



(12) Invention Patent
Application



(10) Application Publication No. CN
111860253 A
(43) Application Publication Date
10/30/2020

(21) Application No. 202010662803.9

(22) Application Date 07/10/2020

(71) Applicant Dongguan Zhengyang Electronic Machinery Co., Ltd.
Address (No. 1, Donghuan 3rd Street)
Jitigang Village, Huangjiang Town,
Dongguan, Guangdong 523750

(72) Inventor Gu Yixin

(74) Patent Agency Beijing Beyond Attorneys at Law
11332
Agent Meng Jinzhe

(51) Int.Cl.
G06K 9/00 (01/2006)
G06K 9/62 (01/2006)

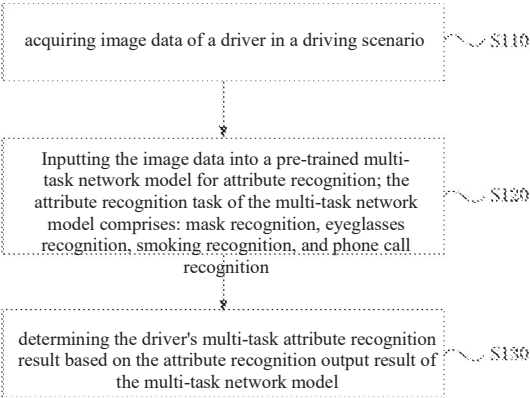
Claims 2 pages Specifications 10 pages Drawings 4 pages

(54) Name of Invention

A multi-task attribute recognition method, apparatus, medium, and device for driving scenarios

(57) Summary

The embodiments of the present application disclose a multi-task attribute recognition method, apparatus, medium, and device for driving scenarios. The method comprises: acquiring image data of a driver in a driving scenario; inputting the image data into a pre-trained multi-task network model for attribute recognition; the attribute recognition task of the multi-task network model comprises: mask recognition, eyeglasses recognition, smoking recognition, and phone call recognition; determining the driver's multi-task attribute recognition result based on the attribute recognition output result of the multi-task network model. By implementing the present technical solution, a multi-task network model can be used to simultaneously recognize multiple features of the driver and output an output result that can comprehensively reflect the driver's driving state, thereby improving the accuracy of driver state recognition.



CN 111860253 A

1. A multi-task attribute recognition method for driving scenarios, wherein the method comprises:
 - acquiring image data of a driver in a driving scenario;
 - inputting the image data into a pre-trained multi-task network model for attribute recognition; the attribute recognition task of the multi-task network model comprises: mask recognition, eyeglasses recognition, smoking recognition, and phone call recognition;
 - determining the driver's multi-task attribute recognition result based on the attribute recognition output result of the multi-task network model.
2. The method according to Claim 1, wherein inputting the image data into a trained multi-task network model for attribute recognition comprises:
 - preprocessing the said image data;
 - inputting the preprocessed image data into a pre-trained multi-task network model for attribute recognition and obtaining attribute recognition results of the multi-task network model for the target tasks of whether a mask is worn, whether eyeglasses are worn, whether there is smoking, and whether a phone call is being made;
 - caching the attribute recognition results of the target task;
 - after caching the attribute recognition results for a specific time period, judging whether the attribute of the target task exists based on the cached attribute recognition results according to the judgment ratio; if so, determining the target task attribute of the sample image.
3. The method according to Claim 2, wherein preprocessing the image data comprises:
 - acquiring a sample image and determining the encoded features of the face in the sample image;
 - determining the driver's face frame using a face recognition model based on the said encoded features;
 - adjusting according to the driver's face frame, then cropping the driver's face to obtain a cropped image, and finally normalizing it.
4. The method according to Claim 1, wherein training the said multi-task network model comprises:
 - inputting the preprocessed image data into the multi-task main network to obtain a feature map, using an OSEM module to extract the feature maps for different attribute recognition tasks, and inputting them into respective attribute recognition modules, wherein the attribute recognition modules are a mask wearing attribute recognition module, an eyeglasses wearing attribute recognition module, a smoking attribute recognition module, and a phone call attribute recognition module;
 - the output of the entire network is a feature map of each task and an attribute recognition result of each task, and the output results of the attribute recognition of each task are whether a mask is worn, whether eyeglasses are worn, whether there is smoking, and whether a phone call is being made;
 - comparing the attribute recognition results to the labeling results, reducing the difference through the corresponding loss function, and inputting the feature maps of each task into the corresponding loss function to optimize network parameters;
 - adjusting the weights of the loss function to optimize the training results of the multi-task network model and reduce the training time.
5. The method according to Claim 4, wherein before training the multi-task network model, the method further comprises:
 - labeling the sample image data with multi-task attributes;
 - determining an initial architecture of the multi-task network model;
 - determining a loss function.
6. The method according to Claim 4, wherein the number of output categories of the attribute recognition tasks of whether a mask is worn, whether there is smoking, and whether a phone call is being made is two categories;
 - The output categories of the attribute recognition task of whether eyeglasses are worn comprise: three categories, wearing IR-blocking glasses, wearing ordinary eyeglasses, and not wearing eyeglasses.
7. The method according to Claim 4, wherein for the attribute recognition module of whether a mask is worn and whether eyeglasses are worn, each module is used for attribute recognition to output the results of whether a mask is worn and whether eyeglasses are worn.

8. The method according to Claim 4, wherein for the attribute recognition modules of whether there is smoking and whether a phone call is being made, each recognition module comprises a subject recognition module and an attention module;

the subject recognition module is used for attribute recognition to output the result of whether there is smoking and whether a phone call is being made;

the attention module comprises an attention proposal sub-network and an attention network, which are used for attribute recognition and optimization of the attribute recognition results of the main network.

9. The method according to Claim 8, wherein the attention proposal sub-network mainly comprises two fully connected layers in series, which are used to output the center point coordinates of the attention area and the edge lengths of the attention area to perform cropping and enlargement processing on the original input image.

10. The method according to Claim 8, wherein the attention network performs network analysis on the cropped and enlarged image and outputs a new result of whether there is smoking or whether a phone call is being made. It can also crop and enlarge the image again as needed, perform network analysis, and output the recognition result.

11. The method according to Claim 10, wherein it is determined to calculate the classification loss and Lrank loss using the fine-grained features generated by the attention network to constrain the attribute analysis results of the main network and the attention network.

12. The method according to Claim 4, wherein after training the multi-task network model, the method further comprises:

selecting the network architecture and model with the best attribute analysis results based on test accuracy;

deleting the network that does not participate in the final attribute recognition output and its parameters to reduce the network model size and forward transmission time.

13. A multi-task attribute recognition apparatus for driving scenarios, wherein the apparatus comprises:

an image data acquisition module, used to acquire image data of a driver from a driving scenario video;

a recognition module, used to input the image data into a pre-trained multi-task network model for attribute recognition; the attribute recognition task of the multi-task network model comprises: mask recognition, eyeglasses recognition, smoking recognition, and phone call recognition;

a multi-task attribute recognition result determination module, used to determine the multi-task attribute recognition result of the driver based on the attribute recognition output result of the multi-task network model.

14. A computer-readable storage medium having stored thereon a computer program, wherein the program, when executed by a processor, implements the multi-task attribute recognition method for driving scenarios according to any of Claims 1-12.

15. A device comprising a memory, a processor, and a computer program stored on the memory and executable by the processor, wherein the processor implements the multi-task attribute recognition method for driving scenarios according to any of Claims 1-12 when executing the computer program.

A multi-task attribute recognition method, apparatus, medium, and device for driving scenarios

Technical Field

[0001] The embodiments of the present application relate to the field of computer technology, specifically to image recognition technology, and more particularly to a method, apparatus, medium, and device for multi-task attribute recognition for driving scenarios.

Background Art

[0002] In ADAS systems, the accurate and timely provision of driver feature attributes is a key factor in achieving the intelligence of safe driving systems. A common approach is to use different method models to recognize different driver attributes.

[0003] With regard to the driver attribute recognition method, deep learning-based methods are generally used to recognize certain attribute features of the driver. The traditional method is to combine multiple models to recognize various attribute features of the driver. However, first of all, the traditional method has a single scenario and struggles to deal with in-cabin scenarios with occlusion, poor lighting, multiple people, missing important attributes, etc. Secondly, if multiple models are used to recognize driver attributes, the system's response speed will slow down, model size will increase, hardware requirements will increase, and in the case of multiple people, it is impossible to ensure that all attribute features are from the same person. Finally, for the recognition of smoking and phone calls, most current methods use anchor-based neural network models, which not only cannot be integrated with other attributes into a single model, but also increase computational and labeling difficulty.

[0004] Therefore, it is necessary to design a method that encompasses all important driver attribute features while also having good real-time performance and accuracy.

Summary of the Invention

[0005] The embodiments of the present application provide a multi-task attribute recognition method, apparatus, medium, and device for driving scenarios, which can adopt a multi-task network model to simultaneously recognize multiple features of the driver and output an output result that can comprehensively reflect the driver's driving state, thereby improving the accuracy of driver state recognition.

[0006] In a first aspect, the embodiments of the present application provide a multi-task attribute recognition method for driving scenarios, the method comprising:

[0007] Acquiring image data of a driver in a driving scenario camera video;

[0008] Inputting the image data into a pre-trained multi-task network model for attribute recognition; the attribute recognition task of the multi-task network model comprises: mask recognition, eyeglasses recognition, smoking recognition, and phone call recognition;

[0009] Determining the driver's multi-task attribute recognition result based on the attribute recognition output result of the multi-task network model.

[0010] Further, inputting the image data into a trained multi-task network model for attribute recognition comprises:

[0011] Preprocessing the image data;

[0012] Inputting the preprocessed image data into a pre-trained multi-task network model for attribute recognition and obtaining attribute recognition results of the multi-task network model for the target tasks of whether a mask is worn, whether eyeglasses are worn, whether there is smoking, and whether a phone call is being made;

[0013] Caching the attribute recognition results of the target task;

[0014] After caching the attribute recognition results for a specific time period, judging whether the attribute of the target task exists based on the cached attribute recognition results according to the judgment ratio; if so, determining the target task attribute of the sample image.

- [0015] Further, preprocessing the image data comprises:
- [0016] Acquiring a sample image and determining the encoded features of the face in the sample image;
- [0017] Determining the driver's face frame using a face recognition model based on the encoded features;
- [0018] Adjusting according to the driver's face frame, then cropping the driver's face to obtain a cropped image, and finally normalizing it.
- [0019] Further, training the multi-task network model comprises:
- [0020] Inputting the preprocessed image data into the multi-task main network to obtain a feature map, using an OSEM module to extract the feature maps for different attribute recognition tasks, and inputting them into respective attribute recognition modules, wherein the attribute recognition modules are a mask wearing attribute recognition module, an eyeglasses wearing attribute recognition module, a smoking attribute recognition module, and a phone call attribute recognition module;
- [0021] The output of the entire network is a feature map of each task and an attribute recognition result of each task, and the output results of the attribute recognition of each task are whether a mask is worn, whether eyeglasses are worn, whether there is smoking, and whether a phone call is being made;
- [0022] Comparing the attribute recognition results to the labeling results, reducing the difference through the corresponding loss function, and inputting the feature maps of each task into the corresponding loss function to optimize network parameters;
- [0023] Adjusting the weights of the loss function to optimize the training results of the multi-task network model and reduce the training time.
- [0024] Further, before training the multi-task network model, the method further comprises:
- [0025] Labeling the sample image data with multi-task attributes;
- [0026] Determining an initial architecture of the multi-task network model;
- [0027] Determining a loss function.
- [0028] Further, the number of output categories of the attribute recognition tasks of whether a mask is worn, whether there is smoking, and whether a phone call is being made is two categories;
- [0029] The output categories of the attribute recognition task of whether eyeglasses are worn comprise: three categories, wearing IR-blocking glasses, wearing ordinary eyeglasses, and not wearing eyeglasses.
- [0030] Further, for the attribute recognition module of whether a mask is worn and whether eyeglasses are worn, each module is used for attribute recognition to output the results of whether a mask is worn and whether eyeglasses are worn.
- [0031] Further, for the attribute recognition modules of whether there is smoking and whether a phone call is being made, each recognition module comprises a subject recognition module and an attention module;
- [0032] The subject recognition module is used for attribute recognition to output the result of whether there is smoking and whether a phone call is being made;
- [0033] The attention module comprises an attention proposal sub-network and an attention network, which are used for attribute recognition and optimization of the attribute recognition results of the main network;
- [0034] Further, the attention proposal sub-network mainly comprises two fully connected layers in series, which are used to output the center point coordinates of the attention area and the edge lengths of the attention area to perform cropping and enlargement processing on the original input image.
- [0035] Further, the attention network performs network analysis on the cropped and enlarged image and outputs a new result of whether there is smoking or whether a phone call is being made. It can also crop and enlarge the image again as needed, perform network analysis, and output the recognition result.
- [0036] Further, it is determined to calculate the classification loss and Lrank loss using the fine-grained features generated by the attention network to constrain the attribute analysis results of the main network and the attention network.
- [0037] Further, after training the multi-task network model, the method further comprises:

- [0038] Selecting the network architecture and model with the best attribute analysis results based on test accuracy;
- [0039] Deleting the network that does not participate in the final attribute recognition output and its parameters to reduce the network model size and forward transmission time.
- [0040] In a second aspect, the embodiments of the present application provide a multi-task attribute recognition apparatus for driving scenarios, the apparatus comprising:
- [0041] An image data acquisition module, used to acquire image data of a driver in a driving scenario;
- [0042] A recognition module, used to input the image data into a pre-trained multi-task network model for attribute recognition; the attribute recognition task of the multi-task network model comprises: mask recognition, eyeglasses recognition, smoking recognition, and phone call recognition;
- [0043] A multi-task attribute recognition result determination module, used to determine the multi-task attribute recognition result of the driver based on the attribute recognition output result of the multi-task network model.
- [0044] In a third aspect, the embodiments of the present application provide a computer-readable storage medium having stored thereon a computer program, the program, when executed by a processor, implementing the multi-task attribute recognition method for driving scenarios as described in the embodiments of the present application.
- [0045] In a fourth aspect, the embodiments of the present application provide a device comprising a memory, a processor, and a computer program stored on the memory and executable by the processor, the processor implementing the multi-task attribute recognition method for driving scenarios as described in the embodiments of the present application when executing the computer program.
- [0046] The technical solution provided by the embodiments of the present application acquires image data of a driver in a driving scenario; inputs the image data into a pre-trained multi-task network model for attribute recognition; the attribute recognition task of the multi-task network model comprises: mask recognition, eyeglasses recognition, smoking recognition, and phone call recognition; determining the driver's multi-task attribute recognition result based on the attribute recognition output result of the said multi-task network model. By using the technical solution provided by the embodiments of the present application, a multi-task network model can be used to simultaneously recognize multiple features of the driver and output an output result that can comprehensively reflect the driver's driving state, thereby improving the accuracy of driver state recognition.

Description of Drawings

- [0047] FIG. 1 is a flow chart of a multi-task attribute recognition method for driving scenarios provided by an embodiment of the present application;
- [0048] FIG. 2 is a flow chart of an attribute recognition process of a multi-task network model provided by an embodiment of the present application;
- [0049] FIG. 3 is a schematic diagram of training process of a multi-task network model provided by an embodiment of the present application;
- [0050] FIG. 4 is a schematic diagram of an attention proposal sub-network provided by an embodiment of the present application;
- [0051] FIG. 5 is a schematic diagram of the structure of a multi-task attribute recognition apparatus for driving scenarios provided by an embodiment of the present application;
- [0052] FIG. 6 is a schematic diagram of the structure of a device provided by an embodiment of the present application.

Detailed Description

[0053] The present application is described in further detail below with reference to the drawings and examples. It will be understood that the specific embodiments described herein are only used to explain the present application and not to limit the present application. It should also be noted that for ease of description, only some, but not all, structures related to the present application are shown in the drawings.

[0054] Before discussing the exemplary embodiments in greater detail, it should be noted that some of the exemplary embodiments are described as processes or methods that are depicted as flow charts. While the flow charts describes the steps as a sequential process, many of the steps can be performed in parallel, concurrently, or simultaneously. In addition, the order of the steps can be rearranged. A process may be terminated when its operations are completed, but may also have additional steps not included in the drawings. The process may correspond to a method, a function, a procedure, a subroutine, a subprogram, and the like.

[0055] FIG. 1 is a flow chart of a multi-task attribute recognition method for driving scenarios provided by an embodiment of the present application. This embodiment is applicable to the case of driver state recognition. The method can be performed by a multi-task attribute recognition apparatus for driving scenarios provided by an embodiment of the present application. The apparatus can be implemented by software and/or hardware and can be integrated into devices such as smart terminals.

[0056] As shown in FIG. 1, the multi-task attribute recognition method for driving scenarios comprises:

[0057] S110, acquiring image data of a driver in a driving scenario.

[0058] The driver's image data can be acquired through an ordinary camera, an infrared camera, etc., which can be set at the rearview mirror position of the vehicle, or other positions, as long as the driver's face image can be acquired. According to the needs of actual application scenarios, an infrared 940 camera is used to build an acquisition hardware and software platform to collect video data in the vehicle cabin.

[0059] In this embodiment, it can be understood that after the image data is acquired, operations such as cleaning, storage, and labeling can be performed on the image data. Specifically, the collected data is batch extracted into image frames, and the image data is cleaned and target frames are labeled. The use of infrared cameras and ordinary cameras for data collection can not only ensure stable imaging output in complex environments and the stability of image features, which is conducive to the learning of network models, but also greatly alleviate or avoid problems such as backlighting.

[0060] In this solution, the collected RGB image can be input into a backbone network to extract the encoded features of the image; a facial recognition neural network for driving scenarios is used to analyze the encoded features. If above a set threshold, the face frame coordinate value closest to the previous result is given; if below a set threshold, it directly outputs: no driver, and ends; the face frame coordinates are proportionally enlarged, the input image is cropped, and the cropped image is normalized and resized.

[0061] S120, inputting the image data into a pre-trained multi-task network model for attribute recognition; the attribute recognition task of the multi-task network model comprises: mask recognition, eyeglasses recognition, smoking recognition, and phone call recognition.

[0062] The acquired image data may be input into a pre-trained multi-task network model. The recognition task of the multi-task network model comprises: mask recognition, eyeglasses recognition, smoking recognition, and phone call recognition. Different recognition tasks can be performed at the same time. For example, after determining the features of the image, multiple recognition tasks can be completed at the same time. If the corresponding information is recognized, such as recognizing wearing a mask, the information that the driver is wearing a mask will be displayed in the output of the mask recognition task.

[0063] It can be understood that before training the multi-task network model, it is necessary to label the driver based on whether he or she is wearing a mask or eyeglasses, smoking, or making a phone call. However, this method does not require location information and is a weakly supervised anchor-free method. Compared with traditional methods, labeling tasks are reduced. In this way, the labeling information and driver images will be used as input to train the model. During application, no labeling information is required. Only the driver's image is needed to output the judgment results of whether he or she is wearing a mask or eyeglasses, smoking, or making a phone call.

[0064] S130, determining the driver's multi-task attribute recognition result based on the attribute recognition output result of the multi-task network model.

[0065] The output result of the model can be determined according to the output result of each recognition task. After the model outputs the corresponding results, the driver's multi-task attribute recognition results are determined.

[0066] The technical solution provided by the embodiments of the present application acquires image data of a driver in a driving scenario; inputs the image data into a pre-trained multi-task network model for attribute recognition; the attribute recognition task of the multi-task network model comprises: mask recognition, eyeglasses recognition, smoking recognition, and phone call recognition; determining the driver's multi-task attribute recognition result based on the attribute recognition output result of the multi-task network model. By using the technical solution provided by the embodiments of the present application, a multi-task network model can be used to simultaneously recognize multiple features of the driver and output an output result that can comprehensively reflect the driver's driving state, thereby improving the accuracy and real-time performance of driver state recognition.

[0067] FIG. 2 is a flow chart of an attribute recognition process of a multi-task network model provided by an embodiment of the present application. As shown in FIG. 2, the attribute recognition process of the multi-task network model comprises:

[0068] S210, preprocessing the image data.

[0069] The collected and cropped RGB image is input into a backbone network to extract the encoded features of the image. Specifically, the backbone network uses the SE-ResNet 18 network, and the output scale feature map is 7*7. These 7*7 feature maps are then processed through a one-squeeze multi-excitation (OSME) module to extract 4 different types of feature maps, namely, mask, eyeglasses, smoking, and phone call feature maps, which are used to determine the features of different attribute recognition tasks.

[0070] In this technical solution, optionally, preprocessing the image data comprises:

[0071] Acquiring a sample image and determining the encoded features of the face in the sample image;

[0072] Determining the driver's face frame using a face recognition model based on the encoded features;

[0073] Adjusting according to the driver's face frame, then cropping the driver's face to obtain a cropped image, and finally normalizing it.

[0074] A facial recognition neural network for driving scenarios is used to analyze the encoded features. If above a set threshold value, the face frame coordinate value closest to the previous result is given; if below a set threshold, it directly outputs: no driver, and ends. Here, since the range of movement of the driver's face in the vehicle is limited, the face frame closest to the previous result may be determined as the face frame of the driver.

[0075] The face frame coordinates are proportionally enlarged, the input image is cropped, and the cropped image is normalized and resized.

[0076] S220, inputting the preprocessed image data into a pre-trained multi-task network model for attribute recognition and obtaining attribute recognition results of the multi-task network model for the target tasks of whether a mask is worn, whether eyeglasses are worn, whether there is smoking, and whether a phone call is being made.

[0077] In this solution, a multi-task network model can be used to recognize the cropped image to determine whether a mask is worn, whether eyeglasses or IR-blocking glasses are worn, whether there is smoking, and whether a phone call is being made.

[0078] S230, caching the attribute recognition results of the target task.

[0079] In this embodiment, optionally, before inputting the cropped image into the multi-task network model for training, the method further includes:

[0080] Labeling the sample image data with multi-task attributes;

[0081] Determining an initial architecture of the multi-task network model;

[0082] Determining a loss function.

[0083] Accordingly, the initial architecture of the multi-task network model uses the SE-ResNet18 network, and the output scale feature map is a 7*7 feature map. These 7*7 feature maps are then processed through a one-squeeze multi-excitation (OSME) module to extract 4 different types of feature maps, namely, mask, eyeglasses, smoking, and phone call feature maps, which are used to determine the features of different attribute recognition tasks.

[0084] Determining the attribute classification loss generated by the four attribute recognition tasks when performing multi-task attribute recognition; the attribute classification loss is used to compare the differences between the network output results and the labeling results during the training process and reduce these differences through forward and back propagation to the multi-task network model. These attribute classification losses are multiplied by different weights, depending on the difficulty of calculating the attribute features.

[0085] Four sets of 7*7 attention feature maps for 4 categories of different attribute recognition are input into the metric loss function together.

[0086] Specific implementation of the metric loss function: Four sets of 7*7 attention feature maps are stretched into vectors and the similarity of different images in a batch is calculated. In our task, the attention similarity S^{sa} of the same category is higher than the attention similarity S^{da} of different categories, so the metric loss function is:

$$L^{np} = \sum_{pos=1}^{pos} \log(1 + \frac{1}{n} \sum_{n=1}^n (S^{da} - S^{sa}))$$

[0087]

[0088] Where n is the number of the same category in a batch multiplied by the number of different categories. pos is the number of the same category.

[0089] The metric loss function here is mainly to make the attention feature maps of whether there is smoking, whether a phone call is being made, whether a mask is worn, and whether eyeglasses are worn mutually exclusive features. This makes different attribute features different in the attention discrimination area of the original image.

[0090] The preprocessed sample image data is matched one-to-one with the labeled multi-task attributes and arranged in order; the preprocessed sample image is input into the network, the final output is compared one-to-one with the labeling result according to the loss function, and the difference is reduced by repeated forward and backward propagation. Finally, the network converges to a level where the difference between the network output and the labeling result is negligible, and the parameters in the network are the required final results. Loss weights are adjusted according to the multi-task attribute recognition results and convergence speed to optimize the training results and speed of the multi-task network model. Specifically, this solution can use the above method to adjust the weights of the loss function to achieve the purpose of optimizing the multi-task network model.

[0091] S240, after caching the attribute recognition results for a specific time period, judging whether the attribute of the target task exists based on the cached attribute recognition results according to the judgment ratio; if so, determining the target task attribute of the sample image.

[0092] The attribute recognition results are cached for a certain period of time. The cache time can be set as needed. N consecutive frames of images are selected. If the ratio is greater than a certain set ratio, it is determined that the attribute is present. For example: Detecting mask state: The initial state is not wearing a mask. Starting from the Nth frame, the sum of the mask state cache values of the previous N frames, M, is divided by N to obtain the mask wearing ratio. If its value is greater than a certain ratio, it is concluded that the mask wearing state is present. Eyeglasses state, smoking state, and phone call state can be similarly judged.

[0093] In this embodiment, the multi-task network model attribute recognition training process comprises:

[0094] Inputting the preprocessed image data into the multi-task main network to obtain a feature map, using an OSEM module to extract the feature maps for different attribute recognition tasks, and inputting them into respective attribute recognition modules, wherein the attribute recognition modules are a mask wearing attribute recognition module, an eyeglasses wearing attribute recognition module, a smoking attribute recognition module, and a phone call attribute recognition module;

[0095] The output of the entire network is a feature map of each task and an attribute recognition result of each task, and the output results of the attribute recognition of each task are whether a mask is worn, whether eyeglasses are worn, whether there is smoking, and whether a phone call is being made;

[0096] Comparing the attribute recognition results to the labeling results, reducing the difference through the corresponding loss function, and inputting the feature maps of each task into the corresponding loss function to optimize network parameters;

[0097] Adjusting the weights of the loss function to optimize the training results of the multi-task network model and reduce the training time.

[0098] In the above technical solution, optionally, the number of output categories of the attribute recognition tasks of whether a mask is worn, whether there is smoking, and whether a phone call is being made is two categories;

[0099] The output categories of the recognition task of whether eyeglasses are worn comprise: three categories, wearing IR-blocking glasses, wearing ordinary eyeglasses, and not wearing eyeglasses.

[0100] It is well known that there are many different network structures, and the required network structure can be selected. The multi-task network structure used here is SE-Resnet 18, and the initial architecture is adjusted. For example, the input size is set to 112*112, the 7*7 convolution kernel (stride=2) of the first convolution layer is replaced with a 3*3 convolution kernel (stride=1), the first maximum pooling layer is deleted, the stride of the second convolution layer of the first resnet block stage is changed to 2, and in each residual block, a batch normalization layer and dropout are added. The structure is BN layer, Prelu, convolution, BN layer, Prelu layer, convolution, and dropout layer, and the dropout ratio of the dropout layer in the residual block is 0%-20%.

[0101] The output of these 7×7 feature maps will be passed through a one-squeeze multi-excitation (OSME) module, which mainly consists of 4 squeeze-and-excitation networks. The 4 output results of this module are then applied to the previous feature map to obtain feature maps for the four tasks. After the four feature maps are pooled, the feature vectors are input into four attribute recognition modules. The outputs of these four attribute recognition modules are whether a mask is worn, whether eyeglasses are worn, whether there is smoking and the smoking APN, and whether a phone call is being made and the phone call attention proposal sub-network (APN). Whether a mask is worn, whether eyeglasses are worn, whether there is smoking, and whether a phone call is being made are attribute recognition tasks, while the smoking APN and phone call APN will be used as inputs to the secondary attention model to crop and enlarge the image area. The attention proposal sub-network comprises two fully connected layers in series, which are used to output the center point coordinates of the attention area and the edge lengths of the attention area to perform cropping and enlargement processing on the feature map.

[0102] The enlarged image will be input into the first-level attention model to re-judge the attributes: Whether there is smoking and the smoking APN, whether a phone call is being made and the phone calling APN, and the APN generated by the first-level attention model will further crop and enlarge the image and input it into the second-level attention model to re-judge the attributes.

[0103] The first-level attention neural network model and the second-level attention neural network model here have the same network structure as the main network model and will output attribute recognition results as described above. Therefore, there will be 3 attribute recognition results for whether there is smoking and whether a phone call is being made. Finally, the model with the highest accuracy is selected as the final model based on testing (already selected during training) and a unique attribute recognition result is output.

[0104] Here, because smoking and phone calls are difficult to judge, the attention mechanism is further applied to optimize the results, but it is not limited to these two tasks. Which tasks require the use of the attention mechanism can be chosen according to your needs.

[0105] The number of outputs for tasks with the attributes of masks, IR-blocking glasses and eyeglasses, smoking, and phone calls are 2, 3, 2, and 2, respectively.

[0106] In the above technical solution, optionally, after the multi-task network model is trained, the method further comprises:

[0107] Selecting the network architecture and model with the best attribute analysis results based on test accuracy;

[0108] Deleting the network that does not participate in the final attribute recognition output and its parameters to reduce the network model size and forward transmission time.

[0109] FIG. 3 is a schematic diagram of training process of a multi-task network model provided by an embodiment of the present application. As shown in FIG. 3, a facial recognition neural network for driving scenarios is used to analyze the encoded features. If above a set threshold value, the face frame coordinate value closest to the previous result is given; if below a set threshold, it directly outputs: no driver, and ends. When the value is higher than a set threshold, the face frame closest to the previous result is cropped. After the cropped image is input into the backbone network, the mask attributes, eyeglasses attributes, smoking attributes, and phone call attributes can be recognized.

[0110] FIG. 4 is a schematic diagram of an attention proposal sub-network provided by an embodiment of the present application. As shown in FIG. 4,

[0111] Assuming the attention region of interest is a square (i.e., the output of the APN), output x_t , y_t , l_t . Where, after the softmax function, the probability p_l is output.

[0112] Where x_t , y_t , and l_t represent the abscissa and ordinate of the center point of the coordinate domain of the center point of the region and the side length of the square, respectively. The original face is cropped and enlarged according to the output of the APN sub-network to create a local area: Face scale2, as input to the new sub-network, which is the first-level attention model.

[0113] The structural design of the new sub-network is the same as that of the main network. It then outputs the classification network and the attention proposal sub-network APN, and outputs the probability p2 through the softmax function. Like the previous APN network, the APN is trained from the extracted features to obtain attention area information. This is the second-level attention model.

[0114] Highly localized regions are cropped and enlarged based on the output of the secondary attention model: Face scale3, generating face scale3 and outputting probability p3 through the softmax function. As a new input, it is input into the same sub-network as the main network, and finally the classification network is output.

[0115] Through these 3-scale sub-networks, the initial face is transformed into multi-scale fine-grained regions.

[0116] The eyeglasses and mask subtasks directly use the softmax function to represent the probability of the existence of this attribute for each element.

[0117] Loss weights are adjusted according to the attribute recognition results of the at least two tasks to complete the training of the multi-task network model.

[0118] On the basis of the above technical solution, optionally, different tasks can be selected to apply the attention mechanism according to needs, thereby optionally determining the fine-grained feature classification loss and Lrank loss generated by the attention proposal sub-network.

[0119] Fine-grained classification loss is the same as other classification losses, which can be softmax or softmax cross entropy, etc. In this case, the focal loss is used.

[0120] The Lrank loss function here is mainly used to make the classification effect of the second-level attention image better than the classification effect produced by the first-level attention image, so that the attention areas can be transformed into more recognizable attention areas step by step.

[0121] Specifically, the labeling information of driver attributes (mask state, eyeglasses state, smoking state, phone call state) is used to train the multi-task driver attribute recognition task using focal loss.

[0122] Since almost all attributes have an overall sample ratio imbalance problem, the focal loss function is used to reduce the weight of negative samples. The focal loss formula is as follows:

$$loss_{\beta} = \begin{cases} -\delta(1-y')^{\gamma} \log y', y=1 \\ -(1-\delta)y'^{\gamma} \log(1-y'), y=0 \end{cases};$$

[0123]

[0124] Where δ is the balance factor, which is used to balance the uneven ratio of positive and negative samples.

[0125] For the smoking and phone call models, in addition to the classification loss focal loss generated by attribute classification, there is also the inter scale classification loss Lrank generated by fine-grained feature classification.

$$L_{rank}(p_i^{(s)}, p_i^{(s+1)}) = \max\{1, p_i^{(s)}, p_i^{(s+1)} + margin\};$$

[0126]

[0127] $P_t^{(s)}$ represents the classification output of the t-lth layer attention model in the correct categories: Probability $P_t^{(s)}$

[0128] Through experiments, appropriate weights are set for the losses of each task. The total loss formula is as follows:

[0129] $loss = W1 * loss_{f1} + W2 * loss_{f1} + W3 * (loss_{f1} + Lrank) + W4 * (loss_{f1} + Lrank) + W5 * L^{np};$

[0130] Where W1, W2, W3, W4, and W5 are the weights of the loss of each task, which are manually adjusted through experiments to make the losses of each task on the same order of magnitude.

[0131] The multi-task network model provided by the present solution can achieve the purpose of simultaneously performing multi-task attribute recognition on the driver's state and uses multi-granularity to recognize whether the driver is smoking or making a phone call, thereby improving the accuracy of model training.

[0132] FIG. 5 is a schematic diagram of the structure of a multi-task attribute recognition apparatus for driving scenarios provided by an embodiment of the present application. As shown in FIG. 5, the multi-task attribute recognition apparatus for driving scenarios comprises:

[0133] An image data acquisition module 510, used to acquire image data of a driver in a driving scenario;

[0134] A recognition module 520, used to input the image data into a pre-trained multi-task network model for attribute recognition; the attribute recognition task of the multi-task network model comprises: mask recognition, eyeglasses recognition, smoking recognition, and phone call recognition;

[0135] A multi-task attribute recognition result determination module 530, used to determine the multi-task attribute recognition result of the driver based on the attribute recognition output result of the multi-task network model.

[0136] The above product can perform the method provided by the embodiments of the present application and has the corresponding functional modules and beneficial effects of the execution method.

[0137] The embodiments of the present application further provide a storage medium containing computer-executable instructions. When executed by a computer processor, the computer-executable instructions are used to perform a multi-task attribute recognition method for driving scenarios. The method comprises:

[0138] Acquiring image data of a driver in a driving scenario;

[0139] Inputting the image data into a pre-trained multi-task network model for attribute recognition; the attribute recognition task of the multi-task network model comprises: mask recognition, eyeglasses recognition, smoking recognition, and phone call recognition;

[0140] Determining the driver's multi-task attribute recognition result based on the attribute recognition output result of the multi-task network model.

[0141] Storage medium - any of various types of memory devices or storage devices. The term "storage medium" is intended to include: mounted media, such as CD-ROMs, floppy disks, or tape devices; computer system memory or random access memory, such as DRAM, DDR, RAM, SRAM, EDO RAM, Rambus RAM, etc.; non-volatile memory, such as flash memory or magnetic media (e.g., hard disk or optical storage); registers or other similar types of memory elements, etc. The storage medium may also include other types of memory or combinations thereof. Additionally, the storage medium may be located in the computer system in which the program is executed, or may be located in a different second computer system that is connected to the computer system via a network, such as the Internet. The second computer system may provide program instructions to the computer for execution. The term "storage medium" may include two or more storage media that may reside in different locations (such as in different computer systems connected by a network). The storage medium may store program instructions (e.g., embodied as a computer program) executable by one or more processors.

[0142] Of course, the storage medium containing computer-executable instructions provided in an embodiment of the present application is not limited to the multi-task attribute recognition operation for driving scenarios as described above, and can also execute related operations in the multi-task attribute recognition method for driving scenarios provided in any embodiment of the present application.

[0143] The embodiments of the present application provide a device that can integrate the multi-task attribute recognition device for driving scenarios provided by the embodiments of the present application. FIG. 6 is a schematic diagram of the structure of a device provided by an embodiment of the present application. As shown in FIG. 6, the present embodiment provides a device 600 comprising: one or more processors 620; a storage apparatus 610 for storing one or more programs. When the one or more programs are executed by the one or more processors 620, the one or more processors 620 perform the multi-task attribute recognition method for driving scenarios provided by the embodiments of the present application. The method comprises:

[0144] Acquiring image data of a driver in a driving scenario;

[0145] Inputting the image data into a pre-trained multi-task network model for attribute recognition; the attribute recognition task of the multi-task network model comprises: mask recognition, eyeglasses recognition, smoking recognition, and phone call recognition;

[0146] Determining the driver's multi-task attribute recognition result based on the attribute recognition output result of the multi-task network model.

[0147] Of course, those skilled in the art can understand that the processor 620 also implements the technical solution of the multi-task attribute recognition method for driving scenarios provided by any embodiment of the present application.

[0148] The device 600 shown in FIG. 6 is only an example and should not limit the functions and scope of use of the embodiments of the present application.

[0149] As shown in FIG. 6, the apparatus 600 comprises a processor 620, a storage apparatus 610, an input apparatus 630, and an output apparatus 640; the number of processors 620 in the device may be one or more, with one processor 620 shown in FIG. 6 as an example; the processor 620, the storage apparatus 610, the input apparatus 630, and the output apparatus 640 in the device may be connected via a bus or other means, and FIG. 6 shows connection via a bus 650 as an example.

[0150] The storage apparatus 610, as a computer-readable storage medium, can be used to store software programs, computer-executable programs, and module units, such as program instructions corresponding to the multi-task attribute recognition method for driving scenarios in the embodiments of the present application.

[0151] The storage apparatus 610 may primarily comprise a program storage area and a data storage area, wherein the program storage area may store an operating system and application programs required for at least one function; the data storage area may store data created according to the use of the terminal and so on. Furthermore, the storage apparatus 610 may comprise a high-speed random access memory and may also comprise a non-volatile memory, such as at least one disk memory device, flash memory device, or other non-volatile solid-state memory device. In some examples, the storage apparatus 610 may further comprise a memory remotely disposed relative to the processor 620, and the remote memory may be connected via a network. Examples of such networks include, but are not limited to, the Internet, intranets, local area networks, mobile communications networks, and combinations thereof.

[0152] The input apparatus 630 may be used to receive input numbers, character information, or voice information and to generate key signal inputs related to user settings and function controls of the device. The output apparatus 640 may comprise a display screen, a speaker, and other devices.

[0153] The device provided by the embodiments of the present application can use a multi-task network model to simultaneously recognize multiple features of the driver and output an output result that can comprehensively reflect the driver's driving state, thereby improving the accuracy of driver state recognition.

[0154] The multi-task attribute recognition apparatus, storage medium, and device for driving scenarios provided in the above embodiments can perform the multi-task attribute recognition method for driving scenarios provided in any embodiment of the present application and have the corresponding functional modules and beneficial effects of performing the method. For technical details not fully described in the above embodiments, please refer to the multi-task attribute recognition method for driving scenarios provided by any embodiment of the present application.

[0155] Note that the above are only preferred embodiments and technical principles used in this application. Those skilled in the art will understand that the present application is not limited to the specific embodiments described herein, and various obvious changes, readjustments, and substitutions can be made by those skilled in the art without departing from the scope of protection of the present application. Therefore, although the present application is described in more detail through the above embodiments, the present application is not only limited to the above embodiments, but may also include other equivalent embodiments without departing from the present application concept, and the scope of the present application is determined by the scope of the appended claims.

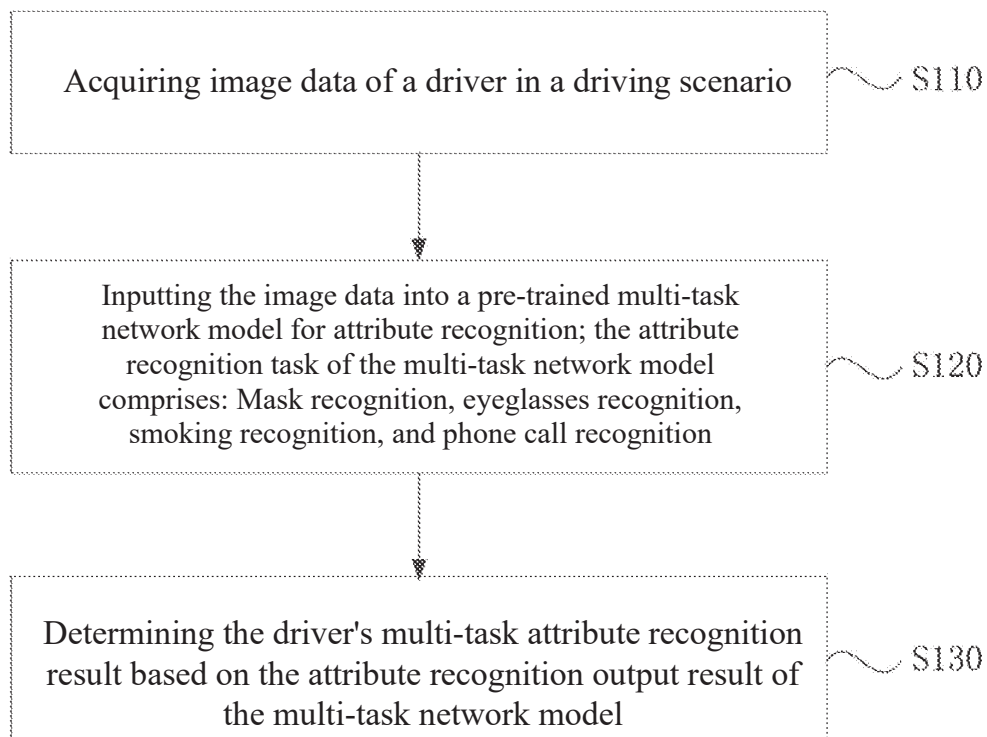


FIG. 1

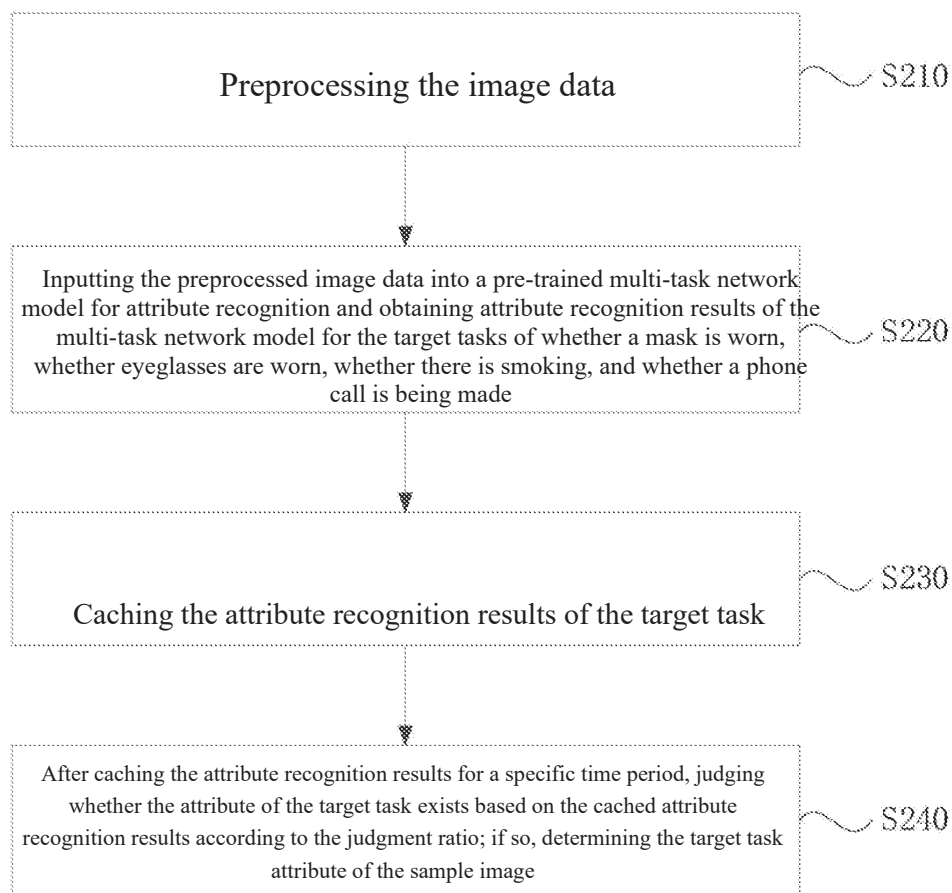


FIG. 2

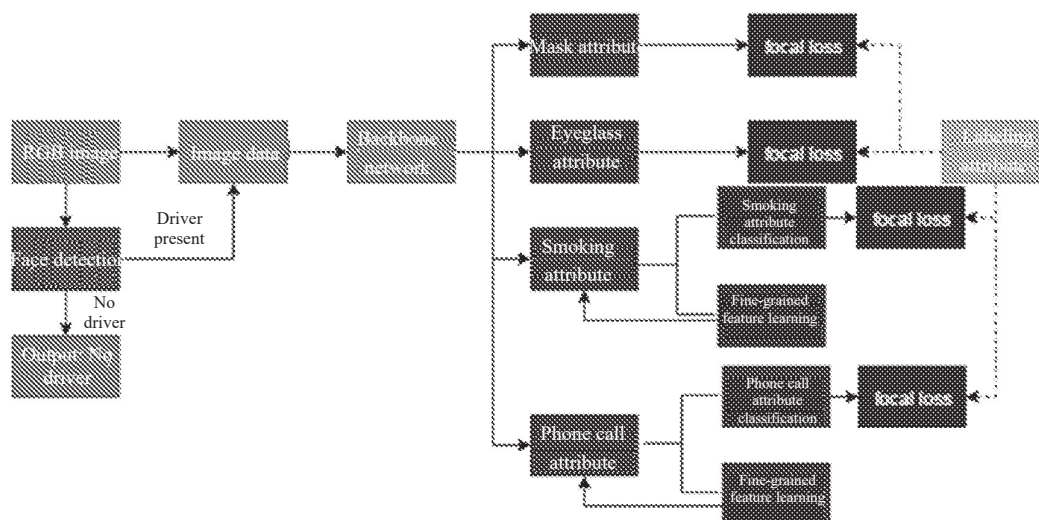


FIG. 3

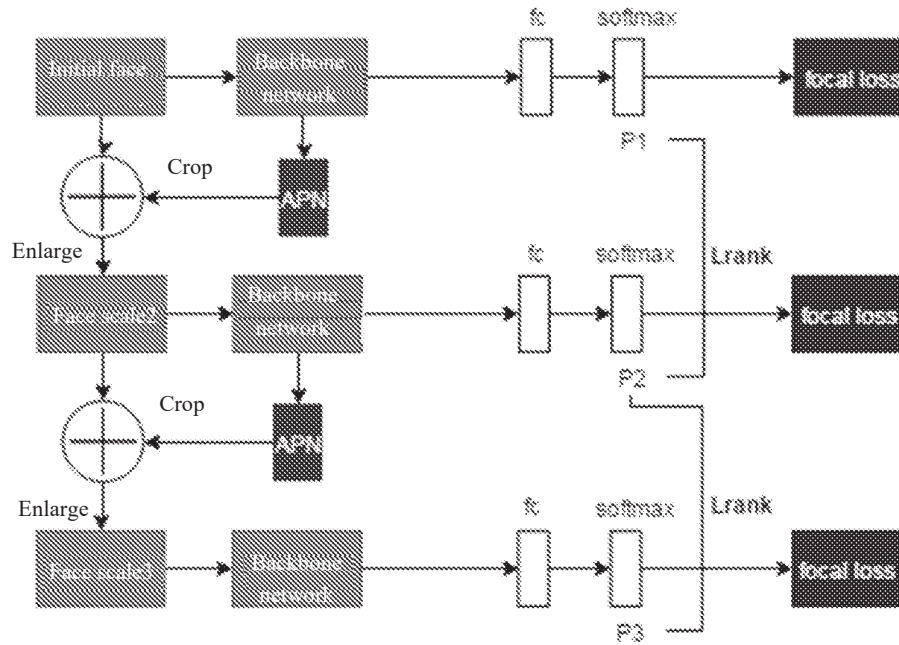


FIG. 4

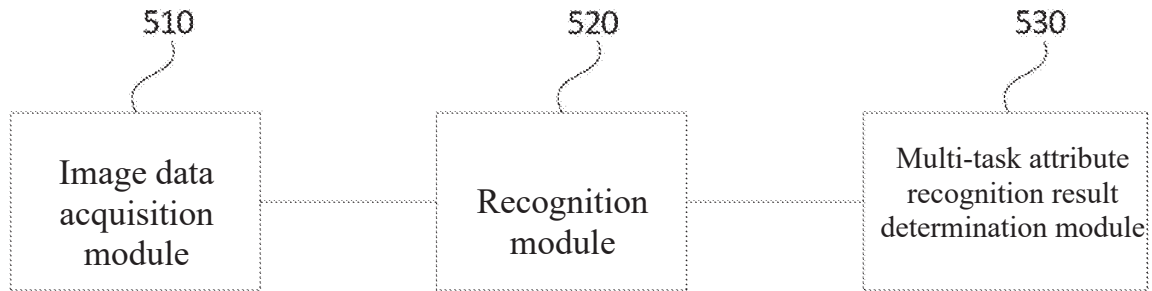


FIG. 5

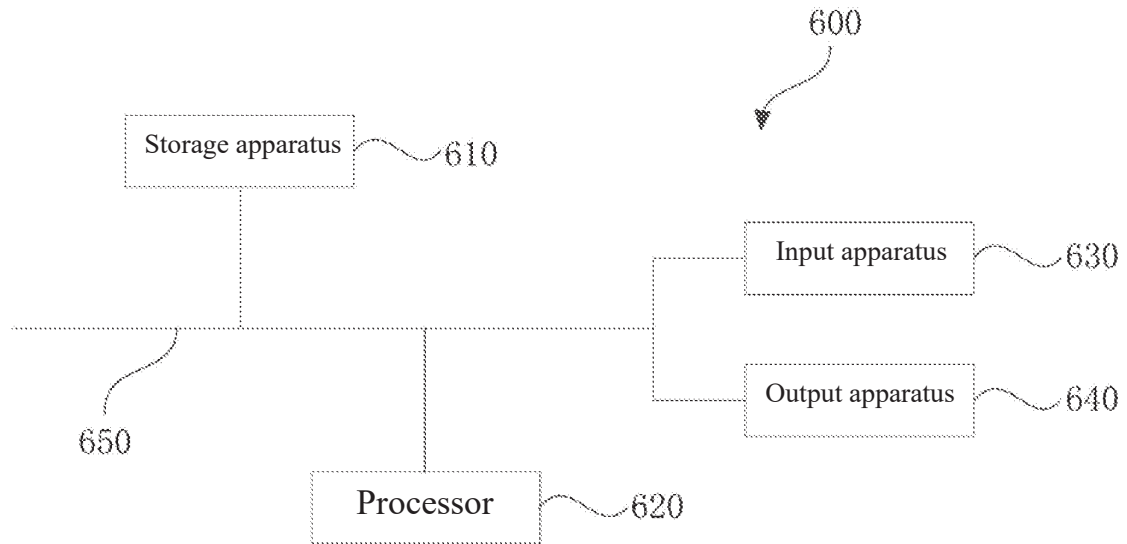


FIG. 6



(21) 申请号 202010662803.9

(22) 申请日 2020.07.10

(71) 申请人 东莞正扬电子机械有限公司

地址 523750 广东省东莞市黄江镇鸡啼岗
村(东环三街1号)

(72) 发明人 顾一新

(74) 专利代理机构 北京品源专利代理有限公司

11332

代理人 孟金喆

(51) Int.Cl.

G06K 9/00 (2006.01)

G06K 9/62 (2006.01)

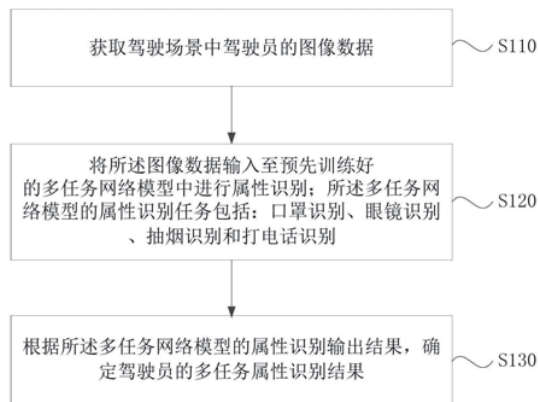
权利要求书2页 说明书10页 附图4页

(54) 发明名称

一种驾驶场景的多任务属性识别方法、装置、介质及设备

(57) 摘要

本申请实施例公开了一种驾驶场景的多任务属性识别方法、装置、介质及设备。该方法包括：获取驾驶场景中驾驶员的图像数据；将所述图像数据输入至预先训练好的多任务网络模型中进行属性识别；所述多任务网络模型的属性识别任务包括：口罩识别、眼镜识别、抽烟识别和打电话识别；根据所述多任务网络模型的属性识别输出结果，确定驾驶员的多任务属性识别结果。通过执行本技术方案，可以采用多任务网络模型，同时对驾驶员的多种特征进行识别，并输出能够整体反应驾驶员驾驶状态的输出结果，从而提高驾驶员状态识别准确度的目的。



1. 一种驾驶场景的多任务属性识别方法,其特征在于,该方法包括:

获取驾驶场景中驾驶员的图像数据;

将所述图像数据输入至预先训练好的多任务网络模型中进行属性识别;所述多任务网络模型的属性识别任务包括:口罩识别、眼镜识别、抽烟识别和打电话识别;

根据所述多任务网络模型的属性识别输出结果,确定驾驶员的多任务属性识别结果。

2. 根据权利要求1所述的方法,其特征在于,将所述图像数据输入至训练好的多任务网络模型中进行属性识别,包括:

对所述图像数据进行预处理;

将预处理后的图像数据输入至预先训练好的多任务网络模型中进行属性识别,得到多任务网络模型对是否戴口罩,是否戴眼镜,是否抽烟以及是否打电话的目标任务的属性识别结果;

缓存对目标任务的属性识别结果;

在缓存了特定时长的属性识别结果后,根据判定比例对缓存的属性识别结果判断目标任务的属性是否存在;若是,则确定样本图像的目标任务属性。

3. 根据权利要求2所述的方法,其特征在于,对所述图像数据进行预处理,包括:

获取样本图像,确定样本图像中人脸的编码特征;

根据所述编码特征利用人脸识别模型确定驾驶人员的人脸框;

根据所述驾驶人员的人脸框,进行调整,再对驾驶人员的人脸裁剪,得到裁剪图像,最后归一化处理。

4. 根据权利要求1所述的方法,其特征在于,所述多任务网络模型的训练,包括:

将预处理后的图像数据输入至多任务主网络得到特征图,采用OSEM模块分别提取针对不同属性识别任务的特征图,分别输入至各自的属性识别模块中,所述属性识别模块分别为是否戴口罩属性识别模块、是否戴眼镜属性识别模块、是否抽烟属性识别模块、是否打电话属性识别模块;

整个网络的输出为每个任务的特征图、每个任务的属性识别结果,各个任务的属性识别的输出结果分别为是否戴口罩,是否戴眼镜,是否抽烟,是否打电话;

将属性识别结果与标注结果相比较,通过相应的损失函数减少其差异,并将各个任务的特征图输入到相应的损失函数中,优化网络参数;

调整损失函数的权重,以优化对多任务网络模型的训练结果和减少训练时间。

5. 根据权利要求4所述的方法,其特征在于,在多任务网络模型的训练之前,所述方法还包括:

对样本图像数据标注多任务属性;

确定所述多任务网络模型的初始架构;

确定损失函数。

6. 根据权利要求4所述的方法,其特征在于,是否戴口罩,是否抽烟以及是否打电话的属性识别任务的输出类别数均为两种类别;

是否戴眼镜的属性识别任务的输出类别数包括:佩戴红外阻断眼镜、佩戴普通眼镜以及不佩戴眼镜三种类别。

7. 根据权利要求4所述的方法,其特征在于,对于是否戴口罩及是否戴眼镜属性识别模

块,每个模块分别用于属性识别输出是否戴口罩,是否戴眼镜的结果。

8.根据权利要求4所述的方法,其特征在于,对于是否抽烟以及是否打电话属性识别模块,每个识别模块包括主体识别模块和注意力模块;

所述主体识别模块用于属性识别输出是否抽烟、是否打电话的结果;

所述注意力模块包括注意力建议子网络与注意力网络,用于属性识别及优化主体网络的属性识别结果。

9.根据权利要求8所述的方法,其特征在于,所述注意力建议子网络主要包括两个串联的全连接层,用于输出注意力区域的中心点坐标,以及注意力区域的边长,以执行对原始输入图像的裁剪和放大处理。

10.根据权利要求8所述的方法,其特征在于,所述注意力网络将裁剪和放大处理的图像经过网络分析,输出新的是否抽烟、是否打电话的结果,也可以根据需要再次裁剪和放大图像,进行网络分析,输出识别结果。

11.根据权利要求10所述的方法,其特征在于,确定利用所述注意力网络产生的细粒度特征计算分类损失、Lrank损失,以约束主体网络和注意力网络的属性分析结果。

12.根据权利要求4所述的方法,其特征在于,在多任务网络模型训练之后,所述方法还包括:

根据测试准确度选择属性分析结果最优的网络架构及其模型;

删掉不参与最后属性识别输出的网络及其参数,减小网络模型大小和前传时间。

13.一种驾驶场景的多任务属性识别装置,其特征在于,该装置包括:

图像数据获取模块,用于从驾驶场景视频中获取驾驶员的图像数据;

识别模块,用于将所述图像数据输入至预先训练好的的多任务网络模型中进行属性识别;所述多任务网络模型的属性识别任务包括:口罩识别、眼镜识别、抽烟识别和打电话识别;

多任务属性识别结果确定模块,用于根据所述多任务网络模型的属性识别输出结果,确定驾驶员的多任务属性识别结果。

14.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,该程序被处理器执行时实现如权利要求1-12中任一项所述的驾驶场景的多任务属性识别方法。

15.一种设备,包括存储器,处理器及存储在存储器上并可在处理器运行的计算机程序,其特征在于,所述处理器执行所述计算机程序时实现如权利要求1-12中任一项所述的驾驶场景的多任务属性识别方法。

一种驾驶场景的多任务属性识别方法、装置、介质及设备

技术领域

[0001] 本申请实施例涉及计算机技术领域,具体涉及图像识别技术,尤其涉及一种驾驶场景的多任务属性识别方法、装置、介质及设备。

背景技术

[0002] 在ADAS系统中,准确及时的提供驾驶员特征属性是实现安全驾驶系统智能化的关键因素。常用方法是用不同的方法模型分别识别不同的驾驶员属性。

[0003] 针对驾驶员属性识别方法,一般采用基于深度学习的方法来识别驾驶员某属性特征,传统方法是结合多个模型识别驾驶员的各个属性特征,但是首先传统方法的场景单一,难以应对存在遮挡、光线不良、多人、重要属性缺漏等的驾驶室内场景。其次若是采用多个模型的方式识别驾驶员属性,系统的反应速度也会变慢,模型大小亦会增大,硬件要求随之增高,并且在多人情况下,无法确保所有属性特征出自同一人。最后针对抽烟和打电话属性的识别,现在方法多数使用anchor-based神经网络模型,这样不仅无法和其他属性融合为一个模型,而且增加了计算难度和标注难度。

[0004] 因此,需要设计出一种囊括所有驾驶员重要属性特征的同时,也能有较好的实时性和准确性的方法。

发明内容

[0005] 本申请实施例提供一种驾驶场景的多任务属性识别方法、装置、介质及设备,可以采用多任务网络模型,同时对驾驶员的多种特征进行识别,并输出能够整体反应驾驶员驾驶状态的输出结果,从而提高驾驶员状态识别准确度的目的。

[0006] 第一方面,本申请实施例提供了一种驾驶场景的多任务属性识别方法,该方法包括:

[0007] 获取驾驶场景摄像头视频中驾驶员的图像数据;

[0008] 将所述图像数据输入至预先训练好的多任务网络模型中进行属性识别;所述多任务网络模型的属性识别任务包括:口罩识别、眼镜识别、抽烟识别和打电话识别;

[0009] 根据所述多任务网络模型的属性识别输出结果,确定驾驶员的多任务属性识别结果。

[0010] 进一步的,将所述图像数据输入至训练好的多任务网络模型中进行属性识别,包括:

[0011] 对所述图像数据进行预处理;

[0012] 将预处理后的图像数据输入至预先训练好的多任务网络模型中进行属性识别,得到多任务网络模型对是否戴口罩,是否戴眼镜,是否抽烟以及是否打电话的目标任务的属性识别结果;

[0013] 缓存对目标任务的属性识别结果;

[0014] 在缓存了特定时长的属性识别结果后,根据判定比例对缓存的属性识别结果判断

目标任务的属性是否存在;若是,则确定样本图像的目标任务属性。

[0015] 进一步的,对所述图像数据进行预处理,包括:

[0016] 获取样本图像,确定样本图像中人脸的编码特征;

[0017] 根据所述编码特征利用人脸识别模型确定驾驶人员的人脸框;

[0018] 根据所述驾驶人员的人脸框,进行调整,再对驾驶人员的人脸裁剪,得到裁剪图像,最后归一化处理。

[0019] 进一步的,所述多任务网络模型的训练,包括:

[0020] 将预处理后的图像数据输入至多任务主网络得到特征图,采用OSEM模块分别提取针对不同属性识别任务的特征图,分别输入至各自的属性识别模块中,所述属性识别模块分别为是否戴口罩属性识别模块、是否戴眼镜属性识别模块、是否抽烟属性识别模块、是否打电话属性识别模块;

[0021] 整个网络的输出为每个任务的特征图、每个任务的属性识别结果,各个任务的属性识别的输出结果分别为是否戴口罩,是否戴眼镜,是否抽烟,是否打电话;

[0022] 将属性识别结果与标注结果相比较,通过相应的损失函数减少其差异,并将各个任务的特征图输入到相应的损失函数中,优化网络参数;

[0023] 调整损失函数的权重,以优化对多任务网络模型的训练结果和减少训练时间。

[0024] 进一步的,在多任务网络模型的训练之前,所述方法还包括:

[0025] 对样本图像数据标注多任务属性;

[0026] 确定所述多任务网络模型的初始架构;

[0027] 确定损失函数。

[0028] 进一步的,是否戴口罩,是否抽烟以及是否打电话的属性识别任务的输出类别数均为两种类别;

[0029] 是否戴眼镜的属性识别任务的输出类别数包括:佩戴红外阻断眼镜、佩戴普通眼镜以及不佩戴眼镜三种类别。

[0030] 进一步的,对于是否戴口罩及是否戴眼镜属性识别模块,每个模块分别用于属性识别输出是否戴口罩,是否戴眼镜的结果。

[0031] 进一步的,对于是否抽烟以及是否打电话属性识别模块,每个识别模块包括主体识别模块和注意力模块;

[0032] 所述主体识别模块用于属性识别输出是否抽烟、是否打电话的结果;

[0033] 所述注意力模块包括注意力建议子网络与注意力网络,用于属性识别及优化主体网络的属性识别结果;

[0034] 进一步的,所述注意力建议子网络主要包括两个串联的全连接层,用于输出注意力区域的中心点坐标,以及注意力区域的边长,以执行对原始输入图像的裁剪和放大处理。

[0035] 进一步的,所述注意力网络将裁剪和放大处理的图像经过网络分析,输出新的是否抽烟、是否打电话的结果,也可以根据需要再次裁剪和放大图像,进行网络分析,输出识别结果。

[0036] 进一步的,确定利用所述注意力网络产生的细粒度特征计算分类损失、Lrank损失,以约束主体网络和注意力网络的属性分析结果。

[0037] 进一步的,在多任务网络模型训练之后,所述方法还包括:

- [0038] 根据测试准确度选择属性分析结果最优的网络架构及其模型；
- [0039] 删掉不参与最后属性识别输出的网络及其参数，减小网络模型大小和前传时间。
- [0040] 第二方面，本申请实施例提供了一种驾驶场景的多任务属性识别装置，该装置包括：
- [0041] 图像数据获取模块，用于获取驾驶场景中驾驶员的图像数据；
- [0042] 识别模块，用于将所述图像数据输入至预先训练好的多任务网络模型中进行属性识别；所述多任务网络模型的属性识别任务包括：口罩识别、眼镜识别、抽烟识别和打电话识别；
- [0043] 多任务属性识别结果确定模块，用于根据所述多任务网络模型的属性识别输出结果，确定驾驶员的多任务属性识别结果。
- [0044] 第三方面，本申请实施例提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如本申请实施例所述的驾驶场景的多任务属性识别方法。
- [0045] 第四方面，本申请实施例提供了一种设备，包括存储器，处理器及存储在存储器上并可在处理器运行的计算机程序，所述处理器执行所述计算机程序时实现如本申请实施例所述的驾驶场景的多任务属性识别方法。
- [0046] 本申请实施例所提供的技术方案，获取驾驶场景中驾驶员的图像数据；将所述图像数据输入至预先训练的多任务网络模型中进行属性识别；所述多任务网络模型的属性识别任务包括：口罩识别、眼镜识别、抽烟识别和打电话识别；根据所述多任务网络模型的属性识别输出结果，确定驾驶员的多任务属性识别结果。通过采用本申请所提供的技术方案，可以采用多任务网络模型，同时对驾驶员的多种特征进行识别，并输出能够整体反应驾驶员驾驶状态的输出结果，从而提高驾驶员状态识别准确度的目的。

附图说明

- [0047] 图1是本申请实施例提供的驾驶场景的多任务属性识别方法的流程图；
- [0048] 图2是本申请实施例提供的多任务网络模型的属性识别过程的流程图；
- [0049] 图3是本申请实施例提供的多任务网络模型的训练过程的示意图；
- [0050] 图4是本申请实施例提供的注意力建议子网络的示意图；
- [0051] 图5是本申请实施例提供的驾驶场景的多任务属性识别装置的结构示意图；
- [0052] 图6是本申请实施例提供的一种设备的结构示意图。

具体实施方式

[0053] 下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释本申请，而非对本申请的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与本申请相关的部分而非全部结构。

[0054] 在更加详细地讨论示例性实施例之前应当提到的是，一些示例性实施例被描述成作为流程图描绘的处理或方法。虽然流程图将各步骤描述成顺序的处理，但是其中的许多步骤可以被并行地、并发地或者同时实施。此外，各步骤的顺序可以被重新安排。当其操作完成时所述处理可以被终止，但是还可以具有未包括在附图中的附加步骤。所述处理可以对应于方法、函数、规程、子例程、子程序等等。

[0055] 图1是本申请实施例提供的驾驶场景的多任务属性识别方法的流程图,本实施例可适用于驾驶员状态识别的情况,该方法可以由本申请实施例所提供的驾驶场景的多任务属性识别装置执行,该装置可以由软件和/或硬件的方式来实现,并可集成于智能终端等设备中。

[0056] 如图1所示,所述驾驶场景的多任务属性识别方法包括:

[0057] S110、获取驾驶场景中驾驶员的图像数据。

[0058] 其中,驾驶员的图像数据获取,可以通过普通摄像头,红外摄像头等,并且可以设置在车辆的后视镜位置,或者其他位置,只要能够获取到驾驶员的人脸图像即可。根据实际应用场景的需要,采用红外940摄像头,搭建采集硬软件平台,在车内驾驶室进行视频数据采集。

[0059] 在本实施例中,可以理解的,在获取到图像数据之后,可以对图像数据进行清理、存储以及标注等操作。具体的,对采集的数据进行图像帧的批量提取,并对图片数据进行清理和目标框的标注。采用红外摄像头和普通摄像头进行数据的采集既可以确保在复杂环境下稳定的成像输出,保证图像特征的稳定性,从而有利于网络模型的学习,也可以很大程度上缓解或避免逆光等问题。

[0060] 本方案中,可以将采集的RGB图像输入主干网络提取图像的编码特征;采用驾驶场景人脸识别神经网络,对编码特征进行分析,若高于设定阈值,给出最接近上一次结果的人脸框坐标值;若低于设定阈值,则直接输出:无驾驶员,并结束;按照人脸框坐标值按比例进行适当扩大,对输入图像进行裁剪,并将裁剪后的图像归一化、统一尺寸。

[0061] S120、将所述图像数据输入至预先训练好的多任务网络模型中进行属性识别;所述多任务网络模型的属性识别任务包括:口罩识别、眼镜识别、抽烟识别和打电话识别。

[0062] 其中,可以将获取到的图像数据输入至预先训练好的多任务网络模型。所述多任务网络模型的识别任务包括:口罩识别、眼镜识别、抽烟识别和打电话识别。其中,不同的识别任务可以同时进行,例如确定图像的特征之后,可以同时完成多个识别任务中,若识别到相应的信息,例如识别到带有口罩,则在口罩识别任务的输出显示出驾驶员带有口罩的信息。

[0063] 可以理解的,在多任务网络模型训练之前,需要根据驾驶员实际是否带有口罩,眼镜,有无抽烟、打电话的动作进行标注,但该方法不需要标注位置信息,是一种弱监督 anchor-free的方法。和传统方法相比,减轻了标注任务。这样标注信息和驾驶员图像将被作为输入,训练模型。在应用过程中不需要标注信息,只需要驾驶员图像即可输出是否带有口罩,眼镜,有无抽烟、打电话的动作判断结果。

[0064] S130、根据所述多任务网络模型的属性识别输出结果,确定驾驶员的多任务属性识别结果。

[0065] 其中,可以根据各个识别任务的输出结果,确定模型的输出结果。并在模型输出相应的结果之后,确定驾驶员的多任务属性识别结果。

[0066] 本申请实施例所提供的技术方案,获取驾驶场景中驾驶员的图像数据;将所述图像数据输入至预先训练好的多任务网络模型中进行属性识别;所述多任务网络模型的属性识别任务包括:口罩识别、眼镜识别、抽烟识别和打电话识别;根据所述多任务网络模型的属性识别输出结果,确定驾驶员的多任务属性识别结果。通过采用本申请所提供的技术方

案,可以采用多任务网络模型,同时对驾驶员的多种特征进行识别,并输出能够整体反应驾驶员驾驶状态的输出结果,从而提高驾驶员状态识别准确度和实时性的目的。

[0067] 图2是本申请实施例提供的多任务网络模型的属性识别过程的流程图,如图2所示,所述多任务网络模型的属性识别过程包括:

[0068] S210、对所述图像数据进行预处理。

[0069] 将采集并剪切处理过的RGB图像输入主干网络提取图像的编码特征。具体的,主干网络采用SE-ResNet18网络,输出尺度的特征图为7*7的特征图。这些7*7的特征图,再经过one squeeze multi-excitation (OSME) 模块,提取到4类不同的特征图,即口罩、眼镜、抽烟和打电话特征图,分别用于判断不同的属性识别任务的特征。

[0070] 在本技术方案中,可选的,对所述图像数据进行预处理,包括:

[0071] 获取样本图像,确定样本图像中人脸的编码特征;

[0072] 根据所述编码特征利用人脸识别模型确定驾驶人员的人脸框;

[0073] 根据所述驾驶人员的人脸框,进行调整,再对驾驶人员的人脸裁剪,得到裁剪图像,最后归一化处理。

[0074] 其中,采用驾驶场景人脸识别神经网络,对编码特征进行分析,若高于设定阈值,给出最接近上一次结果的人脸框坐标值;若低于设定阈值,则直接输出:无驾驶员,并结束。此处,由于驾驶员的人脸在车辆中的移动范围是有限的,因此,可以与最接近上一次结果的人脸框确定为驾驶员的人脸框。

[0075] 将人脸框坐标值按比例进行适当扩大,对输入图像进行裁剪,并将裁剪后的图像归一化、统一尺寸。

[0076] S220、将预处理后的图像数据输入至预先训练好的多任务网络模型中进行属性识别,得到多任务网络模型对是否戴口罩,是否戴眼镜,是否抽烟以及是否打电话的目标任务的属性识别结果。

[0077] 在本方案中,可以利用多任务网络模型对裁剪图像进行识别,确定是否戴口罩,是否戴眼镜或红外阻断眼镜,是否抽烟,是否打电话。

[0078] S230、缓存对目标任务的属性识别结果。

[0079] 本实施例中,可选的,在将裁剪图像输入至多任务网络模型进行训练之前,所述方法还包括:

[0080] 对样本图像数据标注多任务属性;

[0081] 确定所述多任务网络模型的初始架构;

[0082] 确定损失函数。

[0083] 相应的,多任务网络模型的初始架构采用SE-ResNet18网络,输出尺度的特征图为7*7的特征图。这些7*7的特征图,再经过one squeeze multi-excitation (OSME) 模块,提取到4类不同的特征图,即口罩、眼镜、抽烟和打电话特征图,分别用于判断不同的属性识别任务的特征。

[0084] 确定所述四个属性识别任务在执行多任务属性识别时所产生的属性分类损失;所述属性分类损失用于在训练过程中比较网络输出结果与标注结果的差异,并通过前传和反向传播到多任务网络模型中,减小这些差异。而这些属性分类损失将分别乘以不同的权重,这些权重会根据属性特征计算的难易程度而定。

[0085] 其中,4类不同属性识别的4组7*7的注意力特征图将一起输入到度量损失函数中。

[0086] 度量损失的具体实施方式:将4组7*7的注意力特征图拉长为向量,计算一个batch里不同图像的相似度,我们任务相同类别的注意力相似度 S^{sa} 要比不同类别的注意力相似度 S^{da} 高,所以度量损失函数为:

$$[0087] \quad L^{np} = \sum_{pos=1}^{pos} \log(1 + \frac{1}{n} \sum_{n=1}^n (S^{da} - S^{sa}))$$

[0088] 这里的n为一个batch里同类别的个数乘以不同类别的个数。pos为同类别的个数。

[0089] 这里的度量损失函数,主要是为了让是否抽烟、是否打电话、是否戴口罩与是否戴眼镜的注意力特征图成为互斥的特征。使不同的属性特征在原图的注意力辨别区域有所不同。

[0090] 对预处理后的样本图像数据与标注的多任务属性一一对应,按次序排列;预处理后的样本图像输入到网络中,得到的最后输出会与标注结果按照损失函数一一比较,并通过反复前传和反传减小其差异。最后收敛到网络输出与标注结果的差异可忽略的程度,网络里的各个参数即为所需的最后结果。根据多任务属性识别结果和收敛速度调整损失权重,以优化对多任务网络模型的训练结果和速度。本方案具体的,可以采用上述方式来对损失函数的权重进行调整,以达到对多任务网络模型优化的目的。

[0091] S240、在缓存了特定时长的属性识别结果后,根据判定比例对缓存的属性识别结果判断目标任务的属性是否存在;若是,则确定样本图像的目标任务属性。

[0092] 其中对缓存了一定时长的属性识别结果,该缓存时间可根据需要设定,选择连续N帧图像,大于某设定比例则判断该属性存在。例如:检测口罩状态:起始状态为不佩戴口罩,从第N帧开始,将前N帧的图像口罩状态缓存值之和M除以N为佩戴口罩比例,如若其值大于某比例,则得出佩戴口罩状态存在。同理对眼镜状态、抽烟状态、打电话状态进行判断。

[0093] 在本实施例中,所述多任务网络模型属性识别训练过程,包括:

[0094] 将预处理后的图像数据输入至多任务主网络得到特征图,采用OSEM模块分别提取针对不同属性识别任务的特征图,分别输入至各自的属性识别模块中,所述属性识别模块分别为是否戴口罩属性识别模块、是否戴眼镜属性识别模块、是否抽烟属性识别模块、是否打电话属性识别模块;

[0095] 整个网络的输出为每个任务的特征图、每个任务的属性识别结果,各个任务的属性识别的输出结果分别为是否戴口罩,是否戴眼镜,是否抽烟,是否打电话;

[0096] 将属性识别结果与标注结果相比较,通过相应的损失函数减少其差异,并将各个任务的特征图输入到相应的损失函数中,优化网络参数;

[0097] 调整损失函数的权重,以优化对多任务网络模型的训练结果和减少训练时间。

[0098] 在上述技术方案中,可选的,是否戴口罩,是否抽烟以及是否打电话的属性识别任务的输出类别数均为两种类别;

[0099] 是否戴眼镜的识别任务的输出类别数包括:佩戴红外阻断眼镜、佩戴普通眼镜以及不佩戴眼镜三种类别。

[0100] 其中,总所周知,网络结构多种多样,可以选择需要的网络结构,这里用到的多任

务网络结构为SE-Resnet18,并对初始架构进行调整,例如,将输入尺寸设置为112*112,将第一个卷积层的7*7的卷积核(stride=2),替换成3*3的卷积核(stride=1),删掉第一个最大池化层,并把第一个resnet block的stage的第二个卷积层,stride变为2,,并在每一个的残差块中,增加BN层和dropout,结构为BN层、Prelu、卷积、BN层、Prelu层、卷积、dropout层,残差块中的dropout层的dropout比例为0%-20%。

[0101] 对输出的这些 7×7 的特征图将经过one squeeze multi-excitation(OSME)模块,该模块主要为4个压缩和激励网络,这个模块输出的4个结果再将结果作用到之前的特征图中,从而得到四个任务的特征图,四个特征图分别经过池化处理之后,将所述特征向量分别输入至四个属性识别模块,所述四个属性识别模块的输出分别为是否戴口罩,是否戴眼镜,是否抽烟以及抽烟APN,是否打电话以及打电话注意力建议子网络(APN:attention proposal sub-network)。是否戴口罩,是否戴眼镜,是否抽烟以及是否打电话即为属性识别任务,而抽烟APN和打电话APN会作为二级注意力模型的输入,用于对图像进行区域裁剪和放大。所述注意力建议子网络包括两个串联的全连接层,用于输出注意力区域的中心点坐标,以及注意力区域的边长,以执行对特征图的裁剪和放大处理。

[0102] 而放大后的图像将会输入到第一级注意力模型中,重新进行属性判断:是否抽烟以及抽烟APN,是否打电话以及打电话APN,而第一级注意力模型产生的APN将进一步裁剪和放大图像,并输入到第二级注意力模型中,重新进行属性判断。

[0103] 而这里的第一级注意力神经网络模型和第二级注意力神经网络模型,和主网络模型有相同的网络结构,并且会如上所述的都会输出属性识别结果。故而是是否抽烟以及是否打电话会分别有3个属性识别结果,最终根据测试选取准确率最高的模型作为最后的模型(在训练时已经选择好),输出唯一的属性识别结果。

[0104] 这里是由于抽烟和打电话难以判断,所以才进一步运用了注意力机制,优化结果,但并不固定为这两个任务。可以根据自己的需要,选择哪些任务需要运用注意力机制。

[0105] 针对口罩、红外阻断眼镜与普通眼镜、抽烟、打电话属性的任务的输出个数分别为2,3,2,2。

[0106] 在上述技术方案中,可选的,在多任务网络模型训练之后,所述方法还包括:

[0107] 根据测试准确度选择属性分析结果最优的网络架构及其模型;

[0108] 删掉不参与最后属性识别输出的网络及其参数,减小网络模型大小和前传时间。

[0109] 图3是本申请实施例提供的多任务网络模型的训练过程的示意图。如图3所示,采用驾驶场景人脸识别神经网络,对编码特征进行分析,若高于设定阈值,给出最接近上一次结果的人脸框坐标值;若低于设定阈值,则直接输出:无驾驶员,并结束。高于设定阈值时,对最接近上一次结果的人脸框进行裁剪,在图像裁剪之后输入至主干网络之后,可以对口罩属性、眼镜属性、抽烟属性以及打电话属性进行识别。

[0110] 图4是本申请实施例提供的注意力建议子网络的示意图。如图4所示,

[0111] 假设注意力感兴趣区域为正方形(即APN的输出),输出 x_t 、 y_t 、 l_t 。其中,经过softmax函数,输出概率 p_l 。

[0112] 其中, x_t 、 y_t 、 l_t 分别表示区域中心点的坐标域的中心点横坐标、纵坐标、和正方形的边长。将原始人脸根据APN子网络输出裁剪和放大,产生局部区域:人脸scale2,作为新的子网络的输入,这是第一级注意力模型。

[0113] 新的子网络的结构设计和主网络是一样的,再输出分类网络和注意力建议子网络APN,经过softmax函数输出概率p2,APN如上一APN网络一样,从提取到的特征进行训练得到注意力区域信息,这是第二级注意力模型。

[0114] 根据二级注意力模型的输出裁剪和放大的高局部区域:人脸scale3,产生人脸scale3,经过softmax函数输出概率p3。作为新的输入,输入到和主网络一样的子网络中,最后输出分类网络。

[0115] 通过这3个尺度的子网络,将初始人脸变为多尺度的细粒度区域。

[0116] 而眼镜和口罩子任务直接采用softmax函数,代表每个元素的该属性存在的概率。

[0117] 根据所述至少两个任务的属性识别结果调整损失权重,以完成对多任务网络模型的训练。

[0118] 在上述技术方案的基础上,可选的,可以根据需要选择不同的任务应用注意力机制的,从而可选的确定所述注意力建议子网络产生的细粒度特征分类损失、Lrank损失。

[0119] 细粒度分类损失与其他分类损失一样,可以是softmax或者softmax cross entropy等,这里用的是focal loss。

[0120] 这里的Lrank损失函数主要是为了使第二级注意力图像的分类效果优于第一级注意力图像产生的分类效果,使注意力关注区域一级级的变为更有辨识力的注意力区域。

[0121] 具体的,利用驾驶员属性的标注信息(口罩状态、眼镜状态、抽烟状态、打电话状态),利用focal loss训练多任务驾驶员属性识别任务。

[0122] 由于几乎所有属性多存在整体样本比例失衡问题,使用focal loss损失函数降低负样本所占权重。focal loss公式如下:

$$[0123] \quad loss_{fl} = \begin{cases} -\partial(1-y')^{\gamma} \log y', & y=1 \\ -(1-\partial)y'^{\gamma} \log(1-y'), & y=0 \end{cases};$$

[0124] 其中 ∂ 为平衡因子,用来平衡正负样本本身的比例不均。

[0125] 对于抽烟、打电话模型,除了属性分类产生的分类损失focal loss以外,还有细粒度特征分类产生的尺度间分类损失Lrank。

$$[0126] \quad L_{rank}(p_t^{(s)}, p_t^{(s+1)}) = \max\{1, p_t^{(s)}, p_t^{(s+1)} + margin\};$$

[0127] $p_t^{(s)}$ 表示第t-1层注意力模型在正确的categorys的分类输出:概率 $p_t^{(s)}$

[0128] 通过实验,为各个任务的损失设定合适的权重。总的loss的公式如下:

$$[0129] \quad loss = W1 * loss_{fl} + W2 * loss_{fl} + W3 * (loss_{fl} + L_{rank}) + W4 * (loss_{fl} + L_{rank}) + W5 * L^{np};$$

[0130] 其中,W1、W2、W3、W4、W5分别为各个任务损失的权重,通过实验手动调整,使各个任务的损失在同一个数量级上。

[0131] 本方案提供的多任务网络模型,可以实现对驾驶员的状态同时进行多任务属性识别的目的,而且采用多粒度对是否抽烟以及是否打电话的状态进行识别,提高模型训练的准确性。

[0132] 图5是本申请实施例提供的驾驶场景的多任务属性识别装置的结构示意图。如图5所示,所述驾驶场景的多任务属性识别装置包括:

[0133] 图像数据获取模块510,用于获取驾驶场景中驾驶员的图像数据;

[0134] 识别模块520,用于将所述图像数据输入至预先训练好的多任务网络模型中进行

属性识别;所述多任务网络模型的属性识别任务包括:口罩识别、眼镜识别、抽烟识别和打电话识别;

[0135] 多任务属性识别结果确定模块530,用于根据所述多任务网络模型的属性识别输出结果,确定驾驶员的多任务属性识别结果。

[0136] 上述产品可执行本申请实施例所提供的方法,具备执行方法相应的功能模块和有益效果。

[0137] 本申请实施例还提供一种包含计算机可执行指令的存储介质,所述计算机可执行指令在由计算机处理器执行时用于执行一种驾驶场景的多任务属性识别方法,该方法包括:

[0138] 获取驾驶场景中驾驶员的图像数据;

[0139] 将所述图像数据输入至预先训练的多任务网络模型中进行属性识别;所述多任务网络模型的属性识别任务包括:口罩识别、眼镜识别、抽烟识别和打电话识别;

[0140] 根据所述多任务网络模型的属性识别输出结果,确定驾驶员的多任务属性识别结果。

[0141] 存储介质——任何的各种类型的存储器设备或存储设备。术语“存储介质”旨在包括:安装介质,例如CD-ROM、软盘或磁带装置;计算机系统存储器或随机存取存储器,诸如DRAM、DDR RAM、SRAM、EDO RAM,兰巴斯(Rambus)RAM等;非易失性存储器,诸如闪存、磁介质(例如硬盘或光存储);寄存器或其它相似类型的存储器元件等。存储介质可以还包括其它类型的存储器或其组合。另外,存储介质可以位于程序在其中被执行的计算机系统中,或者可以位于不同的第二计算机系统中,第二计算机系统通过网络(诸如因特网)连接到计算机系统。第二计算机系统可以提供程序指令给计算机用于执行。术语“存储介质”可以包括可以驻留在不同位置中(例如在通过网络连接的不同计算机系统中)的两个或更多存储介质。存储介质可以存储可由一个或多个处理器执行的程序指令(例如具体实现为计算机程序)。

[0142] 当然,本申请实施例所提供的一种包含计算机可执行指令的存储介质,其计算机可执行指令不限于如上所述的驾驶场景的多任务属性识别操作,还可以执行本申请任意实施例所提供的驾驶场景的多任务属性识别方法中的相关操作。

[0143] 本申请实施例提供了一种设备,该设备中可集成本申请实施例提供的驾驶场景的多任务属性识别装置。图6是本申请实施例提供的一种设备的结构示意图。如图6所示,本实施例提供了一种设备600,其包括:一个或多个处理器620;存储装置610,用于存储一个或多个程序,当所述一个或多个程序被所述一个或多个处理器620执行,使得所述一个或多个处理器620实现本申请实施例所提供的驾驶场景的多任务属性识别方法,该方法包括:

[0144] 获取驾驶场景中驾驶员的图像数据;

[0145] 将所述图像数据输入至预先训练的多任务网络模型中进行属性识别;所述多任务网络模型的属性识别任务包括:口罩识别、眼镜识别、抽烟识别和打电话识别;

[0146] 根据所述多任务网络模型的属性识别输出结果,确定驾驶员的多任务属性识别结果。

[0147] 当然,本领域技术人员可以理解,处理器620还实现本申请任意实施例所提供的驾驶场景的多任务属性识别方法的技术方案。

[0148] 图6显示的设备600仅仅是一个示例,不应对本申请实施例的功能和使用范围带来

任何限制。

[0149] 如图6所示,该设备600包括处理器620、存储装置610、输入装置630和输出装置640;设备中处理器620的数量可以是一个或多个,图6中以一个处理器620为例;设备中的处理器620、存储装置610、输入装置630和输出装置640可以通过总线或其他方式连接,图6中以通过总线650连接为例。

[0150] 存储装置610作为一种计算机可读存储介质,可用于存储软件程序、计算机可执行程序以及模块单元,如本申请实施例中的驾驶场景的多任务属性识别方法对应的程序指令。

[0151] 存储装置610可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序;存储数据区可存储根据终端的使用所创建的数据等。此外,存储装置610可以包括高速随机存取存储器,还可以包括非易失性存储器,例如至少一个磁盘存储器件、闪存器件、或其他非易失性固态存储器件。在一些实例中,存储装置610可进一步包括相对于处理器620远程设置的存储器,这些远程存储器可以通过网络连接。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0152] 输入装置630可用于接收输入的数字、字符信息或语音信息,以及产生与设备的用户设置以及功能控制有关的键信号输入。输出装置640可包括显示屏、扬声器等设备。

[0153] 本申请实施例提供的设备,可以采用多任务网络模型,同时对驾驶员的多种特征进行识别,并输出能够整体反应驾驶员驾驶状态的输出结果,从而提高驾驶员状态识别准确度的目的。

[0154] 上述实施例中提供的驾驶场景的多任务属性识别装置、存储介质及设备可执行本申请任意实施例所提供的驾驶场景的多任务属性识别方法,具备执行该方法相应的功能模块和有益效果。未在上述实施例中详尽描述的技术细节,可参见本申请任意实施例所提供的驾驶场景的多任务属性识别方法。

[0155] 注意,上述仅为本申请的较佳实施例及所运用技术原理。本领域技术人员会理解,本申请不限于这里所述的特定实施例,对本领域技术人员来说能够进行各种明显的变化、重新调整和替代而不会脱离本申请的保护范围。因此,虽然通过以上实施例对本申请进行了较为详细的说明,但是本申请不仅仅限于以上实施例,在不脱离本申请构思的情况下,还可以包括更多其他等效实施例,而本申请的范围由所附的权利要求范围决定。

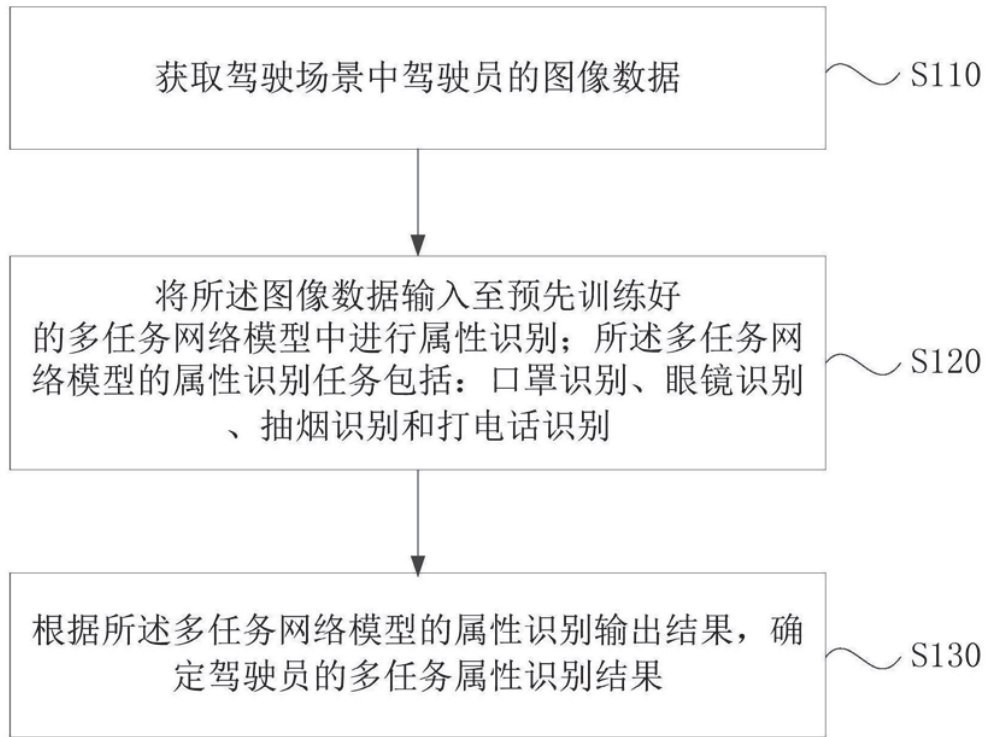


图1

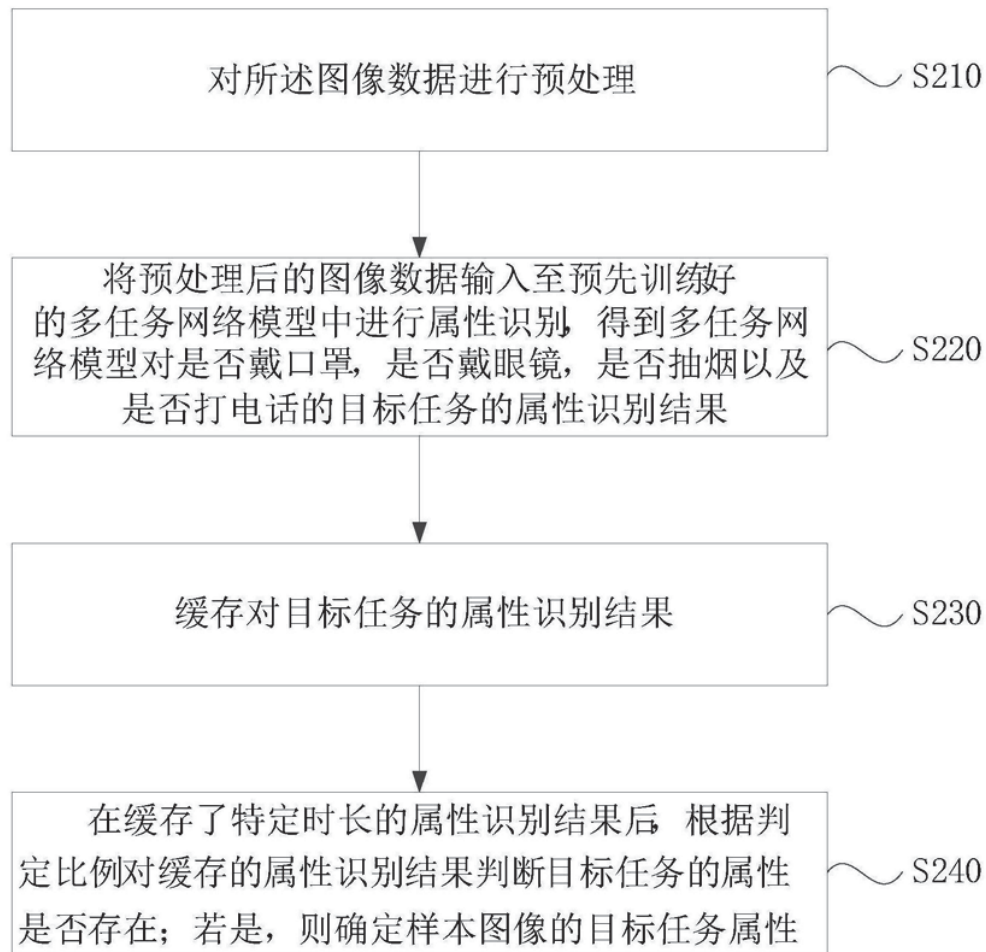


图2

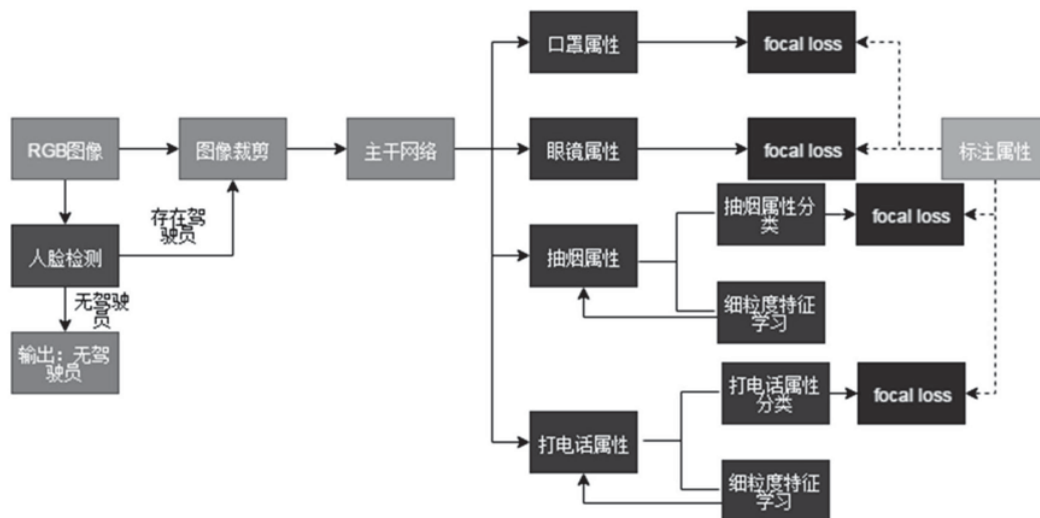


图3

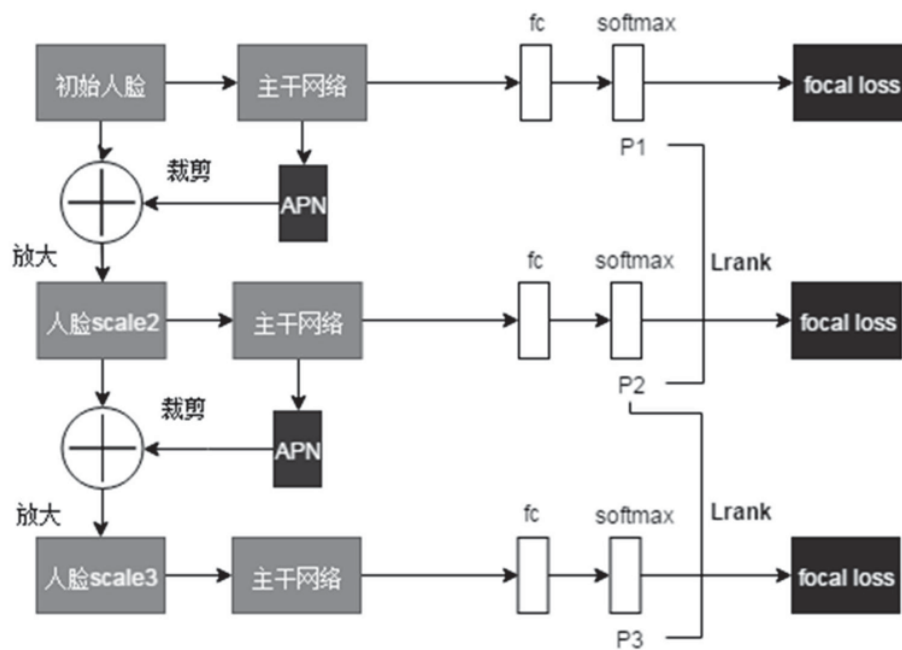


图4



图5

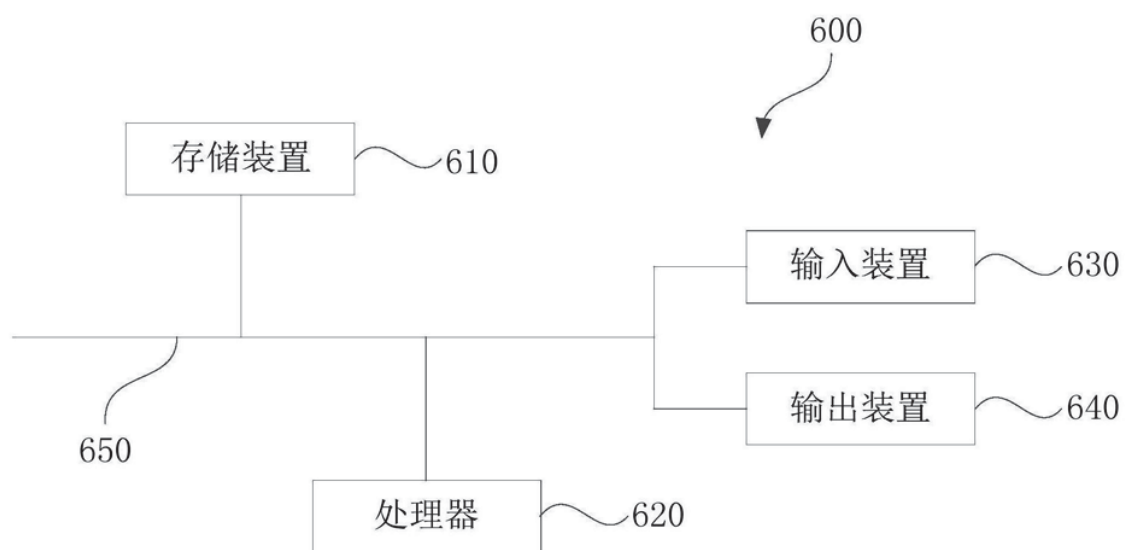


图6

Date: October 2, 2025

To whom it may concern:

I, **Jason Lesowitz**, a translator fluent in the English and Chinese languages, on behalf of TransPerfect Translations, Inc., do solemnly and sincerely declare that the following is, to the best of my knowledge and believe, a true and correct translation of the document(s) listed below in a form that reflects the intention and meaning of the original text.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

The document(s) designated as:

CN111860253A.pdf



Signature

Jason Lesowitz

Print