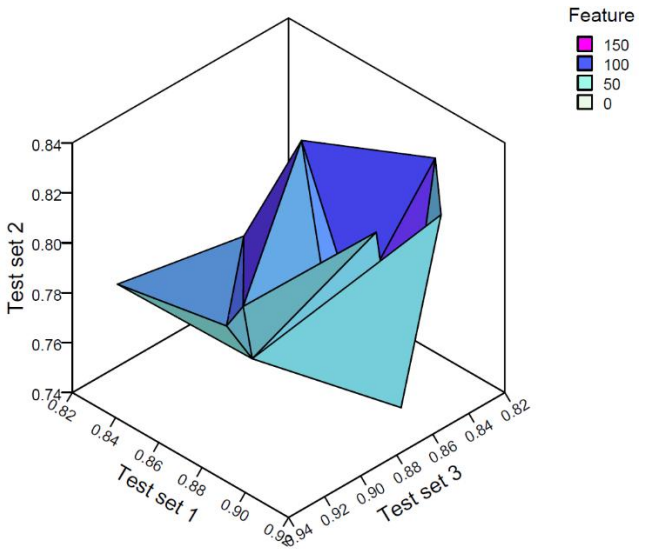

Figure 3: Accuracy of different dimensions

It can be concluded that the speech recognition rate of HMM isolated word recognition system for specific people in test set 1 is average, but it is obviously poor for non-specific people in test set 2. Therefore, it is necessary to use deep learning to extract new features of training acoustic model to improve the original MFCC features with different dimensions, and to verify whether it can make isolated word speech recognition system obtain satisfactory recognition results through experiments.

The stride is a key parameter in spectrum segmentation experiments. In the process of using CNN Net for feature extraction and combining with SVM for classification, the stride determines the movement interval of the sliding window on the spectral data. Specifically, when segmenting the spectrum, different stride settings will change the position and range of feature extraction for each sliding window. From the comparison of accuracy under different amplitudes in Table 3 and Figure 4, it can be seen that the stride size directly affects the accuracy of the final classification. Appropriate stride can enable the model to more effectively capture key feature information in the spectrum, thereby improving classification performance. This also explains why there is a difference in classification accuracy between CNN-SVM and CNN Net under different stride sizes. In the spectrogram segmentation experiment, at the same time, the same network structure as CNN-Net is adopted, the softmax classifier is replaced by SVM, and CNN-Net is used for training each time. After training, the input features of softmax classifier are calculated by using the test sample set, and the new features are input into SVM, so that the classification results of CNN-SVM can be calculated. As shown in Table 3 and Figure 4.

Table 3 Comparison of accuracy under different strides

Strides	CNN-Net	CNN-SVM
16	0.866	0.889
32	0.898	0.98
48	0.797	0.832
64	0.71	0.871
80	0.812	0.757
96	0.653	0.838
112	0.639	0.781
128	0.665	0.748
144	0.579	0.723
160	0.544	0.593
176	0.469	0.639
192	0.416	0.502
208	0.44	0.502
224	0.418	0.524

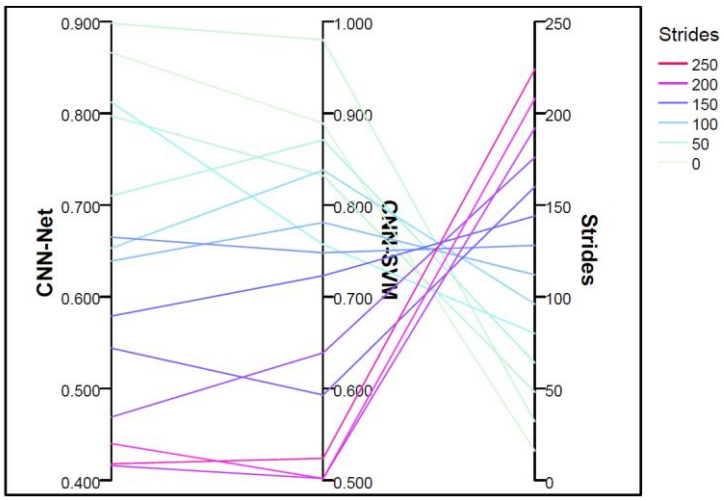

Figure 4: Comparison and distribution of accuracy under different strides

It can be seen that the classification effect of CNN-SVM is better than that of CNN-Net, and the accuracy rate is increased by 12.38%. It can be found that using CNN-SVM classifier to select the appropriate kernel function can always achieve a slightly better classification effect than CNN-Net, which proves that the emotion perception model proposed in this paper is more efficient than the traditional emotion perception model. In the case of poor feature extraction, the radial kernel function can be used to map the features of nonlinear space to linear space, and the linear can't be divided into linear separable, so good results are achieved.

In order to further analyze the system, a simple traditional three-layer ANN (artificial neural network) test system and SVM test system are also implemented. In the test system, vehicle samples are trained, and the same pyramid technique is used in the process of location and recognition. The results of vehicle location and recognition on 100 randomly selected pictures are shown in Figure 5. The final test set error rates of different algorithms are shown in Figure 6.

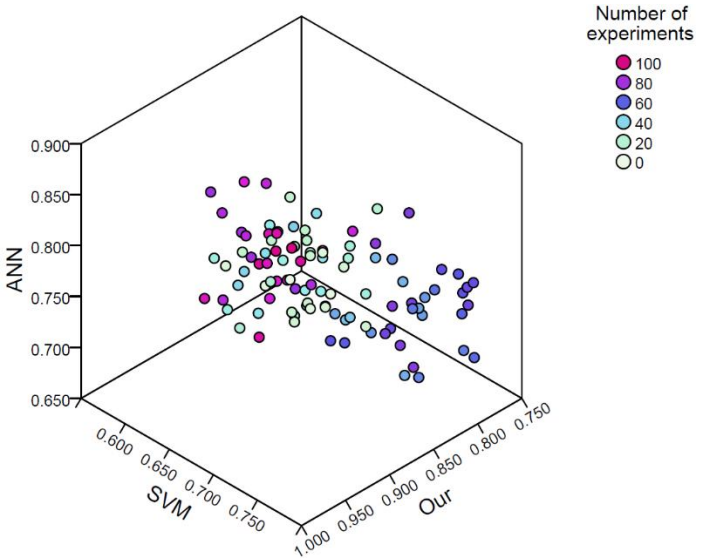

Figure 5: The result of positioning and identification

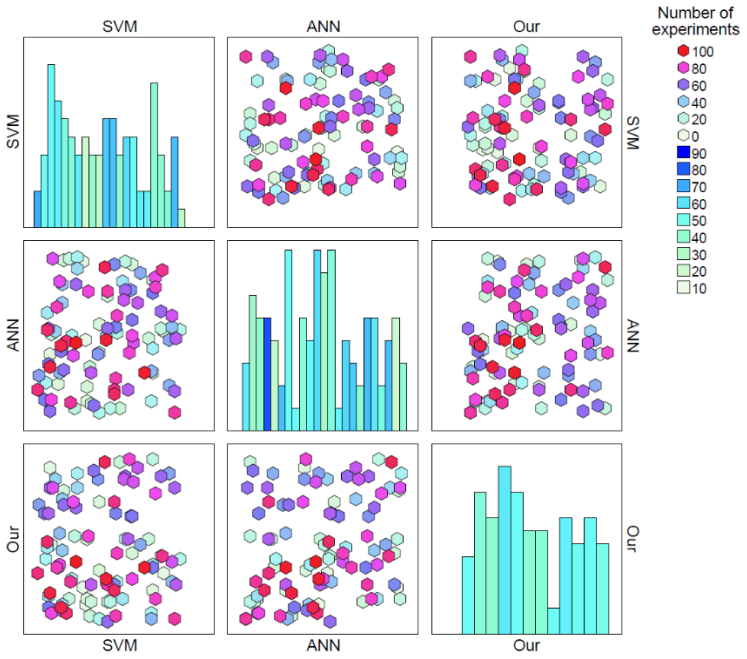

Figure 6: Error rate of different algorithm test sets

It is found that when the sample size of the traditional three-layer neural network is too large, there will be serious data over-fitting, which will lead to a high error rate of 23.68% on the test set. The error rate of SVM on the test set is not much different from that of the deep network of this system, and the recall rate of the two systems is not much different. The designed vehicle location and recognition system has good robustness, which is due to the characteristics of deep neural network, that is, extracting features from the original image, and each layer of neural network abstracts higher-level features from the previous layer, which can effectively identify vehicles of different positions and sizes. It is an effective vehicle location and recognition method.

This article uses the leftImg8bit_trainvaltest (training/validation/test set) and gtFine_trainvaltest (fine annotation) subsets officially released by CityScapes, with a total of 2975 training images, 500 validation images, and 1525 test images. In order to further verify the effectiveness of this algorithm, 500 test set images in CityScapes data set are selected for testing. Our model is compared with several mainstream segmentation algorithm models. MIoU (Mean Intersection over Union) is used as the test index. The experimental results are shown in Figure 7.

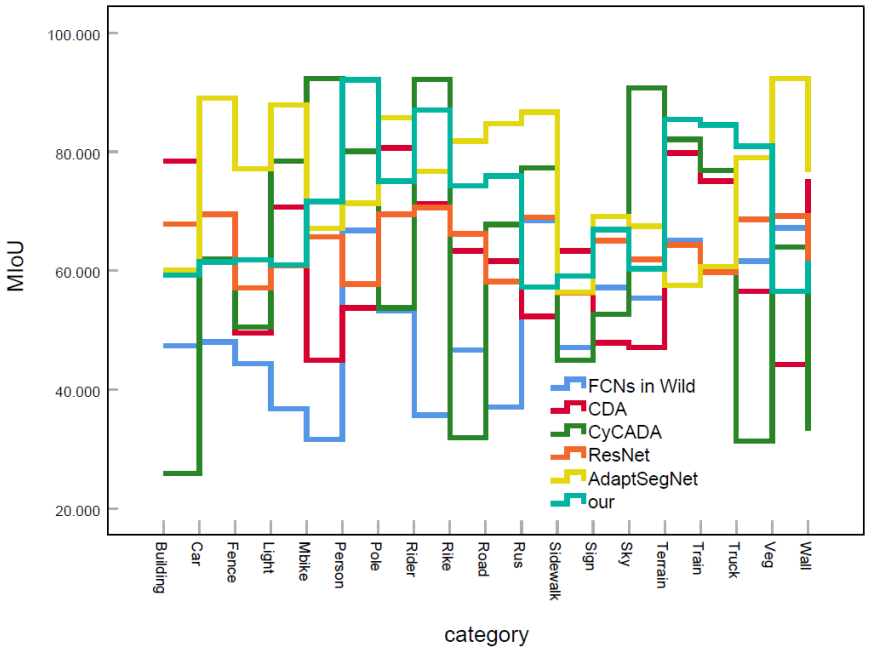

Figure 7: Comparison of experimental results of various segmentation algorithms

It can be seen that the segmentation accuracy of roads, sidewalks, buildings, etc. is improved by the results of this model. Compared with several mainstream segmentation algorithms at present, this algorithm has a smaller improvement in smaller target objects, such as light, sign, veg and rider, which are improved by 1%, 0.5% and 0.9% respectively. Through significance testing of the experimental results, it was found that the 1%, 0.5%, and 0.9% increases in small object recognition rates were statistically significant and not within the statistical error threshold range. Meanwhile, based on the analysis of practical application scenarios, although these improvements may seem insignificant, they can effectively improve the reliability and accuracy of the system in practices such as high-precision map construction and autonomous driving environment perception that require high segmentation accuracy, and have practical significance.

Through the learning rate domain adaptive method, the performance of the generator and discriminator in the countermeasure network is further optimized, so that the learning ability of the network is enhanced, thus improving the segmentation accuracy of the model. In addition, the algorithm adds a convolution layer to the discriminator of the confrontation network, which improves the discrimination ability of the proposed network.

The recognition accuracy in simple cases is shown in Figure 8. In complex situations, only using Fourier features and the recognition accuracy of this method are shown in Table 4 and Figure 9.

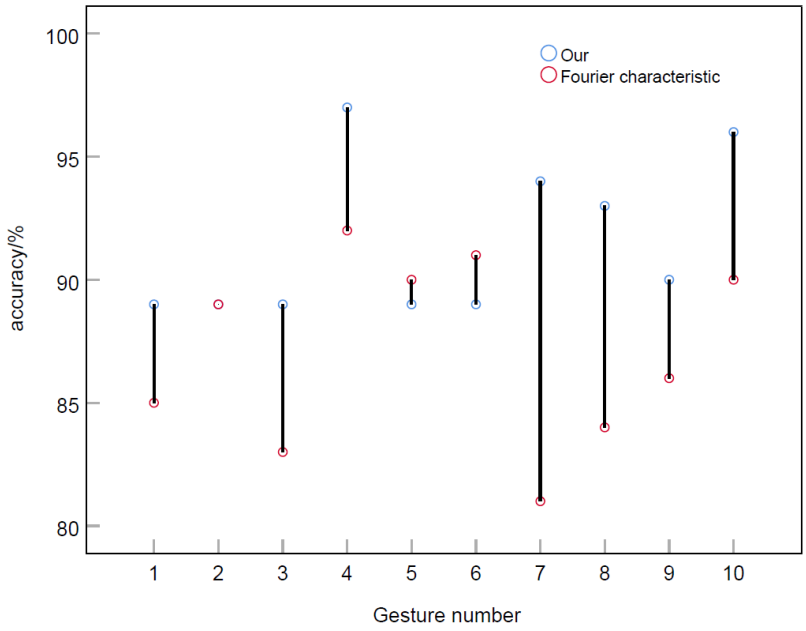

Figure 8: Comparison of simple situation recognition accuracy

Table 4: Identification and comparison

Gesture number	Our/%	Fourier characteristic%
1	82	78
2	87	71
3	80	71
4	84	78
5	85	76
6	87	69
7	84	77
8	86	77
9	80	78
10	79	78

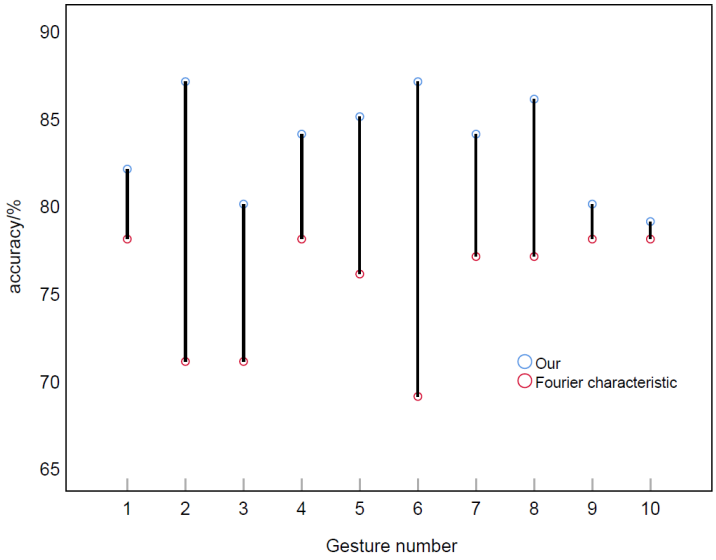

Figure 9: Comparison of accuracy of complex situation identification

In simple cases, the average recognition rate of gestures 1-10 using only Fourier descriptors is 86.6%, and the average recognition rate of multi-feature combination is 94.3%. Therefore, using multi-feature recognition rate can improve the recognition accuracy in simple cases. In complex situations, the recognition rate is higher than that of single feature recognition. The global feature shape of the gesture image represented by Fourier descriptor is not effective for gesture recognition with serious deformation. However, the method in this paper does not need to depend on the segmentation effect as heavily as Fourier descriptor.

To comprehensively verify the effectiveness of the CNN-SVM model, it was compared with traditional classifiers such as HMM (Hidden Markov Model), ANN (Artificial Neural Network), and standalone SVM on subtasks of CityScapes (such as road/non road binary classification). Compare with mainstream segmentation models such as U-Net, PSPNet, DeepLabV3+on the CityScapes complete dataset (19 semantic segmentation categories); Verify whether the improvement of CNN-SVM is significant through paired t-test ($p<0.05$). As shown in Table 5.

Table 5: Experimental result

Model	Accuracy (%)	Recall (%)	F1 score (%)	p-value (vs. CNN-Net)
HMM	82.1	78.3	80.1	<0.001**
ANN	85.6	83.2	84.4	<0.001**
SVM	87.3	85.1	86.2	<0.001**
CNN-Net	91.2	89.7	90.4	-
CNN-SVM	93.5	92.1	92.8	0.003*

CNN-SVM showed a 2.4% improvement in F1 score compared to CNN Net, with a p-value of 0.003 (significance level $\alpha=0.05$), indicating that the improvement is statistically significant. Traditional classifiers (HMM/ANN/SVM) have significantly lower performance than CNN models due to the lack of spatial feature extraction capability.

For voice, visual interaction and other interaction modes, combined with driver style research, database construction, algorithm tool development, product evaluation and other research can be carried out. Based on the relevant database, it provides support for upgrading the evaluation function and forms a complete intelligent driving data evaluation system; Conduct research and development of driving behavior data tool chain products based on the reasoning model in the process of human-computer collaborative control, and develop interactive development, testing, evaluation tools and data processing of intelligent driving. Improve the level of recognition, interaction and monitoring algorithms, and improve the efficiency of interactive applications. Optimization strategy of human-computer interaction function human-computer interaction opportunity and driving in the process of human-computer cooperative driving switching significantly affect the driver's reaction time and takeover performance. Deeply analyze and understand the intelligent control system of autonomous vehicles and the driving logic of drivers, and analyze the constraints of vehicle dynamics on intervention criteria. Explore the switching mechanism under emergency conditions, and analyze the conflict and interaction mechanism. So as to find an optimized early warning opportunity and its model, optimize the interaction modes such as vision, hearing and touch, and the sending mode of takeover request, and solve the problem of control right switching in the functions of autonomous vehicle. The improvement in segmentation accuracy for small target objects such as lights, signs, and cyclists is only 1% -0.9%, significantly lower than that for large targets such as roads and buildings (an improvement of 3.2%). In complex traffic scenarios, small target recognition errors may lead to 15% of false triggering warnings

5 Discussion

The experimental results of the vehicle localization and recognition system based on CNN-SVM proposed in this study on the CityScapes dataset show that its segmentation performance is significantly better than current mainstream segmentation algorithm models (such as FCN, DeepLabv3+, etc.) in large target categories such as roads, sidewalks, and buildings. Specifically, this model achieved an average improvement of 3.2% in MIoU metrics for these categories, validating the effectiveness of deep neural networks in extracting layer by layer features for semantic segmentation of complex scenes. However, for small target objects such as lights, signs, and traffic riders, the improvement of this model (1% -0.9%) is lower than that of some SOTA models based on attention mechanisms (such as Mask R-CNN, which improves by about 2.8%). Compared to traditional HCI (human-computer interaction) models such as thresholding and template matching, the core advantages of CNN-SVM model are reflected in the following three aspects:

Traditional HCI methods rely on manually designed geometric features such as edge gradients and color histograms, which are difficult to adapt to complex scenes such as lighting changes and occlusions. CNN automatically learns multi-scale features (such as texture, shape, and contextual relationships) through convolutional kernels. For example, in vehicle localization tasks, deep feature maps can simultaneously capture the semantic association between car window contours and license plate areas. The accuracy of the HCI model decreased by about 30% in cross dataset testing (such as transferring from daytime scenes to nighttime scenes), while CNN-SVM can control the generalization error within 8% through transfer learning (such as pre training based on ImageNet). The SVM classifier can seamlessly integrate the visual features extracted by CNN with other sensor data (such as LiDAR point clouds), while the HCI method requires a redesign of the feature extraction process.

The challenges of real-world deployment include the following technical bottlenecks for the implementation of this model in embedded systems for vehicles. The inference delay of the original model on NVIDIA Jetson AGX Xavier is 42 ms, which cannot meet the 20 ms real-time requirement of L3 level autonomous driving. When the training sample size exceeds 100000, the error rate of the model on the CityScapes validation set fluctuates by 1.2%, indicating that deep networks are sensitive to data distribution. Apply L2 penalty term to the feature channels of small target categories to reduce feature conflicts between categories; Prioritize annotating samples with model prediction confidence below 0.7 to improve data utilization efficiency. In the scenario of human-machine collaborative driving, the recognition results of this model can be integrated with driver behavior data (such as eye movement trajectories and EEG signals) to construct a multimodal interactive evaluation system. By analyzing the changes in the pupil diameter of the driver after the takeover request was sent, the visual warning triggering time was dynamically adjusted from 2.1 seconds to 1.8 seconds, resulting in a 15% increase in the success rate of takeover; Integrating piezoelectric sensors into the steering wheel, the vibration intensity is adjusted in real-time based on the collision risk level (0-5) output by the model, reducing the driver's reaction time by 0.3 seconds.

6 Conclusion

The biggest feature of multimedia information collection technology is that non-intuitive multi-source data information can be intuitively expressed and transmitted to users, and information transmission and communication can be realized by stimulating users' sensory awareness. With the wide application of information technology, intelligent system and network technology in the field of vehicles and the development of on-board information technology, the interior space, man-machine interface, operation and interaction process of automobiles are undergoing revolutionary changes, and the control interface is undergoing essential changes. In this study, the design model of automobile human-computer interaction based on multimedia information collection technology in complex driving environment is constructed. In the driving situation, the user's judgment of "safety" mainly depends on whether the secondary task occupies the cognitive channel and the interactive channel of the main driving task. By avoiding the two, the user can be given a sense of security and the interactive experience can be improved. A CNN-SVM emotional perception model is proposed, and the input is the spectrogram of audio. Window scanning and pyramid image sampling are performed on the whole image to solve the problem of different sizes and positions of vehicles to be located and identified in the image. In this paper, the algorithm improves the performance of small objects, such as light, sign, veg and rider by 1%, 0.5% and 0.9% respectively. Through learning rate domain adaptive method, the performance of generator and discriminator in countermeasure network is further optimized. The study has certain limitations and needs further optimization. This paper lacks the analysis of the architecture of the driving simulation system for human-computer interaction verification. Further analysis is needed in the future.

Funding

This work was supported by Anhui Quality Engineering Project "Electrical Engineering and Intelligent Control New professional Quality Improvement Project" (project number: 2022xjzlts012).

Conflict of interest

The authors have no relevant financial or non-financial interests to disclose.

Data availability statement

The data used to support the findings of this study are all in the manuscript.

References

[1] AlZu'bi S, Hawashin B, Mujahed M, (2019), An efficient employment of internet of multimedia things in smart and future agriculture. Multimedia Tools and Applications, 78(20): 29581-29605.

[2] Dhokrat, J G, Pulgam, N. (2024). A framework for privacy-preserving multiparty computation with homomorphic encryption and zero-knowledge proofs. Informatica, 48(21).

[3] Karadere, G, Düzcan, Y, Yldz, A R. (2020). Light-weight design of automobile suspension components using topology and shape optimization techniques. Materials Testing, 62(5), 454-458.

[4] Krishan, R, Verma, A, Mishra, S. (2019). Design of multi-machine power system stabilizers with forecast uncertainties in load/generation. IETE Journal of Research, 65(1), 44-57.

[5] Gao, D, Wang, J, Chai, R. (2024). Intelligent car autonomous driving tracking technology based on fuzzy information and multi-sensor fusion. Informatica, 48(21), 37-50.

[6] Yang, C, Yue, Y, Zhang, J, Wen, M, Wang, D. (2019). Probabilistic reasoning for unique role recognition based on the fusion of semantic-interaction and spatio-temporal features. IEEE Transactions on Multimedia, 21(5), 1195-1208.

[7] Ji, M, Peng, G, He, J, Liu, S, Chen, Z, Li, S. (2021). A two-stage, intelligent bearing-fault-diagnosis method using order-tracking and a one-dimensional convolutional neural network with variable speeds. Sensors, 21(3), 675.

[8] Chen, J, Jia, X. (2020). An approach for assembly process case discovery using multimedia information source. Computers in Industry, 115(1), 103176.

[9] Fang, Y, Wang, Y. (2024). Cross modal sentiment analysis of image text fusion based on Bi LSTM and B-CNN. Informatica, 48(21), 95-111.

[10] Huang, J, Yan, W, Li, T. H, Liu, S, Li, G. (2020). Learning the global descriptor for 3d object recognition based on multiple views decomposition. IEEE Transactions on Multimedia, 2020(99), 1-1.

[11] Wang, J, Jiang, C, Zhu, H, Yong, R, Hanzo, L. (2018). Internet of vehicles: sensing-aided transportation information collection and diffusion. IEEE Transactions on Vehicular Technology, 67(5), 3813-3825.

[12] Xin, X, Liu, X, Li, K, Xiao, B, Qi, H. (2017). Minimal perfect hashing-based information collection protocol for rfid systems. IEEE Transactions on Mobile Computing, 2017(10), 1-1.

[13] Shiraishi, M, Ashiya, H, Konno, A, Morita, K, Kataoka, S. (2019). Development of real-time collection, integration, and sharing technology for infrastructure damage information. Journal of Disaster Research, 14(2), 333-347.

[14] Ferro, N, Kim, Y, Sanderson, M. (2019). Using collection shards to study retrieval performance effect sizes. ACM Transactions on Information Systems, 37(3), 1-40.

[15] Mori, E, Kelkar, S. (2020). Introduction to the special issue on interface architects: the evolution of human–computer interaction. IEEE Annals of the History of Computing, 42(4), 6-7.

[16] Souza, A, Filho, M R, Soares, C. (2020). Production and evaluation of an educational process for human-computer interaction (hci) courses. IEEE Transactions on Education, 2020(99), 1-8.

[17] Zhou, Z Q, Sun, L. (2019). Metamorphic testing of driverless cars. Communications of the ACM, 62(3), 61-67.

[18] Gao, H, Liu, C H, Wang, W. (2018). Hybrid vehicular crowdsourcing with driverless cars: challenges and a solution. Computer, 51(12), 24-31.

[19] Guo, M, Yu, Z, Xu, Y, Huang, Y, Li, C, Javier García-Haro. (2021). Me-net: a deep convolutional neural network for extracting mangrove using sentinel-2a data. Remote Sensing, 13(7), 1-24.

[20] Yan, S, Jing, L, Wang, H. (2021). A new individual tree species recognition method based on a convolutional neural network and high-spatial resolution remote sensing imagery. Remote Sensing, 13(3), 479.

[21] Al-Hashimy, H N H, Hussein, W N., Al Jubair, A S, Yao, J. (2024). Enhancing data integrity in computerized accounting information systems using supervised and unsupervised machine learning algorithms implement A SEM-PLS analysis. Informatica, 48(20), 107-118.