Patent application

Patent name: A driver behavior monitoring system using body part coordinates and AIGC

Type of report to be declared: Patent

Inventor of the patent: Ha Minh Tuan, Zhang Ai Ping

Writer and technician of the technical disclosure document: Ha Minh Tuan, Zhang Ai Ping

Phone: +84 0977536826 Fax: E-mail: minhtuanha031@gmail.com

Abstract

During driving, drivers may engage in various abnormal behaviors such as using their phone, talking to passengers, drinking, eating, and more. These behaviors divert their attention away for a period of time and can potentially lead to traffic accidents. Solutions for detecting these behaviors often involve the use of wearable devices or employing deep learning models to identify such behaviors in images. However, these methods are either inconvenient or can only detect a limited set of behaviors that the deep learning model has been trained on, whereas the behaviors that can lead to traffic accidents are extremely diverse. To solve this problem, this patent proposed a driver behavior monitoring system using body part coordinates and AIGC. The major techniques introduced in this patent are as follows. (1) Using two cameras for original data collection; (2) Training an AIGC model to generate driver images with different situations. (3) Using original images and generated images to individually design a driver front view dataset and a driver side view dataset. (4) Training a driver drowsiness detection model using the driver front view dataset. (5) Training a model for detecting a driver using a smartphone while driving using the driver side view dataset. (6) Monitoring the driver behavior in side view images using body part coordinates. With the proposed system, the behaviors of the driver are fully monitored.

1. The technical problem to be solved with this patent

1.1 Collecting a large number of images of a driver in different situations

Deep learning-based object detection techniques require large amounts of data to effectively train the model. This is especially important because diverse data helps the model better understand the characteristics of different objects and situations. Collecting images from different angles is a way to ensure diversity in the data, providing the model with many different perspectives of the same situation and helping the model develop the ability to recognize objects. under many different conditions.

To collect complete information about the driver's state while driving, it is necessary to use some cameras to capture images of a driver from different views. For example, driver front view images may contain information on some objects such as closed eyes, head tilt, or yawning mouth. On the contrary, driver front view images can show abnormal behaviors such as answering the phone or drinking water more clearly. This helps the model detect unsafe driving behaviors. However, collecting images of drivers making abnormal behavior may be difficult, as the behavior is often uncommon and not easily recorded. Furthermore, the driver may feel uncomfortable or unwilling to recreate dangerous situations. Therefore, using AIGC (Artificial Intelligence Generative Content) data generation models to generate images of drivers in new situations from a small number of original images is a reasonable solution. AIGC data generation models are capable of generating completely new and diverse data based on conditions and situations learned from the original data. They can simulate rare situations and provide good data for training models to detect objects and driver abnormal behaviors. This improves the ability to detect and evaluate the driver's body posture in a variety of situations, thereby improving driving safety and performance.

* 1. Driver drowsiness detection

Detecting driver drowsiness is important because it can help ensure traffic safety and avoid accidents. It makes drivers lose focus on driving, which can lead to slow reactions, loss of ability to evaluate traffic situations, and increased risk of accidents. Signs of a tired driver may include: eyes closed, head tilted down, and yawning. There are many techniques applied to detect driver drowsiness. (1) Analyze driving behavior, i.e. monitor driving style, or unstable steering. (2) Use physiological sensors to measure physiological signs such as heart rate, breathing rate, or skin temperature to detect drowsiness. (3) Calculate the ratio of the driver's eye and mouth areas. in driver images. The above techniques can detect driver drowsiness early, but they are easily affected by noise when the driver needs to perform operations in emergency situations or have similar expressions. The advantage of detecting driver drowsiness using deep learning models compared to conventional image analysis algorithms is the ability to learn and recognize signs of drowsiness from large datasets. The deep learning model can learn complex features and is capable of automatically learning from data, helping to improve the accuracy of drowsiness detection.

* 1. Detecting the driver using a smartphone while driving

Detecting drivers using a smartphone while driving is an important task because this behavior can cause serious traffic accidents. This behavior of the driver is usually detected by an image classification model. However, because the image of each driver in each car is different, their posture using the phone is also very diverse. Therefore, image classification models are not trained with rich enough datasets, so they cannot predict new situations well. In this project, the case of a driver using a smartphone is detected by the side view camera, because this camera can observe the driver's posture more fully compared to a front view camera. On the other hand, a smartphone detection model and a body posture estimation model are used for this task. This is to avoid wrong detection when the driver is not holding the phone when receiving a call or failed detection when the camera does not observe the phone when the driver holds it to their ear.

1.4 Detecting other abnormal behaviors

Driver abnormal behaviors in the cabin that cameras can observe include using a smartphone when driving, not looking at the road, driving with one hand, or signs of fatigue. However, the number of abnormal behaviors that can lead to traffic accidents is limitless because drivers can perform many different behaviors in the cabin, and these behaviors can vary from person to person. This makes predicting and detecting abnormal behavior difficult and complex. Common techniques to detect abnormal behavior include using cameras, on-board sensors, and deep learning. However, these methods still have disadvantages such as errors and dependence on training data. Deep learning models can only detect a finite number of abnormal behaviors because they are trained on sample data, and collecting sample data for all possible behaviors is impossible. Therefore, they may not be able to recognize some new anomalous behaviors that do not appear in the training data set. The model for estimating the driver's body posture is capable of detecting important body parts such as the driver's head and hands. To identify driver abnormal behavior, it is necessary to determine each driver's standard posture and monitor its changes over time. Each person may have their own standard posture, and it can change based on the driving situation and psychological state. By monitoring the coordinates of the driver's head and hands, the model can detect abnormal expressions based on changes from the standard posture. This helps identify distracting or dangerous behaviors and improves driving safety.

2. Detailed introduction of the technical background and description of the existing technical solutions that most closely resemble the present patent

2.1 AIGC-based data generation technique

AIGC (Artificial Intelligence Generated Content) based data generation technique is a method that uses artificial intelligence (AI) to create new data, images, or content without human intervention. This is an area of research that has been growing in recent years and has many important applications. Some prominent products and techniques in this field include ChatGPT, Stable Diffsion, StyleGAN, etc. The methods these products use to generate completely new data are often based on neural networks and machine learning algorithms. They learn from existing training data and then generate new data by automatically mapping similar graphs, patterns, or structures. The applications of AIGC imaging are very diverse. AIGC techniques can be used to create images and graphic content for websites, mobile applications, video games, and advertising. They can also be used to create simulated medical images and simulate clinical situations for training and education in the medical field. Additionally, AIGC is used to generate training and testing data for artificial intelligence models, such as chatbots and image recognition systems. Advantages of using the AIGC technique include the ability to generate large amounts of data quickly and efficiently, the ability to control the nature of the data produced, and the ability to aggregate the data for many different applications. However, it is also important to note that careful selection of algorithms and parameter tuning is required to ensure data accuracy and avoid generating erroneous information.

**2.2 Human body posture estimation model**

Human body posture estimation model is a tool or software designed to analyze, measure, and evaluate human postures and movements. The goal of this model is to understand and simulate human body postures in a variety of contexts and applications. It has a wide range of important applications in many fields, including healthcare, sports, industry, and education. In the medical field, human body posture estimation models can be used to monitor patient posture over long periods of time. This can be helpful in evaluating and managing health problems such as herniated discs or treating osteoarthritis. In industry, human body posture estimation models can help ensure working safety in dangerous environments. For example, it can be used to detect and prevent unsafe positions during work, helping to avoid accidents and injuries. Data is the most important factor in building this model, and it can be collected from many different sources such as cameras, Kinect, or posture measuring devices.

2.3 The closest available techniques to this patent

In patent CN115983352B, a method of generating data for an automated driving system is described. This method uses radiation fields and adversarial generative networks to collect data from traffic accident scenes and create a training dataset. The advantage of this technique is to improve the reliability of the automated driving system in complex situations. However, this technique has limitations in simulating situations with large differences in the position and size of objects in the image.

In patent CN110059582B, a multi-scale attention convolution neural network is used to identify driver behavior through collecting and enhancing driver image data. The advantage of this model is the combination of multi-scale and attention to improve behavior recognition. However, the model requires a large and diverse amount of data for training only recognizes pre-classified behaviors in the training set, and cannot detect other behaviors that can cause traffic accidents.

In AU2021103045A4, a driver fatigue detection technique using a Raspberry Pi single-circuit computer and computer vision to monitor blink patterns and eye movements is proposed. The advantage of this technique is the ability to detect fatigue early and help prevent traffic accidents. However, its accuracy can be affected by factors such as the driver wearing glasses or low light conditions.

3. The disadvantages of existing technological inference by causality and the purpose of the patent

For monitoring driver behavior such as drowsiness, yawning, drinking water, using the phone, etc. many techniques have been published. Specifically, driver fatigue is detected with wearable sensors, analyzing heart rate, or wheel movement. In addition, computer vision-based driver fatigue detection techniques with eye and mouth area estimation are also applied. In particular, the height and width ratio of the image area containing the eyes and mouth is calculated to establish the threshold. Besides, other abnormal behaviors are also detected using image classification models. To enrich the dataset, traditional data augmentation techniques are also applied. These methods still have some limitations as follows.

1. Low accuracy: Because the original dataset cannot cover all situations that a driver might realistically encounter in the driver's cabin, traditional data augmentation methods are often applied to enrich this dataset. These techniques often focus on changing factors such as the size and position of objects in the images or adjusting image properties such as brightness, contrast, and adding noise. However, these methods cannot generate completely new images of the driver, so they cannot fully simulate rare situations in real life. As a result, the trained model has low accuracy and cannot operate stably.
2. High cost of designing datasets: To collect a large enough amount of data, many cameras are installed on different vehicles, in different locations. Then multiple drivers need to sit in the cabin of these vehicles performing different actions that are captured by cameras. Although data collection is simple, it consumes large costs, human resources, and time.
3. Driver abnormal behavior is diverse, so deep learning models cannot learn all factual situations. Although they can identify common behaviors, situations that are rare, complex, or unlike any of the learned data samples can make it difficult for them.
4. Fatigue is hard to detect by a machine vision-based technique when the driver is wearing sunglasses or the mouth is covered with their hand. This happens because these techniques rely on recognizing facial expressions and other signs. Wearing sunglasses or covering the mouth with a hand may hide important information that the system needs to identify driver fatigue.
5. Using wearable sensors to detect driver fatigue is not only inconvenient but also increases the cost of the system. Wearing sensors on the body can cause discomfort and difficulty for the driver, affecting concentration and driving safety. Wearing and removing the sensors before and after each drive requires time and effort. This can make using the system frustrating, especially for drivers who have to drive continuously for long periods of time. Furthermore, high-quality sensors are often not cheap, and integrating them into vehicle systems requires effort and money. This can increase the value of the car or system and make it an unattractive choice for many consumers.
6. These functions, which are driver drowsiness detection and behavior monitoring, have not been integrated into a unique system. Integrating these functions into a single system can provide many benefits, including improving traffic safety and reducing the risk of accidents due to driver distraction. Continued research and development of solutions to optimize this integration is critical to enhancing the performance and availability of future smart automotive systems.

4. The detailed construction of the technical solutions of this patent

4.1 Original dataset design

The system structure proposed in this patent is illustrated in Figure 1. In it, images of the driver in different situations are collected by the Image acquisition subsystem. The image processing unit includes body posture estimation, driver drowsiness detection, steering wheel and smartphone detection models. Data and information about the system are stored in data storage. When abnormal situations are detected, the system will send a warning signal to the driver.

The original data collection process is illustrated in Figure 2. One camera is mounted in front of the driver to capture images of the driver's face and hands for drowsy detection (Step 1). Another camera is installed to capture side view images of the driver, including the head, shoulders, and arms (Step 2). After images of drivers in various situations are collected, such as using a smartphone, wearing sunglasses, eating, drinking, etc., they are labeled (Step 3) and selected for the AIGC model training set (Step 4). Figures 3 and 4 illustrate image collection for driver abnormal behavior detection and driver drowsiness detection techniques, respectively.

4.2 Training an AIGC model and generating new images of a driver in different situations using the original dataset

The process for AIGC model training and image generation using the original dataset is shown in Figure 5. After setting the initial parameters (Step 5), the AIGC model is trained with the original dataset (Step 6). It should be noted that the images in the training dataset are collected from two different views. In particular, front view images help the model learn the Normal and Drowsy states of the driver, while the features of other driver behaviors are learned from the side view images. After the AIGC model is trained (Step 7), a prompt is then fed into the AIGC model as input (Step 8) to generate images of a driver in the respective situations (Step 9).

4.3 Driver front view dataset design and driver drowsiness detection technique

Previous image processing-based methods typically extract the driver's eyes and mouth regions from the original images. Then, the state of the driver is determined based on the proportions between the height and the width of these two objects. This method is difficult when the driver wears sunglasses and covers his/her mouth with a hand when yawning. To solve this problem, this patent uses a driver drowsiness detection model as follows (Figure 6). Firstly, original driver front view images and corresponding images generated by the AIGC model are used to design a driver front view dataset for an object detection model. The labels used for this model are opened-eye, closed-eye, yawning-mouth, steering wheel-holding-hand (holdng-hand for short), mouth-covering-hand, and sunglasses (Steps 10-12). After that, the model is trained using the front view dataset (Step 13).

After training, the model is used to detect driver drowsiness as shown in Figure 7. Firstly, the objects in the driver front view image are detected (Step 14). If objects like closed-eye or yawning-mouth are detected (Step 15), or sunglasses and mouth-covering-hand are detected (Step 18), i.e. driving in a drowsy state (Steps 16, 19). The system then sends an alert signal to the driver (Step 17). Otherwise, no action is taken (Steps 20, 21).

4.4 Driver side view dataset design and the technique for detecting a driver using a smartphone while driving

In previous studies, the situation of a driver using a smartphone while driving is often detected by an image classification model. However, because the characteristics of the driver and the location of the smartphone in the images are extremely diverse, such a model may produce wrong classifications.

To overcome this problem, this patent uses the coordinates of the hands and smartphone in the image to determine the situation (Figure 8). First, the original driver side view images and the corresponding images generated by the AIGC model are used to design a side view dataset (Steps 22, 23). After that, a smartphone and the steering wheel detection model is trained using the side view dataset (Step 24). The proposed system needs to be connected to the vehicle control system to realize the vehicle's status. While the vehicle is being driven by the driver (Step 25), the input driver side view image is predicted using a body posture estimation model (Figure 9) and a smartphone detection model (Steps 26, 27). As a result, the coordinates of the smartphone and hands in the images are obtained (Step 28). If the distance between the coordinates of the smartphone and the coordinates of the hands is less than the preset value (Step 29), that means the driver is using a smartphone while driving, and the system will issue an alerting signal (Step 31). Otherwise, no action is taken (Step 30). The result of detecting a driver using a smartphone in side view images is illustrated in Figure 10.

4.5 Driver behavior monitoring technique based on body part coordinates

Similar to the situation of a driver using a smartphone while driving, the abnormal behaviors of a driver, such as looking backward, drinking water, talking to passengers, etc., are often detected by an image classification model. However, the driver's abnormal postures that can lead to accidents are so diverse that the classification model cannot learn all the features of every situation. This leads to wrong results of prediction.

To overcome that challenge, this patent uses body part coordinates (Figure 9) to classify driver behaviors (Figure 11). During driving, the driver's posture is monitored by the driver side view camera. With the coordinates of the steering wheel known, the standard posture of the driver in the images can be determined including the coordinates of the following parts, right ear, right eye, nose, shoulders, two hands holding the steering wheel, and the distances between the nose and hands are in range of the length of respective arms +/- 10%. Note that the standard posture of a driver must ensure the above characteristics, while the coordinates of the body parts may change. It can be seen that the coordinates of the body parts of drivers at their standard posture are not the same. Even the coordinates of the body parts of the same driver in the standard posture at each point in time are different. Therefore, during driving, the driver's standard posture is continuously determined and updated in data storage (Step 32).

During driving, the current coordinates of the body parts in driver side view images are compared with the ones in the standard posture (Step 33). If the image shows the left ear without the right ear (Step 34), the driver is looking backward (Step 35). If both eyes and ears are visible in the image (Step 36), the driver is talking to passengers (Step 37). If the deviation of hands or nose to the standard locations is larger than preset thresholds (Step 38), the driver may be doing an abnormal behavior such as drinking, eating, making up, smoking, etc (Step 39). When the system detects the above cases, an alerting signal will be sent to the driver (Step 40). Otherwise, no action is taken (Step 41). The results of monitoring driver behavior in side view images are illustrated in Figures 12A-12D.

5. The main points to be protected by the patent

(1) A driver behavior monitoring system based on body part coordinates and AIGC

6. Deduce the advantages of this patent by reasoning

Novelty:

* Using an object detection model to detect driver drowsiness.
* Based on the body part coordinates, detect abnormal behaviors of the driver.

Creativity:

* Using an AIGC model to generate images of a driver in different situations.
* Detecting a driver using a smartphone while driving with the prediction results of an object detection model and a body posture estimation model.

Practicality:

* Generating driver images in different situations helps reduce data collection costs.
* Improving the diversity of the dataset, helping the model better predict new situations.
* The comparison with the standard body posture of the driver helps to detect more types of driver abnormal behavior than previous studies.

7. Other alternative that can also achieve the purpose of the patent (tìm hiểu thêm)

(1) Image generation technique based on Generative Adversarial Networks (GANs).

8. Other documents that help patent attorneys understand the technology

8.1 Determining the coordinates of human body parts using a human body posture estimation model

The technique of determining the coordinates of human body parts in an image based on a human body posture estimation model is often called "Human Pose Estimation" or "Body Keypoint Detection." This is one of the important applications of artificial intelligence in image processing and human body posture analysis. The main goal of this technique is to determine the coordinates of key points on the human body, often called "keypoints" or "body parts." Keypoints typically include the head, shoulders, wrists, hips, knees, and eyes. By identifying these keypoints, a histogram of the human body pose in the image can be created.

The Human Pose Estimation technique uses machine learning models to determine the coordinates of keypoints on the image. Some popular methods include:

Convolutional neural networks (CNN): CNNs are often used to extract features from images and predict the coordinates of keypoints. The model can be trained on large datasets with images labeled with keypoints.

Supervised and unsupervised deep learning: There are both supervised and unsupervised methods for identifying keypoints. Supervised methods use labeled data sets, while unsupervised methods try to find keypoints without label information..

8.2 Comparison of traditional data augmentation techniques and image generation techniques using the AIGC model

Data augmentation techniques based on image processing and image generation techniques using the AIGC (Artificial Intelligence Generated Content) model are two different methods to create new data in the field of image processing. Below is a comparison between them.

* Method: Image Processing-based data augmentation technique creates new data by changing the original data through image transformations such as rotation, zoom, flip, noise, or cropping. The image generation technique using the AIGC model uses a machine learning model, including neural networks and algorithms, to automatically generate new image data.
* Ability to generate new data: Data augmentation techniques are limited by predefined image transformations, which cannot generate completely new or diverse data. The AIGC model is capable of generating completely new and diverse data, not limited by specific image transformations.
* Accuracy: Data generated through data augmentation techniques often retain the properties of the original data and can easily be controlled for accuracy if reasonable techniques are used. Data generated using the AIGC model may have lower accuracy and needs to be carefully checked to ensure reliability.

**9. Description of the photos**

Figure **1** illustrates the structure of the proposed system.

Figure **2** illustrates the original data designing process.

Figure **3** illustrates image collection for driver abnormal behavior detection.

Figure **4** illustrates the image collection for driver drowsiness detection

Figure **5** illustrates the image generation using the AIGC technique.

Figure **6** illustrates the flowchart of model training for detecting the objects in driver front view images.

Figure **7** illustrates the flowchart of driver drowsiness detection.

Figure **8** illustrates the technique for detecting a driver using a smartphone while driving using the coordinates of the smartphone and hands.

Figure **9** illustrates a model of the body part.

Figure **10** illustrates a driver using a smartphone while driving detected by the proposed method, where the black rectangle denotes the smartphone in the image.

Figure **11** illustrates the flowchart for detecting abnormal behaviors of the driver.

Figure **12A** illustrates the body part coordinates of a driver in normal status.

Figure **12B** illustrates the body part coordinates of a driver talking to passengers and using a smartphone.

Figure **12A** illustrates the body part coordinates of a driver looking backward.

Figure **12B** illustrates the body part coordinates of a driver making up while driving.

BODY POSTURE ESTIMATION MODEL

DRIVER DROWSINESS DETECTION MODEL

IMAGE PROCESSING UNIT

IMAGE ACQUISITION SUBSYSTEM

ALERTING SUBSYSTEM

DATA STORAGE

STEERING WHEEL AND SMARTPHONE DETECTION MODEL

Figure 1

MOUNTING ONE CAMERA IN FRONT OF THE DRIVER AND ANOTHER ON THE RIGHT SIDE OF THE DRIVER

USING THE DRIVER FRONT VIEW CAMERA TO CAPTURE IMAGES OF A DROWSY DRIVER, WHILE THE DRIVER'S ABNORMAL BEHAVIOR IS CAPTURED BY THE DRIVER SIDE VIEW CAMERA

LABELING ALL THE COLLECTED IMAGES

DESIGNING THE ORIGINAL DATASET FOR THE AIGC MODEL

1

2

3

4

Figure 2



Figure 3

Figure 4



Figure 5

Figure 6

SETTING THE PARAMETERS FOR THE AIGC MODEL

TRAINING THE AIGC MODEL

GIVING A PROMPT TO THE AIGC MODEL AS THE INPUT

GENERATING THE IMAGES OF A DRIVER IN DIFFERENT SITUATIONS

OBTAINING THE AIGC MODEL FOR GENERATING DRIVER IMAGES

5

6

7

8

9

ORIGINAL DRIVER FRONT VIEW IMAGES AND THE CORRESPONDING ONES GENERATED BY AIGC MODEL

DESIGNING A DRIVER FRONT VIEW DATASET FOR AN OBJECT DETECTION MODEL

TRAINING THE MODEL

LABELING THE DATA

10

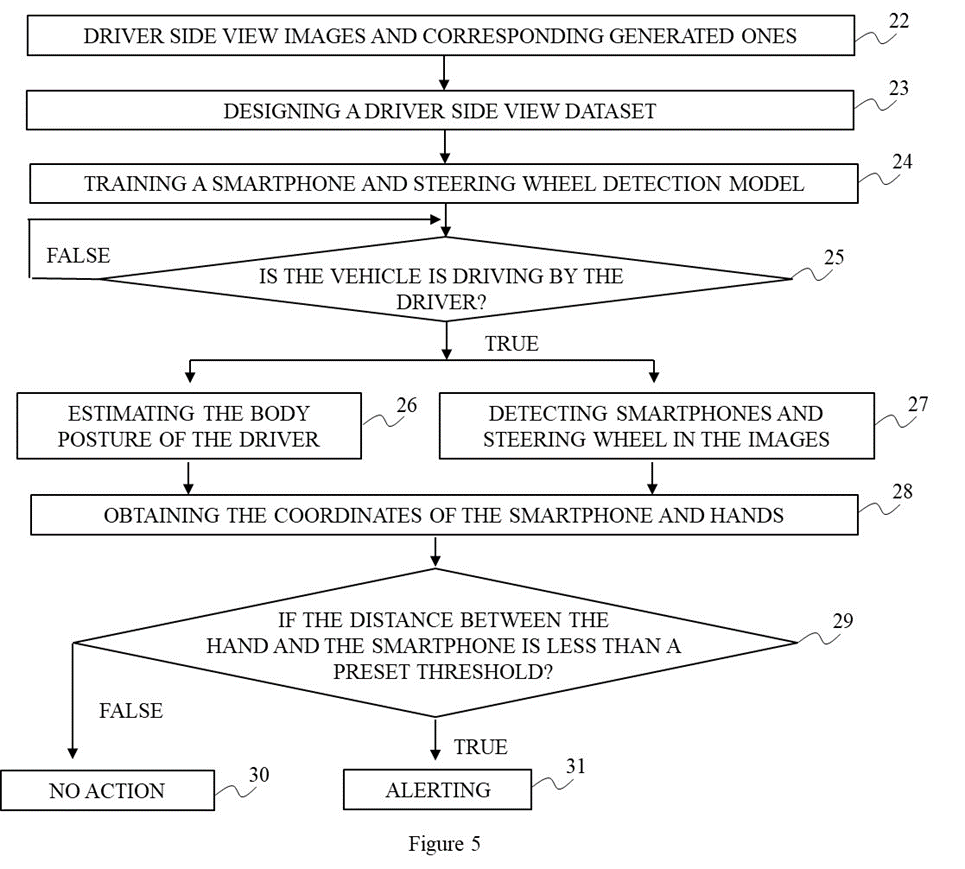
11

12

13

Figure 7

Figure 8



14

ALERTING

DROWSY

DROWSY

DETECTING OBJECTS IN DRIVER FRONT VIEW IMAGES

IF CLOSED-EYE OR YAWNING-MOUTH IS DETECTED?

IF SUNGLASSES AND COVERING-HAND ARE DETECTED?

NORMAL

NO ACTION

TRUE

TRUE

FALSE

FALSE

15

17

16

18

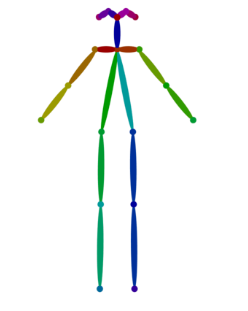
19

20

21

Figure 9

1. Nose
2. Neck
3. Right shoulder
4. Right elbow
5. Right hand
6. Left shoulder
7. Left elbow
8. Left hand
9. Right hip
10. Right knee
11. Right foot
12. Left hip
13. Left knee
14. Left foot
15. Right eye
16. Left eye
17. Right ear
18. Left ear



0

1

2

3

4

5

6

7

8

9

10

11

12

13

14

16

15

17

Figure 10



UPDATING THE DRIVER'S STANDARD POSTURE IN DATA STORAGE

COMPARING THE CURRENT COORDINATES OF THE BODY PARTS WITH THE ONES IN THE STANDARD POSTURE

LEFT EAR DETECTED BUT NO RIGHT

EAR?

LOOKING BACKWARD

BOTH EARS

AND EYES ARE DETECTED?

TALKING TO PASSENGERS

DEVIATION OF HANDS

OR NOSE TO THE STANDARD LOCATIONS IS LARGER THAN

PRESET THRESHOLDS?

OTHER ABNORMAL BEHAVIORS

NO ACTION

ALERTING

FALSE

FALSE

FALSE

TRUE

TRUE

TRUE

Figure 11

32

33

34

35

37

39

40

41

36

38

Figure 12C

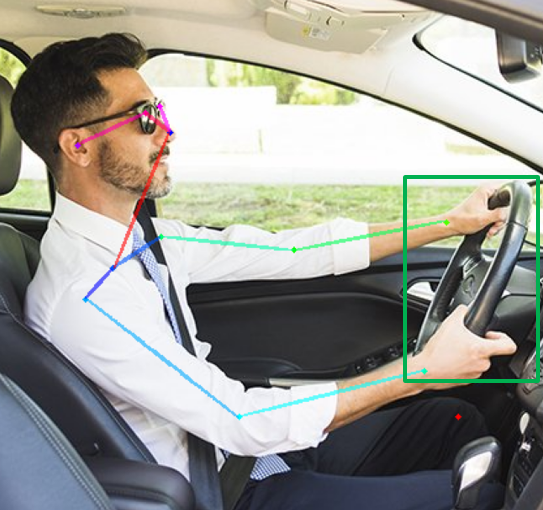
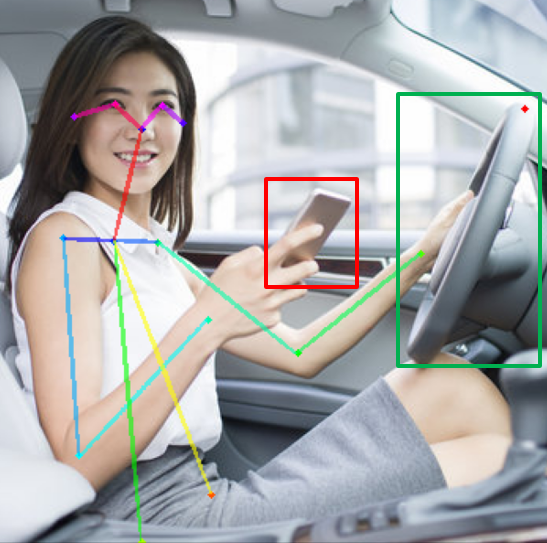
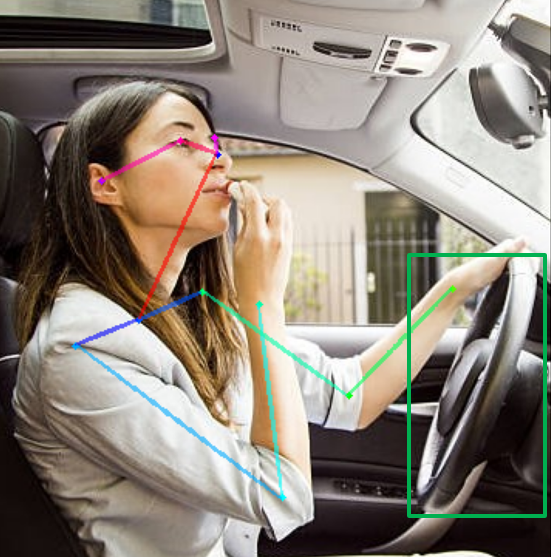


Figure 12D

Figure 12A

Figure 12B