CLAIMS

1. An action recognition method, comprising:

obtaining mouth key points of a face based on a face image;

determining, based on the mouth key points, an image in a first region involves at least part of the mouth key points and comprises an image of an object interacting with mouth; and

determining whether a person corresponding to the face image is smoking based on the image in the first region,

wherein before determining whether the person corresponding to the face image is smoking based on the image in the first region, the method further comprises:

obtaining at least two second key points of the object interacting with mouth based on the image in the first region,

aligning, based on the at least two second key points, the object interacting with mouth in a way that the object interacting with mouth is oriented to a preset direction, and

obtaining an image in a second region involving the object interacting with mouth and oriented to the preset direction, wherein the image in the second region involves at least part of the mouth key points and comprises an image of the object interacting with mouth;

wherein determining whether the person corresponding to the face image is smoking based on the image in the first region comprises:

determining whether the person corresponding to the face image is smoking based on the image in the second region.

2. The method according to claim 1, wherein before determining whether the person corresponding to the face image is smoking based on the image in the first region, the method further comprises:

obtaining at least two first key points of the object interacting with mouth based on the image in the first region; and

screening the image in the first region based on the at least two first key points, to select out the image in the first region in which the object interacting with mouth and having a length greater than or equal to a preset value is involved,

determining whether the person corresponding to the face image is smoking based on the image in the first region comprises:

in response to that the image in the first region passes the screening, determining whether the person corresponding to the face image is smoking based on the image in the first region.

3. The method according to claim 2, wherein screening the image in the first region based on the at least two first key points comprises:

determining key point coordinates corresponding to the at least two first key points in the image in the first region; and

screening the image in the first region based on the key point coordinates corresponding to the at least two first key points.

4. The method according to claim 3, wherein screening the image in the first region based on the key point coordinates corresponding to the at least two first key points comprises:

determining, based on the key point coordinates corresponding to the at least two first key points, a length of the object interacting with mouth which is involved in the image in the first region; and

in response to that the length of the object interacting mouth is greater than or equal to the preset value, determining that the image in the first region passes the screening.

5. The method according to claim 4, further comprising:

in response to that the length of the object interacting with mouth is less than the preset value, determining that the image in the first region fails to pass the screening; and

determining that the image in the first region does not involve a cigarette.

6. The method according to any of claims 3 to 5, wherein before determining the key point coordinates corresponding to the at least two first key points in the image in the first region, the method further comprises:

assigning a serial number to each of the at least two first key points to distinguish the at least two first key points.

7. The method according to any of claims 3 to 6, wherein determining the key point coordinates corresponding to the at least two first key points in the image in the first region comprises:

determining, by a first neural network, the key point coordinates corresponding to the at least two first key points in the image in the first region, wherein the first neural network is trained with a first sample image.

8. The method according to claim 7, wherein the first sample image comprises labelled key point coordinates, and training the first neural network comprises:

inputting the first sample image into the first neural network to obtain predicted key point coordinates corresponding to the at least two first key points;

determining a first network loss based on the predicted key point coordinates and the labelled key point coordinates; and

adjusting a parameter of the first neural network based on the first network loss.

9. The method according to any of claims 2 to 8, wherein obtaining the at least two first key points of the object interacting with mouth based on the image in the first region comprises:

performing a key point recognition for the object interacting with mouth on the image in the first region to obtain

at least two central axis key points on a central axis of the object interacting with mouth, and/or

at least two side key points on each of two sides of the object interacting with mouth.

10. The method according to any of claims 1 to 9, wherein determining whether the person corresponding to the face image is smoking based on the image in the first region comprises:

determining, by a second neural network, whether the person corresponding to the face image is smoking based on the image in the first region, wherein the second neural network is trained with a second sample image.

11. The method according to claim 10, wherein the second sample image is associated with a label of whether the person corresponding to the second sample image is smoking, and training the second neural network comprises:

inputting the second sample image into the second neural network to obtain a prediction of whether a person corresponding to the second sample image is smoking;

obtaining a second network loss based on the prediction and the label, and

adjusting a parameter of the second neural network based on the second network loss.

12. The method according to any of claims 1 to 11, wherein obtaining the mouth key points of the face based on the face image comprises:

performing a face key point extraction on the face image to obtain face key points in the face image;

obtaining the mouth key points based on the face key points.

13. The method according to claim 12, wherein determining the image in the first region based on the mouth key points comprises:

determining a center of a mouth involved in the face image based on the mouth key points;

determining the first region by taking the center of the mouth as a center point of the first region and taking a preset length as a side length or a radius.

14. The method according to claim 13, wherein before determining the image in the first region based on the mouth key points, the method further comprises:

obtaining at least one eyebrow key point based on the face key points;

determining the first region by taking the center of the mouth as the center point of the first region and taking the preset length as the side length or the radius comprises:

determining the first region by taking the center of the mouth as the center point of the first region, and taking a vertical distance from the center of the mouth to a center of an eyebrow as the side length or the radius, wherein the center of the eyebrow is determined based on the at least one eyebrow key point.

15. An action recognition apparatus, comprising:

a mouth key point module, configured to obtain mouth key points of a face based on a face image;

a first region determining module, configured to determine, based on the mouth key points, an image in a first region involves at least part of the mouth key points and comprises an image of an object interacting with mouth; and

a smoking recognizing module, configured to determine whether a person corresponding to the face image is smoking based on the image in the first region,

wherein the apparatus further comprises:

a second key point module, configured to obtain at least two second key points of the object interacting with mouth based on the image in the first region; and

an image aligning module, configured to align, based on the at least two second key points, the object interacting with mouth in a way that the object interacting with mouth is oriented to a preset direction, and obtain an image in a second region involving the object interacting with mouth and oriented to the preset direction, wherein the image in the second region involves at least part of the mouth key points and comprises an image of the object interacting with mouth,

wherein the smoking recognizing module is configured to, when determining whether the person corresponding to the face image is smoking based on the image in the first region, determine whether the person corresponding to the face image is smoking based on the image in the second region.

16. An electronic device, comprising: a processor, wherein the processor comprises an action recognition apparatus according to claim 15.

17. An electronic device, comprising:

a memory configured to store executable instructions; and

a processor configured to communicate with the memory to execute the executable instructions to perform operations in an action recognition method according to any of claims 1 to 14.

18. A computer readable storage medium, configured to store computer readable instructions, wherein the instructions are executed to perform the action recognition method according to any of claims 1 to 14.

19. A computer program product comprising computer readable codes, wherein the computer readable codes are running on a device to cause a processor in the device to execute instructions for implementing an action recognition method according to any of claims 1 to 14.