METHOD FOR THE EVENT DETECTION AND APPARATUS, ELECTRONIC DEVICE, STORAGE MEDIUM, AND PROGRAM PRODUCT

CROSS-REFERENCE TO RELATED APPLICATION

[0001] The present disclosure is based upon and claims priority to Chinese Patent Application No. 202110236981.X, filed by Shanghai SenseTime Intelligent Technology Co., Ltd. on March 3, 2021 and entitled “METHOD FOR THE EVENT DETECTION AND APPARATUS, ELECTRONIC DEVICE, AND STORAGE MEDIUM”, the disclosure of which is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] The present disclosure relates to the field of computer vision technologies, and in particular to a method and apparatus for an event detection, an electronic device, a storage medium, and a program product.

BACKGROUND

[0003] With the development of economy and the continuous improvement of infrastructure construction, escalators are more and more widely applied in scenarios such as shopping malls, office buildings, and public transportation. The escalators bring convenience to pedestrians, and meanwhile, possible accidents caused by abnormal actions of the pedestrians on the escalators, such as falling, retrograding, and squatting, have attracted more and more attention of people. If the abnormal actions on the escalators cannot be detected in time, and an alarm cannot be given in time to stop the operation of the escalators, it may cause heavy losses to the life and property safety of the pedestrians on the escalators.

SUMMARY

[0004] The present disclosure provides a method and apparatus for an event detection, an electronic device, a storage medium, and a program product.

[0005] According to an aspect of the present disclosure, a method for an event detection is provided, which may include that: a video stream of an escalator is acquired; a human body detection is performed on the video stream to determine a first region where a target object is located in the video steam; a posture recognition is performed on a first region image corresponding to the first region to obtain a posture recognition result of the target object, herein the posture recognition result includes a confidence that a posture of the target object is an abnormal posture; and an event detection result of the target object is determined according to the posture recognition result of the target object.

[0006] In a possible implementation, the operation that an event detection result of the target object is determined according to the posture recognition result of the target object may include that: in a case where a confidence of the target object at a present time is greater than or equal to a confidence threshold value, the event detection result of the target object is determined to be that an abnormal event occurs.

[0007] In this manner, it can be directly determined that an abnormal event occurs when the target object is an abnormal posture, so that the real-time performance of event detection is improved, thereby giving the alarm in time and reducing the risk of safety accidents.

[0008] In a possible implementation, the operation that an event detection result of the target object is determined according to the posture recognition result of the target object may include that: in a case where a confidence of the target object at a present time is greater than or equal to a confidence threshold value and a first duration at which the confidence is greater than or equal to the confidence threshold value is greater than or equal to activate seconds, the event detection result of the target object is determined to be that an abnormal event occurs.

[0009] In this manner, it can be determined that an abnormal event occurs when the target object is an abnormal posture and lasts for a certain duration, so that possible false detection during single-frame detection is reduced and the accuracy of event detection is improved.

[0010] In a possible implementation, the method may further include that: in a case where the event detection result of the target object is that the abnormal event occurs, first alarm information is sent.

[0011] In this manner, timely alarm of the abnormal event can be realized, so as to reduce the risk of safety accidents.

[0012] In a possible implementation, the method may further include that: in a case where the event detection result of the target object is determined to be that an abnormal event occurs, a detection state of the target object is changed into a detected state; and in a case where the detection state of the target object is the detected state, the confidence of the target object at a present time is greater than or equal to the confidence threshold value, and a second duration at which the confidence is greater than or equal to the confidence threshold value is greater than cooldown seconds, the event detection result of the target object is determined to be that an abnormal event lasts; and in a case where the event detection result of the target object is that the abnormal event lasts, second alarm information is sent.

[0013] In this manner, it can be determined that an abnormal event lasts in a case where the target object is an abnormal posture and the cooldown seconds is reached, so that repeated alarm is performed by taking the cooldown seconds as a time interval to repeatedly remind a device for monitoring the operation of an escalator or relevant personnel in charge of the operation of the escalator, thereby reducing the cases where relevant personnel do not observe the alarm information, and reducing the risks of safety accidents.

[0014] In a possible implementation, the method may further include that: in a case where the detection state of the target object is the detected state, the confidence of the target object at a present time is less than the confidence threshold value, and a third

duration at which the confidence is less than the confidence threshold value is greater than a deactivate seconds, the event detection result of the target object is determined to be that an abnormal event is ended.

[0015] In this manner, it can be determined that the abnormal event is ended when the target object is not an abnormal posture and the deactivate seconds is reached, so as to complete the whole process of event detection.

[0016] In a possible implementation, the operation that a first region image corresponding to the first region is determined may include that: the first region is expanded according to a preset expansion coefficient to obtain a second region; and an image of the second region is clipped out from video frames of the video stream, to obtain the first region image.

[0017] In this manner, the performance of a classifier of a posture recognition network can be improved, and the accuracy of subsequent posture recognition can be improved.

[0018] In a possible implementation, the operation that the posture recognition is performed on the first region image corresponding to the first region may include that: posture recognition is performed on the first region image corresponding to the first region through a posture recognition network. The method further includes that: training is performed according to a preset training set to obtain the posture recognition network, herein an object posture marked by a sample image in the training set includes at least one of the following: an abnormal posture, a normal posture, or other postures; and the other postures include the postures except the normal posture and the abnormal posture.

[0019] In this manner, the robustness of posture recognition can be improved, and the accuracy of posture recognition can be improved.

[0020] In a possible implementation, the abnormal event may include a squatting event.

[0021] According to an aspect of the present disclosure, an apparatus for an event detection is provided, which may include: a video stream acquisition part, a region determination part, a posture recognition part and an event detection part. Herein, the

video stream acquisition part is configured to acquire a video stream of an escalator; a region determination part is configured to perform a human body detection on the video stream to determine a first region where a target object is located in the video steam; a posture recognition part is configured to perform a posture recognition on a first region image corresponding to the first region to obtain a posture recognition result of the target object, herein the posture recognition result includes a confidence that a posture of the target object is an abnormal posture; and an event detection part is configured to determine an event detection result of the target object according to the posture recognition result of the target object.

[0022] In a possible implementation, the event detection part may further be configured to: in a case where the confidence of the target object at a present time is greater than or equal to a confidence threshold value, determine the event detection result of the target object to be that an abnormal event occurs.

[0023] In a possible implementation, the event detection part may further be configured to: in a case where the confidence of the target object at a present time is greater than or equal to a confidence threshold value and a first duration at which the confidence is greater than or equal to the confidence threshold value is greater than or equal to the activate seconds, determine the event detection result of the target object to be that an abnormal event occurs.

[0024] In a possible implementation, the apparatus may further include: a first alarm part, configured to: in a case where the event detection result of the target object is that the abnormal event occurs, send first alarm information.

[0025] In a possible implementation, the apparatus may further include: a state change part, configured to: in a case where the event detection result of the target object is that an abnormal event occurs, change a detection state of the target object into a detected state. The event detection part may further be configured to: in a case where the detection state of the target object is the detected state, the confidence of the target object at a present time is greater than or equal to the confidence threshold value, and a second duration at which the confidence is greater than or equal to the confidence threshold value is greater than cooldown seconds, determine the event detection result of the target

object to be that an abnormal event lasts. The apparatus may further include: a first alarm part, configured to: in a case where the event detection result of the target object is that the abnormal event lasts, send second alarm information.

[0026] In a possible implementation, the event detection part is further configured to: in a case where the detection state of the target object is the detected state, the confidence of the target object at a present time is less than the confidence threshold value, and a third duration at which the confidence is less than the confidence threshold value is greater than the deactivate seconds, determine the event detection result of the target object to be that an abnormal event is ended.

[0027] In a possible implementation, the apparatus may further include: an expansion part and a clipping part. The expansion part is configured to: expand the first region according to a preset expansion coefficient to obtain a second region; and the clipping part is configured to: clip out an image of the second region from video frames of the video stream, to obtain the first region image.

[0028] In a possible implementation, the posture recognition part may further be configured to: perform the posture recognition on the first region image corresponding to the first region through a posture recognition network. The apparatus may further include: a training part, configured to perform training according to a preset training set to obtain the posture recognition network, herein an object posture marked by a sample image in the training set includes at least one of the following: an abnormal posture, a normal posture, and other postures; and the other postures comprise postures except the normal posture and the abnormal posture.

[0029] In a possible implementation, the abnormal event may include a squatting event.

[0030] According to an aspect of the present disclosure, an electronic device is provided, including a processor, and a memory configured to store a processor executable instruction. The processor is configured to call the instruction stored in the memory to implement the abovementioned method.

[0031] According to an aspect of the present disclosure, a computer-readable storage

medium is provided, in which a computer program instruction may be stored, and the computer program instruction is executed by a processor to implement the abovementioned method.

[0032] According to an aspect of the present disclosure, a computer program product is provided, including a computer-readable code. When the computer-readable code runs on a device, a processor in the device executes an instruction for implementing the abovementioned method for the event detection.

[0033] According to the embodiments of the present disclosure, a region where an object in the video stream can be detected; an image of the region is recognized to determine a posture of the object; the event detection result is determined according to the posture of the object, which can improve accuracy of event detection, so as to reduce the probability of accidents, and reduce the cost of manual detection.

[0034] It is to be understood that the above general description and the following detailed description are only exemplary and explanatory and not intended to limit the present disclosure. According to the following detailed descriptions made to exemplary embodiments with reference to the drawings, other features and aspects of the present disclosure may become clear.

BRIEF DESCRIPTION OF THE DRAWINGS

[0035] For describing the technical solutions of the embodiments of the present disclosure more clearly, the drawings required to be used in the embodiments will be simply introduced below. The drawings, which are incorporated in and constitute a part of this specification, illustrate embodiments consistent with the present disclosure and, together with the specification, serve to explain the technical solutions of the present disclosure. It is to be understood that the following drawings only illustrate some embodiments of the present disclosure and thus should not be considered as limitation to the scope. Those of ordinary skill in the art may also obtain other related drawings according to these drawings without creative work.

[0036] FIG. 1 illustrates a flowchart of a method for an event detection according to the embodiments of the present disclosure.

[0037] FIG. 2a illustrates a schematic diagram of a detection state change of a target object according to the embodiments of the present disclosure.

[0038] FIG. 2b illustrates a schematic diagram of a process framework of a method for an event detection according to the embodiments of the present disclosure.

[0039] FIG. 3 illustrates a block diagram of an apparatus for an event detection according to the embodiments of the present disclosure.

[0040] FIG. 4 illustrates a block diagram of an electronic device according to the embodiments of the present disclosure.

[0041] FIG. 5 illustrates a block diagram of another electronic device according to the embodiments of the present disclosure.

DETAILED DESCRIPTION

[0042] Each exemplary embodiment, feature, and aspect of the present disclosure will be described below with reference to the drawings in detail. The same reference numerals in the drawings represent components with the same or similar functions. Although each aspect of the embodiments is illustrated in the drawings, the drawings are not required to be drawn to scale, unless otherwise specified.

[0043] Herein, special term "exemplary" refers to "use as an example, embodiment, or description". Herein, any "exemplarily" described embodiment may not be explained to be superior to or better than other embodiments.

[0044] The term "and/or" in this specification describes only an association relationship for describing associated objects and represents that three relationships may exist. For example, A and/or B may represent the following three cases: Only A exists, both A and B exist, and only B exists. In addition, term "at least one" in the present

disclosure represents any one of multiple or any combination of at least two of multiple. For example, including at least one of A, B, or C may represent including any one or more elements selected from a set formed by A, B, or C.

[0045] In addition, for describing the present disclosure better, many specific details are presented in the following specific implementations. It is understood by those skilled in the art that the present disclosure may still be implemented even without some specific details. In some examples, methods, means, components and circuits known very well to those skilled in the art are not described in detail, so as to highlight the subject of the present disclosure.

[0046] The method for the event detection according to the embodiments of the present disclosure may be applied to scenarios such as shopping malls, office buildings, and public transportation. Based on a deep learning method, a video stream in a region where an escalator in a scenarios is located is processed and analyzed, which can detect an abnormal event (such as falling, retrograding, and squatting) of an object (such as pedestrian) in real time, locate a region where the abnormal event occurs on the escalator, and give an alarm in time, so as to stop the operation of the escalator, thereby reducing the risk of safety accidents.

[0047] FIG. 1 illustrates a flowchart of a method for an event detection according to the embodiments of the present disclosure. As illustrated in FIG. 1, the method for the event detection includes the following steps.

[0048] At S11, a video stream of an escalator is acquired.

[0049] At S12, a human body detection is performed on the video stream to determine a first region where a target object is located in the video steam.

[0050] At S13, posture recognition is performed on a first region image corresponding to the first region to obtain a posture recognition result of the target object. The posture recognition result includes the confidence that a posture of the target object is an abnormal posture.

[0051] At S14, an event detection result of the target object is determined according to

the posture recognition result of the target object.

[0052] In a possible implementation, the method for the event detection may be executed by electronic devices such as a terminal device or a server. The terminal device may be one of the following: a mobile phone, a pad, a computer with a wireless transceiving function, a palm computer, a desk computer, a personal digital assistant, a portable media player, a smart speaker, a navigation apparatus, wearable devices (such as a smart watch, smart glasses, a smart necklace), a pedometer, a digital TV, a Virtual Reality (VR) terminal device, an Augmented Reality (AR) terminal device, a wireless terminal in industrial control, a wireless terminal in self driving, a wireless terminal in remote medical surgery, a wireless terminal in a smart grid, a wireless terminal in transportation safety, a wireless terminal in a smart city, a wireless terminal in a smart home, a vehicle, an on-board device and an on-board module in an Internet of vehicles system, etc. The method may be implemented in a mode that a processor calls a computer-readable instruction stored in a memory. Or, the method may be executed by a server.

[0053] For example, at least one image collection device is arranged at a position where an escalator to be detected is located, for example, at least one camera facing the escalator is arranged, to collect a video stream of the escalator, and detect an object (such as a pedestrian) taking the escalator in video stream pictures. No limits are made to an installation position of the image collection device, a collection mode of the video stream, and a specific region corresponding to the video stream in the present disclosure.

[0054] In a possible implementation, at S11, a video stream of an escalator may be acquired. For example, the video stream uploaded by the image collection device is received, and the collected video stream is decoded, to obtain a decoded video stream (which may also be called a picture stream).

[0055] In a possible implementation, at S12, human detection may be performed on the decoded video stream, to determine a human body box in each video frame of the video stream; and the human body box in each video frame is followed to determine a human body box of a pedestrian (which may be called a target object) that belongs to the same identity, that is, a first region where the target object is located in the video stream

is determined.

[0056] A manner of human body detection may be, for example, human body key point recognition, human body contour detection, etc. A manner of human body following may be, for example, that objects belonging to the same identity are determined according to an intersection-over-union (IoU) of human body boxes in adjacent video frames. Those skilled in the art should understand that human body detection and human body following can be realized in any mode in a related technology, no limits are made thereto in the present disclosure.

[0057] In a possible implementation, human body detection and human body following may be performed on each video frame of the video stream; or the video stream may be sampled at a certain time interval, and human body detection and human body following are performed on the sampled video stream; and a key frame in the video stream may also be acquired, and human body detection and human body following are performed on the key frame. No limits are made thereto in the present disclosure.

[0058] In a possible implementation, a first region may be represented as vertex coordinates of an upper left corner (x1, y1) and a lower right corner (x2, y2) of a human body box (also referred to as a Region of Interest (ROI)), and a following Identify (ID) of the target object, and no limits are made to a representation mode of the first region in the present disclosure.

[0059] In a possible implementation, a first region image corresponding to the first region may be clipped out from each video frame; and at S13, posture recognition is performed on the first region image by a trained posture recognition network, to obtain a posture recognition result of the target result. The posture recognition network may be, for example, a convolutional neural network. No limits are made to a specific network structure and a training manner of the posture recognition network in the present disclosure.

[0060] In a possible implementation, the posture recognition result includes the confidence that the posture of the target object is an abnormal posture, so as to determine, in subsequent processing, whether the abnormal event occurs. The posture recognition

result may also include the confidence that the posture of the target object is a normal posture and the confidence that the posture of the target object is other postures except the normal posture and the abnormal posture. No limits are made to specific content of the posture recognition result in the present disclosure.

[0061] In a possible implementation, the abnormal posture may be, for example, a squatting posture and a falling posture, etc. The normal posture may be, for example, a standing posture. Other postures may be, for example, the postures of lying down and lying prone, or the postures that cannot be recognized. No limits are made to specific types of various postures in the present disclosure.

[0062] In a possible implementation, at S14, the event detection result of the target object may be determined according to the posture recognition result of the target object. The event detection result may include that an abnormal event occurs or an abnormal event does not occur. For example, when the confidence of the abnormal posture exceeds a threshold value, or the confidence exceeds the threshold value and lasts for a certain duration, the event detection result is determined to be that the abnormal event occurs. The abnormal event may be determined according to the type of the abnormal posture, for example, a squatting event and a falling event. No limits are made thereto in the present disclosure.

[0063] In a possible implementation, if the event detection result is that an abnormal event occurs, alarm information can be generated and sent. The alarm information may include a reminder of an abnormal event, and may also include a region where the target object to which the abnormal event occurs is located, so as to facilitate locating by relevant personnel. No limits are made thereto in the present disclosure.

[0064] In a possible implementation, the alarm information may be, for example, sent to a control apparatus of the escalator, so that the control apparatus stops the operation of the escalator. The alarm information may also be sent to a device (such as a terminal device and/or a display device) that can be seen by relevant personnel in charge of the operation of the escalator, so that the relevant personnel stop the operation of the escalator and go to the escalator for rescue. No limits are made to the contents of the alarm information in the present disclosure.

[0065] According to the embodiments of the present disclosure, a region where an object in the video stream can be detected; an image of the region is recognized to determine a posture of the object; the event detection result is determined according to the posture of the object, which can improve accuracy of event detection, so as to reduce the probability of accidents, and reduce the cost of manual detection.

[0066] The method for the event detection of the embodiments of the present disclosure is described hereafter.

[0067] As previously described, a video stream of a region where an escalator is located may be collected through a camera, and the collected video stream is transmitted to an electronic device, such as a local front-end server or a cloud server. The electronic device may decode the received video stream, to obtain a decoded video stream.

[0068] At S12, human body detection and human body following may be performed on the decoded video stream through a detection network, to detect a human body box in each video frame of the video stream; and the human body box of the pedestrian belonging to the same identity is followed, to obtain a first region where the target object is located in the video stream. The detection network may be a convolutional neural network, and no limits are made to a network structure of the detection network in the present disclosure.

[0069] After the first region is obtained, a first region image corresponding to the first region may be acquired. An image of the first region may be directly clipped out from the video frames of the video stream, and expansion processing may also be performed on the first region.

[0070] In a possible implementation, the operation that the first region image corresponding to the first region is determined according to the method for the event detection of the embodiments of the present disclosure includes the following operations.

[0071] The first region is expanded according to a preset expansion coefficient to obtain a second region, herein the second region is the expanded first region.

[0072] An image of the second region is clipped out from video frame(s) of the video

stream, to obtain the first region image.

[0073] As previously described, the first region may be represented as vertex coordinates of an upper left corner (x1, y1) and a lower right corner (x2, y2) of a human body box, and a following ID of the target object. An expansion coefficient may be preset, and the first region is expanded according to the expansion coefficient. The upper left corner and the lower right corner of the expanded second region are set as (x_1', y_1') and (x_2', y_2'), respectively, then:

$$x_1' = \max\left(0, x_1 - e(x_2 - x_1)\right)$$

$$y_1' = \max\left(0, y_2 - e(y_2 - y_1)\right)$$

$$x_2' = \min\left(W - 1, x_2 + e(x_2 - x_1)\right)$$

$$y_2' = \min\left(H - 1, y_2 + e(y_2 - y_1)\right) \qquad \text{Formula (1);}$$

[0074] In Formula (1), W and H respectively represent the width and the height of the video frame, and e represents the expansion coefficient. The value of the expansion coefficient e is, for example, 0.1. No limits are made to a specific value of the expansion coefficient e in the present disclosure.

[0075] According to Formula (1), the expansion of the first region in an image range of the video frame can be realized, so as to obtain a second region. Then, an image of the second region may be clipped out from the corresponding video frame to obtain a first region image, so as to facilitate subsequent processing.

[0076] In this manner, the performance of a classifier of a posture recognition network can be improved, and the accuracy of subsequent posture recognition can be improved.

[0077] In a possible implementation, posture recognition may be performed on the first region image by a posture recognition network at S13, so as to obtain a posture recognition result of the target result. The posture recognition network may include a multi-classifier, so as to recognize the posture of the target object as an abnormal posture, a normal posture, or other postures. Before deploying the posture recognition network,

the posture recognition network may be trained.

[0078] In a possible implementation, the method for the event detection of the embodiments of the present disclosure may further include the following operation.

[0079] Training is performed according to a preset training set to obtain the posture recognition network, herein an object posture marked by a sample image in the training set includes at least one of the following: an abnormal posture, a normal posture, and other postures; and the other postures include the postures except the normal posture and the abnormal posture.

[0080] For example, in order to improve the robustness of the classifier of the posture recognition network, an open set training mode can be adopted. That is, the object postures marked by the sample image further include other postures except the abnormal posture and the normal posture. For example, in a case where the abnormal posture and the normal posture are respectively a squatting posture and a standing posture, other postures may include the postures of lying down, lying prone, etc., may also include the situations that the posture cannot be determined, for example, a lower body is completed covered, too blurred, etc., or may also include that a detection box covers a plurality of pedestrian and that who is a main pedestrian cannot be determined. No limits are made to both posture types included in other postures and specific situations without a posture in the present disclosure.

[0081] During training, sample images in a training set may be respectively input into a posture recognition network to be trained for processing, to obtain a sample recognition result. According to a difference between the sample recognition result and the sample images, a network loss of the posture recognition network to be trained is determined; parameters of the posture recognition network to be trained are reversely adjusted according to the network loss; a trained posture recognition network is obtained in a case of network convergence through a plurality of iterations.

[0082] In this manner, the robustness of posture recognition can be improved, and the accuracy of posture recognition can be improved.

[0083] In a possible implementation, the posture recognition result may include the

confidence s that the posture of the target object is an abnormal posture. At S14, event detection may be performed according to the posture recognition result of the target object, to determine an event detection result of the target object.

[0084] In a possible implementation, a state machine may be maintained for each following ID (that is, target object), and the state of the state machine is updated according to the posture recognition result of each video frame, so as to realize event detection.

[0085] FIG. 2a illustrates a schematic diagram of a detection state change of a target object according to the embodiments of the present disclosure. As illustrated in FIG. 2a, the detection state of the target object may include an idle state (IDLE) 21, a state under detecting (DETECTING) 22, and a detected state (DECTECTED) 23. During detecting, the detection state may be changed according to various conditions in FIG. 2a, and a corresponding event detection result is output in a case of satisfying a certain condition.

[0086] In FIG. 2a, all of the Activate seconds (AC), the Cooldown seconds (CD), the Deactivate seconds (DA), and the confidence threshold value T are the parameters set in advance. The Detected Time (DT), the Undetected Time (UT), and the Trigger Time (TT) are timestamps, which are used to determine whether the activate seconds, the deactivate seconds, and the cooldown seconds are reached, and these timestamps are dynamically updated during processing; now represents a timestamp (that is, a present time) of a video frame at present; SET represents the updating of the timestamp, and OUT represents an output result of an event.

[0087] The values of the activate seconds AC and the deactivate seconds DA are, for example, 0 to 3 seconds, the value of the cooldown seconds CD is, for example, 1 to 10 seconds, the value of the confidence threshold value T is, for example, 0.5, and no limits are made thereto in the present disclosure.

[0088] In a possible implementation, S14 may include that: in a case where the confidence of the target object at a present time is greater than or equal to a confidence threshold value, an event detection result of the target object is determined to be that an abnormal event occurs.

[0089] For example, if there is no ROI (that is, the first region) in the video frame at a present time, or the confidence s that the target object is an abnormal posture is less than the confidence threshold value T, that is, a condition 211 (No ROI or s<T) is satisfied, then the detection state of the target object is in an idle state 21.

[0090] In a possible implementation, in a case where the target object is in the idle state 21, and the activate seconds AC is not set or the activate seconds AC is set as 0, if the confidence of the target object at a present time is greater than or equal to the confidence threshold value T, that is, a condition 212 (S>=T and AC=0) is satisfied, then the output result is START, that is, the event detection result of the target object is determined to be that an abnormal event occurs.

[0091] In this case, the detection state of the target object may be changed into a detected state 23, so as to realize state updating; both the trigger time TT and the undetected time UT are updated into a present time “now”, the trigger time TT and the undetected time UT are start to be calculated, so as to facilitate subsequent determination.

[0092] In this manner, it can be directly determined that an abnormal event occurs when the target object is an abnormal posture, so that the real-time performance of event detection is improved, thereby alarming can be performed in time and reducing the risk of safety accidents.

[0093] In a possible implementation, S14 may include the following operation.

[0094] In a case where the confidence of the target object at a present time is greater than or equal to a confidence threshold value and a first duration at which the confidence is greater than or equal to the confidence threshold value is greater than or equal to the activate seconds, the event detection result of the target object is determined to be that an abnormal event occurs. The activate seconds may be understood as an activate duration.

[0095] As illustrated in FIG. 2a, in a case where the target object is in an idle state 21 and the activate seconds AC is greater than 0, if the confidence of the target object at a present time is greater than or equal to the confidence threshold value T, a condition 213 (s>=T and AC>0) is satisfied, then no result is output, that is, the event detection result is that an abnormal event does not occur. In this case, the detection state of the target object

may be changed into a state under detecting 22, and the detected Time DT is updated into a present time “now”.

[0096] In a possible implementation, in a case where the target object is in the state under detecting 22, if the confidence of the target object at a present time is less than the confidence threshold value T, or there is no ROI in the video frame at a present time, that is, a condition 221 (No ROI or s<T) is satisfied, then no result is output, that is, the event detection result is that an abnormal event does not occur. In this case, the detection state is changed into an idle state 21.

[0097] In a possible implementation, in a case where the target object is in a state under detecting 22, if the confidence of the target object at a present time is greater than or equal to the confidence threshold value T, and a first duration (now-DT) at which the confidence is greater than or equal to the confidence threshold value T is less than the activate seconds AC, that is, a condition 222 (s>=T and now-DC<AC) is satisfied, then no result is output, that is, the event detection result is that an abnormal event does not occur. In this case, the detection state is kept in the state under detecting 22.

[0098] In a possible implementation, in a case where the detection state of the target object is the state under detecting 22, if the confidence of the target object at a present time is greater than or equal to the confidence threshold value T, and a first duration (now-DT) at which the confidence is greater than or equal to the confidence threshold value T is greater than or equal to the activate seconds AC, that is, a condition 223 (s>=T and now-DC>=AC) is satisfied, then an output result is START, that is, the event detection result of the target object is determined to be that an abnormal event occurs.

[0099] In this case, the detection state of the target object may be changed into a detected state 23, so as to realize state updating; both the trigger time TT and the undetected time UT are updated into a present time “now”, so as to facilitate subsequent determination.

[00100] In this manner, it can be determined that an abnormal event occurs when the target object is an abnormal posture and lasts for a certain duration, so that possible false detection during single-frame detection is reduced and the accuracy of event detection is

improved.

[00101] In a possible implementation, the method for the event detection according to the embodiments of the present disclosure further includes that: in a case where the event detection result of the target object is that the abnormal event occurs, first alarm information is sent.

[00102] That is to say, in a case where the output result is START, that is, the event detection result of the target object is that the abnormal event occurs, corresponding alarm information (called first alarm information) is sent. The alarm information may include a reminder that an abnormal event occurs, and may also include a region where the target object to which the abnormal event occurs is located, so as to facilitate locating by relevant personnel. No limits are made thereto in the present disclosure.

[00103] In a possible implementation, the first alarm information may be, for example, sent to a control apparatus of the escalator, so that the control apparatus stops the operation of the escalator. The alarm information may also be sent to a device (such as a terminal device and/or a display device) that can be seen by relevant personnel in charge of the operation of the escalator, so that the relevant personnel stop the operation of the escalator and go to the escalator for rescue. No limits are made to the contents of the first alarm information in the present disclosure.

[00104] In this manner, timely alarm of the abnormal event can be realized, so as to reduce the risk of safety accidents.

[00105] In a possible implementation, the method for the event detection according to the embodiments of the present disclosure further includes the following operation.

[00106] In a case where the event detection result of the target object is that an abnormal event occurs, the detection state of the target object is changed into a detected state.

[00107] As previously described, if the condition 223 (s>=T and now-DC>=AC) is satisfied, then an output result is START, that is, the event detection result of the target object is determined to be that an abnormal event occurs. In this case, the detection state

of the target object is changed into a detected state 23, so as to facilitate detection continuously in subsequent processing.

[00108] In a possible implementation, the method for the event detection further includes the following operations.

[00109] In a case where the detection state of the target object is the detected state, the confidence of the target object at a present time is greater than or equal to the confidence threshold value, and a second duration at which the confidence is greater than or equal to the confidence threshold value is greater than the cooldown seconds, the event detection result of the target object is determined to be that an abnormal event lasts.

[00110] In a case where the event detection result of the target object is that the abnormal event lasts, second alarm information is sent. The cooldown seconds may be understood as a cooldown duration.

[00111] For example, in a case where the detection state of the target object is the detected state 23, if the confidence of the target object at a present time is greater than or equal to the confidence threshold value T, and a first duration (now-TT) at which the confidence is greater than or equal to the confidence threshold value T is less than or equal to the cooldown seconds CD, then it can be considered as being within the cooldown seconds CD after the previous output, and no result is output.

[00112] In a possible implementation, if the confidence of the target object at a present time is greater than or equal to the confidence threshold value T, and the second duration (now-TT) that the confidence is greater than or equal to the confidence threshold value T is greater than the cooldown seconds CD, that is, a condition 231 ((s>=T, and if now-TT>CD) is satisfied, then an output result is ONGING, and the event detection result of the target object is determined to be that the abnormal event lasts.

[00113] In a possible implementation, if the confidence of the target object at a present time is greater than or equal to the confidence threshold value T, then the undetected time UT is updated into a present time “now”, that is, the timestamp of the undetected time UT is reset, so as to restart to calculate the undetected time. If the confidence of the target object at a present time is greater than or equal to the confidence threshold value T, and

the second duration now-TT that the confidence is greater than or equal to the confidence threshold value T is greater than the cooldown seconds CD, then the cooldown seconds CD may be updated into a present time “now”, that is, the timstamp of the cooldown seconds CD is reset, so as to restart to calculate the cooldown seconds.

[00114] In a possible implementation, in a case where the event detection result of the target object is that an abnormal event lasts, corresponding alarm information may be generated and sent (called second alarm information). The second alarm information may include a reminder that an abnormal event lasts, and may also include a region where the target object that the abnormal event lasts is located, so as to facilitate locating by relevant personnel. No limits are made thereto in the present disclosure.

[00115] In this manner, it can be determined that the abnormal event lasts in a case where the target object is an abnormal posture and the cooldown seconds is reached, so that repeated alarm is performed by taking the cooldown seconds as a time interval to repeatedly remind a device for monitoring the operation of an escalator or relevant personnel in charge of the operation of the escalator, thereby reducing cases that relevant personnel do not observe the alarm information, and reducing the risks of safety accidents.

[00116] In a possible implementation, the method may also include the following operations.

[00117] In a case where the detection state of the target object is the detected state, the confidence of the target object at a present time is less than the confidence threshold value, and a third duration at which the confidence is less than the confidence threshold value is greater than a deactivate seconds, the event detection result of the target object is determined to be that the abnormal event is ended. The deactivate seconds may be understood as a deactivate duration.

[00118] For example, in a case where the detection state of the target object is the detected state 23, if the confidence of the target object at a present time is less than the confidence threshold value T or there is no ROI, and the third duration (now-UT) that the confidence is less than the confidence threshold value is less than or equal to the

deactivate seconds, that is, the condition 232 ((No ROI or s<T) and now-UT<=DA) is satisfied, then it can be considered as being within the deactivate seconds DA after the previous output, and no result is output.

[00119] In a possible implementation, in a case where the detection state of the target object is the detected state 23, if the confidence of the target object at a present time is less than the confidence threshold value T or there is no ROI, and the third duration (now-TT) that the confidence is less than the confidence threshold value is greater than the deactivate seconds, that is, a condition 233 ((No ROI or s<T) and now-UT>DA) is satisfied, an output result is END, that is, the event detection result of the target object is that the abnormal event is ended.

[00120] In a possible implementation, in a case where the event detection result is that the abnormal event is ended, event end information may be generated and sent, so as to remind a device for monitoring the operation of an escalator or relevant personnel in charge of the operation of the escalator; or the event end information may also not be generated. No limits are made thereto in the present disclosure.

[00121] In a possible implementation, in a case that the event detection result of the target object is that the abnormal event is ended, the detection state of the target object may be changed into an idle state 21, so as to realize state updating; and the timestamp of the undetected time UT is updated nil (none), so as to facilitate subsequent re-determination.

[00122] In a possible implementation, if the detection state of the target object is in an idle state 21, and there is no ROI of the target object for a certain period, the state machine of the target object may be released, or the state machine is applied to other target objects. No limits are made thereto in the present disclosure.

[00123] In this manner, it can be determined that the abnormal event is ended when the target object is not in an abnormal posture and the deactivate seconds is reached, so as to complete the whole process of event detection.

[00124] FIG. 2b illustrates a schematic diagram of a process framework of a method for an event detection provided by the embodiments of the present disclosure. As

illustrated in FIG. 2b, the process framework includes a pedestrian detection and following module, a walking, running, and retrograding recognition module, and a squatting recognition module.

[00125] After the input video stream is decoded, the input video stream is input to the pedestrian detection and following module to obtain the first region of each pedestrian (expressed as the vertex coordinates of the upper left corner and the lower right corner of each pedestrian detection box, and a following ID of the corresponding pedestrian). The pedestrian detection box may be a rectangular box. Pedestrian detection is usually implemented by a deep neural network. A pedestrian following algorithm includes a deep learning-based algorithm, an IoU matching-based algorithm, etc. No limits are made to the use of which detection and following algorithm is in the present disclosure.

[00126] Posture recognition may be performed on a first region image corresponding to the first region through a trained posture recognition network to obtain a posture recognition result of the target object, that is, the posture recognition result is output through the posture recognition network. For example, the posture recognition network may be a squatting recognition module, so as to output a squatting event through the squatting recognition module.

[00127] The walking, running, and retrograding recognition module of the embodiments of the present disclosure may analyzes the motion speed of a pedestrian by using a video stream, and compare the motion speed of the pedestrian with the speed of an escalator, so as to determine whether the pedestrian has a behavior of walking, running, or retrograding, and then, for example, output an walking, running, and retrograding event.

[00128] According to the method for the event detection of the embodiments of the present disclosure, a video stream of a region where an escalator is located may be detected and followed, so as to determine a human body region of the same object. Posture recognition is performed on the image of this region to determine a posture of the object; event detection is performed according to the posture of the object, so as to determine whether a squatting event occurs; and finally, the region determined as the squatting event and a corresponding human body box are returned, and an alarm is

performed, which can improve the accuracy of event detection, and reduce the risk of safety accidents.

[00129] According to the method for the event detection of the embodiments of the present disclosure, the event detection can be realized by only using a monocular camera without other sensors (such as an infrared sensor), which can improve the efficiency of event detection.

[00130] The method for the event detection of the embodiments of the present disclosure can be applied to the field of safety protection, and applied to related alarming products for detecting a pedestrian dangerous behavior on an escalator, such as the escalator self-service system deployed in application scenarios such as large shopping malls and supermarkets, subway stations, and office buildings, to realize automatic detection and alarm of dangerous behaviors such a squatting event on the escalator, so as to take measures to slow down the escalator or stop the escalator and reduce the manual detection.

[00131] It can be understood that various method embodiments mentioned in the present disclosure may be combined to form combined embodiments without departing from principles and logics. For saving the space, elaborations are omitted in the present disclosure. It can be understood by those skilled in the art that, in the above-mentioned method of the specific implementations, a specific execution sequence of each step should be determined by functions and probable internal logic thereof.

[00132] In addition, the present disclosure further provides an event detection apparatus, an electronic device, a computer-readable storage medium, and a program. All of the above can be used to implement any event detection method provided by the present disclosure. Corresponding technical solutions, descriptions may refer to the corresponding records of the method section, which will not be elaborated herein.

[00133] FIG. 3 illustrates a block diagram of an apparatus for an event detection according to the embodiments of present the present disclosure. As illustrated in FIG. 3, the apparatus includes a video stream acquisition part 31, a region determination part 32, a posture recognition part 33, and an event detection part 34.

[00134] The video stream acquisition part 31 is configured to acquire a video stream of an escalator.

[00135] The region determination part 32 is configured to perform human body detection on the video stream, to determine a first region where a target object is located in the video steam.

[00136] The posture recognition part 33 is configured to perform posture recognition on a first region image corresponding to the first region to obtain a posture recognition result of the target object, where the posture recognition result includes the confidence that a posture of the target object is an abnormal posture.

[00137] An event detection part 34 is configured to determine an event detection result of the target object according to the posture recognition result of the target object.

[00138] In a possible implementation, the event detection part may further be configured to: in a case where the confidence of the target object at a present time is greater than or equal to a confidence threshold value, determine the event detection result of the target object to be that an abnormal event occurs.

[00139] In a possible implementation, the event detection part may further be configured to: in a case where the confidence of the target object at a present time is greater than or equal to a confidence threshold value and a first duration at which the confidence is greater than or equal to the confidence threshold value is greater than or equal to activate seconds, determine the event detection result of the target object to be that an abnormal event occurs.

[00140] In a possible implementation, the apparatus may further include: a first alarm part, configured to: in a case where the event detection result of the target object is that the abnormal event occurs, send first alarm information.

[00141] In a possible implementation, the apparatus further includes: a state changing part, configured to change the detection state of the target object into the detected state in a case where the event detection result of the target object is that the abnormal event occurs. Herein, the event detection part is further configured to: in a case where the

detection state of the target object is the detected state, the confidence of the target object at a present time is greater than or equal to the confidence threshold value, and a second duration at which the confidence is greater than or equal to the confidence threshold value is greater than cooldown seconds, determine the event detection result of the target object to be that an abnormal event lasts. The apparatus further includes: a first alarm part, configured to: in a case where the event detection result of the target object is that the abnormal event lasts, send second alarm information.

[00142] In a possible implementation, the event detection part is further configured to: in a case where the detection state of the target object is the detected state, the confidence of the target object at a present time is less than the confidence threshold value, and a third duration at which the confidence is less than the confidence threshold value is greater than a deactivate seconds, determine the event detection result of the target object to be that an abnormal event is ended.

[00143] In a possible implementation, the apparatus further includes: an expansion part and a clipping part. The expansion part is configured to: expand the first region according to a preset expansion coefficient to obtain a second region; and the clipping part is configured to: clip out an image of the second region from a video frame of the video stream, to obtain a first region image.

[00144] In a possible implementation, the posture recognition part is further configured to: perform the posture recognition on the first region image corresponding to the first region through a posture recognition network; and the apparatus further includes: a training part, configured to perform training according to a preset training set to obtain the posture recognition network, herein an object posture marked by a sample image in the training set includes at least one of: an abnormal posture, a normal posture, and other postures; and the other postures include the postures except the normal posture and the abnormal posture.

[00145] In a possible implementation, the abnormal event includes a squatting event.

[00146] In some embodiments, the functions or modules of the apparatus provided by the embodiments of the present disclosure can be used to execute the method described

in the above method embodiments, and its specific implementation may refer to the description of the above method embodiment. For simplicity, it will not be elaborated herein.

[00147] The embodiments of the present disclosure further provide a computer-readable storage medium, which stores a computer program instruction that when is executed by a processor, implements the above method. The computer-readable storage medium may be a nonvolatile computer readable storage medium.

[00148] The embodiments of the present disclosure further provide an electronic device, including a processor, and a memory configured to store a processor executable instruction. The processor is configured to call the instruction stored in the memory to implement the above method.

[00149] The embodiments of the present disclosure further provide a computer program product, including a computer-readable code. When the computer-readable code runs on a device, a processor in the device executes an instruction for implementing the method for the event detection provided by any one of the above embodiments.

[00150] The embodiments of the present disclosure further provide a computer program product configured to store a computer-readable instruction. When the instruction is executed, a computer executes the operations of the method for the event detection provided by any one of the abovementioned embodiments.

[00151] The electronic device may be provided as a terminal, a server, or other forms of devices.

[00152] FIG. 4 is a block diagram of an electronic device according to the embodiments of the present disclosure. For example, the electronic device 800 may be a mobile phone, a computer, a digital broadcast terminal, a message receiving and transmitting device, a gaming console, a tablet device, a medical device, fitness equipment, a personal digital assistant, or the like.

[00153] Referring to FIG. 4, the electronic device 800 may include one or more of the following components: a processing component 802, a memory 804, a power component

806, a multimedia component 808, an audio component 810, an Input/Output (I/O) interface 812, a sensor component 814, and a communication component 816.

[00154] The processing component 802 typically controls an overall operation of the electronic device 800, such as the operations associated with displays, telephone calls, data communications, camera operations, and recording operations. The processing component 802 may include one or more processors 820 to execute instructions to perform all or part of the steps in the above-mentioned method. Moreover, the processing component 802 may include one or more modules which facilitate interaction between the processing component 802 and the other components. For example, the processing component 802 may include a multimedia module to facilitate interaction between the multimedia component 808 and the processing component 802.

[00155] The memory 804 is configured to store various types of data to support the operation of the electronic device 800. Examples of such data include instructions, contact data, phonebook data, messages, pictures, video, etc, for any applications or methods operated on the electronic device 800. The memory 804 may be implemented by any type of volatile or non-volatile memory devices, or a combination thereof, such as a Static Random Access Memory (SRAM), an Electrically Erasable Programmable Read-Only Memory (EEPROM), an Erasable Programmable Read-Only Memory (EPROM), a Programmable Read-Only Memory (PROM), Read-Only Memory (ROM), a magnetic memory, a flash memory, and a magnetic disk or a compact disk.

[00156] The power component 806 provides power for various components of the electronic device 800. The power component 806 may include a power management system, one or more power supplies, and other components associated with generation, management, and distribution of power for the electronic device 800.

[00157] The multimedia component 808 includes a screen providing an output interface between the electronic device 800 and a user. In some embodiments, the screen may include a Liquid Crystal Display (LCD) and a Touch Panel (TP). If the screen includes the TP, the screen may be implemented as a touch screen to receive input signals from the user. The TP includes one or more touch sensors to sense touches, swipes, and gestures on the TP. The touch sensors may not only sense a boundary of a touch or swipe action

but also detect a duration and pressure associated with the touch or swipe action. In some embodiments, the multimedia component 808 includes a front camera and/or a rear camera. A front camera and/or a rear camera may receive external multimedia data when the electronic device 800 is in an operation mode, for example, a photographing mode or a video mode. Each of the front camera and the rear camera may be a fixed optical lens system or have focusing and optical zooming capabilities.

[00158] The audio component 810 is configured to output and/or input an audio signal. For example, the audio component 810 includes a Microphone (MIC). The MIC is configured to receive an external audio signal when the electronic device 800 is in the operation mode, such as a call mode, a recording mode, and a voice recognition mode. The received audio signal may further be stored in the memory 804 or sent through the communication component 816. In some embodiments, the audio component 810 further includes a speaker configured to output the audio signal.

[00159] The I/O interface 812 provides an interface between the processing component 802 and a peripheral interface module, and the peripheral interface module may be a keyboard, a click wheel, a button, etc. The button may include, but is not limited to: a home button, a volume button, a starting button, and a locking button.

[00160] The sensor component 814 includes one or more sensors configured to provide state assessment in various aspects for the electronic device 800. For example, the sensor component 814 may detect an on/off state of the electronic device 800 and relative positioning of components. For example, the components is a display and a small keyboard of the electronic device 800. The sensor component 814 may also detect a change in a position of the electronic device 800 or a component of the electronic device 800, presence or absence of contact between the user and the electronic device 800, orientation or acceleration/deceleration of the electronic device 800 and a change in temperature of the electronic device 800. The sensor component 814 may include a proximity sensor configured to detect presence of an object nearby without any physical contact. The sensor component 814 may also include a light sensor, such as a Complementary Metal Oxide Semiconductor (CMOS) or a Charge Coupled Device (CCD) image sensor, configured for use in an imaging application. In some

embodiments, the sensor component 814 may also include an acceleration sensor, a gyroscope sensor, a magnetic sensor, a pressure sensor, or a temperature sensor.

[00161] The communication component 816 is configured to facilitate wired or wireless communication between the electronic device 800 and another device. The electronic device 800 may access a communication-standard-based wireless network, such as a Wireless Fidelity (Wi-Fi) network, the 2nd-Generation (2G) wireless communication technology, or the 3rd-Generation (3G) wireless communication technology, or a combination thereof. In one exemplary embodiment, the communication component 816 receives a broadcast signal or broadcast associated information from an external broadcast management system through a broadcast channel. In an exemplary embodiment, the communication component 816 further includes a Near Field Communication (NFC) module to facilitate short-range communication. For example, the NFC module may be implemented based on a Radio Frequency Identification (RFID) technology, an Infrared Data Association (IrDA) technology, an Ultra Wide Band (UWB) technology, a Bluetooth (BT) technology, and another technology.

[00162] In an exemplary embodiment, the electronic device 800 may be implemented by one or more Application Specific Integrated Circuits (ASICs), Digital Signal Processors (DSPs), Digital Signal Processing Devices (DSPDs), Programmable Logic Devices (PLDs), Field Programmable Gate Arrays (FPGAs), controllers, micro-controllers, microprocessors, or other electronic elements, and is configured to execute the above-mentioned method.

[00163] In the exemplary embodiment, a nonvolatile computer-readable storage medium is also provided, for example, a memory 804 including a computer program instruction. The computer program instruction may be executed by a processor 820 of an electronic device 800 to implement the above-mentioned method.

[00164] FIG. 5 illustrates a block diagram of another electronic device according to the embodiments of the present disclosure. For example, the electronic device 1900 may be provided as a server. Referring to FIG. 5, the electronic device 1900 includes a processing component 1922, which further includes one or more processors, and a memory resource represented by a memory 1932, configured to store an instruction

executable for the processing component 1922, for example, an application. The application stored in the memory 1932 may include one or more modules each of which corresponds to a set of instructions. In addition, the processing component 1922 is configured to execute instructions to execute the above-mentioned method.

[00165] The electronic device 1900 may further include a power component 1926 configured to execute power management of the electronic device 1900, a wired or wireless network interface 1950 configured to connect the electronic device 1900 to a network, and an I/O interface 1958. The electronic device 1900 may operate an operating system stored in the memory 1932, for example, a Microsoft server operating system (Windows ServerTM), a graphical user interface based operating system (Mac OS XTM) introduced by Apple, a multi-user multi-process computer operating system (UnixTM), a free and open source code Unix-like operating system (LinuxTM), an open source code Unix-like operating system (FreeBSDTM), or the like.

[00166] In the illustrative embodiment, a nonvolatile computer-readable storage medium is also provided, for example, a memory 1932 including a computer program instruction. The computer program instruction may be executed by a processing component 1922 of an electronic device 1900 to implement the abovementioned method.

[00167] The present disclosure may be systems, methods, and/or computer program products. The computer program products may include a computer-readable storage medium, in which a computer-readable program instruction configured to enable a processor to implement each aspect of the present disclosure is carried.

[00168] The computer-readable storage medium may be a tangible device capable of retaining and storing an instruction used by an instruction execution device. The computer-readable storage medium may be, (but not limited to), an electric storage device, a magnetic storage device, an optical storage device, an electromagnetic storage device, a semiconductor storage device, or any appropriate combination thereof. More specific examples (non-exhaustive list) of the computer-readable storage medium include a portable computer disk, a hard disk, a RAM, a ROM, an EPROM (or a flash memory), an SRAM, a Compact Disc Read-Only Memory (CD-ROM), a Digital Video Disk (DVD), a memory stick, a floppy disk, a mechanical encoding device, for example, a

punched card or in-slot raised structure with an instruction stored therein, and any appropriate combination thereof. Herein, the computer-readable storage medium is not explained as a transient signal, for example, a radio wave or another freely propagated electromagnetic wave, an electromagnetic wave propagated through a wave guide or another transmission medium (for example, a light pulse propagated through an optical fiber cable) or an electric signal transmitted through an electric wire.

[00169] The computer-readable program instruction described here may be downloaded from the computer-readable storage medium to each computing/processing device or downloaded to an external computer or an external storage device through a network such as the Internet, a Local Region Network (LAN), a Wide Region Network (WAN) and/or a wireless network. The network may include a copper transmission cable, optical fiber transmission, wireless transmission, a router, a firewall, a switch, a gateway computer, and/or an edge server. A network adapter card or network interface in each computing/processing device receives the computer-readable program instruction from the network and forwards the computer-readable program instruction for storage in the computer-readable storage medium in each computing/processing device.

[00170] The computer program instruction configured to execute the operations of the present disclosure may be an assembly instruction, an Instruction Set Architecture (ISA) instruction, a machine instruction, a machine related instruction, a microcode, a firmware instruction, state setting data or a source code or target code edited by one or any combination of more programming languages, the programming language including an object-oriented programming language such as Smalltalk and C++ and a conventional procedural programming language such as "C" language or a similar programming language. The computer-readable program instruction may be completely executed in a computer of a user or partially executed in the computer of the user, executed as an independent software package, executed partially in the computer of the user and partially in a remote computer, or executed completely in the remote server or a server. Under the condition that the remote computer is involved, the remote computer may be concatenated to the computer of the user through any type of network including an LAN or a WAN, or may be concatenated to an external computer (for example, concatenated by an Internet service provider through the Internet). In some embodiments, an electronic

circuit such as a programmable logic circuit, an FPGA, or a Programmable Logic Array (PLA) may be customized by use of state personal information of a computer-readable program instruction, and the electronic circuit may execute the computer-readable program instruction, thereby implementing each aspect of the present disclosure.

[00171] Herein, each aspect of the present disclosure is described with reference to flowcharts and/or block diagrams of the method, device (system) and computer program product according to the embodiments of the present disclosure. It is to be understood that each block in the flowcharts and/or the block diagrams and a combination of each block in the flowcharts and/or the block diagrams may be implemented by computer-readable program instructions.

[00172] These computer-readable program instructions may be provided for a universal computer, a dedicated computer or a processor of another programmable data processing device, thereby generating a machine to further generate a device that realizes a function/action specified in one or more blocks in the flowcharts and/or the block diagrams when the instructions are executed through the computer or the processor of the other programmable data processing device. These computer-readable program instructions may also be stored in a computer-readable storage medium, and through these instructions, the computer, the programmable data processing device and/or another device may work in a specific mode, so that the computer-readable medium including the instructions includes a product including instructions for implementing each aspect of the function/action specified in one or more blocks in the flowcharts and/or the block diagrams.

[00173] These computer program instructions can also be loaded to a computer or another programmable data processing device, so that a series of operating steps are performed on the computer or the another programmable data processing device to produce a computer-implemented process, and therefore instructions executed on the computer or the another programmable data processing device provide steps for implementing the function/action specified in one or more blocks in the flowcharts and/or the block diagrams.

[00174] The flowcharts and block diagrams in the drawings illustrate probably

implemented system architectures, functions, and operations of the system, method, and computer program product according to multiple embodiments of the present disclosure. On this aspect, each block in the flowcharts or the block diagrams may represent part of a module, a program segment or an instruction, and part of the module, the program segment or the instruction includes one or more executable instructions configured to realize a specified logical function. In some alternative implementations, the functions marked in the blocks may also be realized in a sequence different from those marked in the drawings. For example, two continuous blocks may actually be executed substantially concurrently and may also be executed in a reverse sequence sometimes, which is determined by the involved functions. It is further to be noted that each block in the block diagrams and/or the flowcharts and a combination of the blocks in the block diagrams and/or the flowcharts may be implemented by a dedicated hardware-based system configured to execute a specified function or operation or may be implemented by a combination of a special hardware and a computer instruction.

[00175] The computer program product may be specifically realized by means of hardware, software or a combination thereof. In an optional embodiment, the computer program product is specifically embodied as a computer storage medium, and in another optional embodiment, the computer program product is specifically embodied as software products, such as a Software Development Kit (SDK).

[00176] Each embodiment of the present disclosure has been described above. The above descriptions are exemplary, non-exhaustive and also not limited to each disclosed embodiment. Many modifications and variations are apparent to those of ordinary skill in the art without departing from the scope and spirit of each described embodiment of the present disclosure. The terms used herein are selected to explain the principle and practical application of each embodiment or improvements in the technologies in the market best or enable others of ordinary skill in the art to understand each embodiment disclosed herein.

INDUSTRIAL APPLICABILITY

[00177] The embodiments of the present disclosure disclose a method for an event detection and apparatus, an electronic device, a storage medium, and a product program. The method for an event detection includes that: a video stream of an escalator is acquired; human body detection is performed on the video stream, to determine a first region where a target object is located in the video steam; posture recognition is performed on a first region image corresponding to the first region to obtain a posture recognition result of the target object, herein the posture recognition result includes the confidence that a posture of the target object is an abnormal posture; and an event detection result of the target object is determined according to the posture recognition result of the target object. By the method, the accuracy of event detection can be improved, so as to reduce the probability of accidents and reduce the cost of manual detection.

CLAIMS

1. A method for an event detection, comprising:

acquiring a video stream of an escalator;

performing a human body detection on the video stream to determine a first region where a target object is located in the video steam;

performing a posture recognition on a first region image corresponding to the first region to obtain a posture recognition result of the target object, wherein the posture recognition result comprises a confidence that a posture of the target object is an abnormal posture; and

determining an event detection result of the target object according to the posture recognition result of the target object.

2. The method of claim 1, wherein the determining the event detection result of the target object according to the posture recognition result of the target object comprising:

in a case where a confidence of the target object at a present time is greater than or equal to a confidence threshold value, determining the event detection result of the target object to be that an abnormal event occurs.

3. The method of claim 1, wherein the determining the event detection result of the target object according to the posture recognition result of the target object comprising:

in a case where a confidence of the target object at a present time is greater than or equal to a confidence threshold value and a first duration at which the confidence is greater than or equal to the confidence threshold value is greater than or equal to activate seconds, determining the event detection result of the target object to be that an abnormal event occurs.

4. The method of any one of claims 1 to 3, further comprising:

in a case where the event detection result of the target object is that the abnormal event occurs, sending first alarm information.

5. The method of any one of claims 1 to 4, further comprising:

in a case where the event detection result of the target object is that an abnormal event occurs, changing a detection state of the target object into a detected state; and

in a case where the detection state of the target object is the detected state, the confidence of the target object at a present time is greater than or equal to the confidence threshold value, and a second duration at which the confidence is greater than or equal to the confidence threshold value is greater than cooldown seconds, determining the event detection result of the target object to be that an abnormal event lasts; and

in a case where the event detection result of the target object is that the abnormal event lasts, sending second alarm information.

6. The method of claim 5, further comprising:

in a case where the detection state of the target object is the detected state, a confidence of the target object at a present time is less than the confidence threshold value, and a third duration at which the confidence is less than the confidence threshold value is greater than deactivate seconds, determining the event detection result of the target object to be that the abnormal event is ended.

7. The method of any one of claims 1 to 6, wherein determining the first region image corresponding to the first region comprises:

expanding the first region according to a preset expansion coefficient to obtain a second region; and

clipping out an image of the second region from video frames of the video stream, to obtain the first region image.

8. The method of any one of claims 1 to 7, wherein the performing the posture recognition on the first region image corresponding to the first region comprises: performing the posture recognition on the first region image corresponding to the first region through a posture recognition network;

the method further comprises: performing training according to a preset training set to obtain the posture recognition network, wherein an object posture marked by a sample image in the training set comprises at least one of: an abnormal posture, a normal posture, or other postures; and the other postures comprise postures except the normal posture and the abnormal posture.

9. The method of any one of claims 2 to 6, wherein the abnormal event comprises a squatting event.

10. An apparatus for an event detection, comprising:

a video stream acquisition part, configured to acquire a video stream of an escalator;

a region determination part, configured to perform a human body detection on the video stream to determine a first region where a target object is located in the video steam;

a posture recognition part, configured to perform a posture recognition on a first region image corresponding to the first region to obtain a posture recognition result of the target object, wherein the posture recognition result comprises a confidence that a posture of the target object is an abnormal posture; and

an event detection part, configured to determine an event detection result of the target object according to the posture recognition result of the target object.

11. The apparatus of claim 10, wherein the event detection part is further configured to: in a case where a confidence of the target object at a present time is greater than or equal to a confidence threshold value, determine the event detection result of the target object to be that an abnormal event occurs.

12. The apparatus of claim 10, wherein the event detection part is further configured to: in a case where a confidence of the target object at a present time is greater than or equal to a confidence threshold value and a first duration at which the confidence is greater than or equal to the confidence threshold value is greater than or equal to activate seconds, determine the event detection result of the target object to be that an abnormal event occurs.

13. The apparatus of any one of claims 10 to 12, further comprising: a first alarm part, configured to: in a case where the event detection result of the target object is that the abnormal event occurs, send first alarm information.

14. The apparatus of any one of claims 10 to 13, further comprising: a state changing part, configured to: in a case where the event detection result of the target object is that an abnormal event occurs, change a detection state of the target object into a detected state, wherein

the event detection part is further configured to: in a case where the detection state of the target object is the detected state, the confidence of the target object at a present time is greater than or equal to the confidence threshold value, and a second duration at which the confidence is greater than or equal to the confidence threshold value is greater than cooldown seconds, determine the event detection result of the target object to be that an abnormal event

lasts; and

the apparatus further comprises: a first alarm part, configured to: in a case where the event detection result of the target object is that the abnormal event lasts, send second alarm information.

15. The apparatus of claim 14, wherein the event detection part is further configured to: in a case where the detection state of the target object is the detected state, a confidence of the target object at a present time is less than the confidence threshold value, and a third duration at which the confidence is less than the confidence threshold value is greater than deactivate seconds, determine the event detection result of the target object to be that an abnormal event is ended.

16. The apparatus of any one of claims 10 to 15, further comprising: an expansion part and a clipping part, wherein

the expansion part is configured to: expand the first region according to a preset expansion coefficient to obtain a second region; and

the clipping part is configured to: clip out an image of the second region from a video frame of the video stream, to obtain the first region image.

17. The apparatus of any one of claims 10 to 16, wherein the posture recognition part is further configured to: perform the posture recognition on the first region image corresponding to the first region through a posture recognition network; and

the apparatus further comprises: a training part, configured to: perform training according to a preset training set to obtain the posture recognition network, wherein an object posture marked by a sample image in the training set comprises at least one of the following: an abnormal posture, a normal posture, or other postures; and the other postures comprise

postures except the normal posture and the abnormal posture.

18. The apparatus of any one of claims 11 to 15, wherein the abnormal event comprises a squatting event.

19. An electronic device, comprising:

a processor; and

a memory, configured to store a processor executable instruction, wherein

the processor is configured to call the instruction stored in the memory to execute the method of any one of claims 1 to 9.

20. A computer-readable storage medium having stored thereon a computer program instruction that when executed by a processor, implements the method of any one of claims 1 to 9.

21. A computer program product, comprising a computer-readable code, wherein when the computer-readable code runs on a device, a processor in the device executes an instruction for implementing the method for event detection of any one of claims 1 to 9.