

IMAGE DETECTION METHOD AND APPARATUS, ELECTRONIC DEVICE, AND STORAGE MEDIUM

CROSS-REFERENCE TO RELATED APPLICATION(S)

[0001] The present application claims priority to Chinese Patent Application No. 202311108453.1, filed on August 30, 2023 and entitled "IMAGE DETECTION METHOD AND APPARATUS, ELECTRONIC DEVICE, AND STORAGE MEDIUM", which is hereby incorporated by reference in its entirety.

FIELD

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

BACKGROUND

[0003] At present, defects of a virtual portrait in an image generated by artificial intelligence are a widespread problem, for example, an abnormal number of fingers, an abnormal body proportion, unnatural body joint connection, and the like, which seriously affect the quality of the image generated by artificial intelligence. In this regard, there is still a lack of an effective discrimination method for discriminating whether the virtual portrait in the image generated by artificial intelligence has a defect.

SUMMARY

[0004] The part of the summary is provided so as to introduce the concept in a brief form, and these concepts will be described in detail in the subsequent part of detailed description. The part of the summary is not intended to identify the key features or necessary features of the claimed technical solution, nor is it intended to limit the scope of the claimed technical solution.

[0005] In a first aspect, according to one or more embodiments of the present disclosure, an image detection method is provided, including:

[0006] determining whether a target object exists in a to-be-detected image;

[0007] extracting, in response to the target object existing in the to-be-detected image, structural key point information of the target object in the to-be-detected image; and

[0008] determining, based on the structural key point information and the to-be-detected image, a classification of the target object in the to-be-detected image.

[0009] In a second aspect, according to one or more embodiments of the present disclosure, an image detection apparatus is provided, including:

[0010] a determining unit, configured to determine whether a target object exists in a to-be-detected image;

[0011] an extracting unit, configured to extract, in response to the target object existing in the to-be-detected

image, structural key point information of the target object in the to-be-detected image; and

[0012] a classifying unit, configured to determine, based on the structural key point information and the to-be-detected image, a classification of the target object in the to-be-detected image.

5 [0013] In a third aspect, according to one or more embodiments of the present disclosure, an electronic device is provided, including: at least one memory and at least one processor; where the memory is configured to store program code, and the processor is configured to call the program code stored in the memory to cause the electronic device to perform the method provided according to one or more embodiments of the present disclosure.

10 [0014] In a fourth aspect, according to one or more embodiments of the present disclosure, a non-transitory computer storage medium is provided, where the non-transitory computer storage medium stores program code, and the program code, when executed by a computer device, causes the computer device to perform the method provided according to one or more embodiments of the present disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

15 [0015] The above-mentioned and other features, advantages, and aspects of various embodiments of the present disclosure will become more apparent in combination with the drawings and with reference to the following specific embodiments. Throughout the drawings, the same or similar reference numerals represent the same or similar elements. It should be understood that the drawings are schematic, and the components and elements are not necessarily drawn to scale.

20 [0016] Fig. 1 is a flowchart of an image detection method provided according to an embodiment of the present disclosure;

[0017] Fig. 2 is a flowchart of an image detection method provided according to another embodiment of the present disclosure;

[0018] Fig. 3 is a flowchart of an image detection method provided according to a yet another embodiment of the present disclosure;

25 [0019] Fig. 4 is a schematic structural diagram of an image detection apparatus provided according to an embodiment of the present disclosure; and

[0020] Fig. 5 is a schematic structural diagram of an electronic device provided according to an embodiment of the present disclosure.

DETAILED DESCRIPTION OF EMBODIMENTS

30 [0021] Embodiments of the present disclosure will now be described in detail with reference to the drawings. Although some embodiments of the present disclosure are shown in the drawings, it should be understood that the present disclosure may be implemented in various forms, and should not be construed as limited to the embodiments set forth herein. Rather, these embodiments are provided for a more thorough and complete

understanding of the present disclosure. It should be understood that the drawings and embodiments of the present disclosure are only for illustrative purposes, and are not intended to limit the protection scope of the present disclosure.

[0022] It should be understood that the steps described in the implementations of the present disclosure may be performed in different orders and/or in parallel. Furthermore, the implementations may include additional steps and/or omit performing the illustrated steps. The scope of the present disclosure is not limited in this respect.

[0023] As used herein, the term "include/comprise" and variations thereof are open-ended inclusions, that is, "include/comprise but not limited to". The term "based on" is "at least partially based on". The term "one embodiment" means "at least one embodiment"; the term "another embodiment" means "at least one another embodiment"; the term "some embodiments" means "at least some embodiments". The term "responsive to" and related terms mean that a signal or event is affected by another signal or event to some extent, but not necessarily completely or directly affected. If event x occurs "responsive to" event y, then x may be directly or indirectly responsive to y. For example, the occurrence of y may ultimately lead to the occurrence of x, but there may be other intermediate events and/or conditions. In other cases, y may not necessarily lead to the occurrence of x, and x may occur even if y has not occurred. Furthermore, the term "responsive to" may also mean "at least partially responsive to".

[0024] The term "determine" broadly covers a wide variety of actions, which may include acquiring, calculating, computing, processing, deriving, researching, looking up (for example, looking up in a table, database, or other data structure), ascertaining, and similar actions, and may also include receiving (for example, receiving information), accessing (for example, accessing data in a memory), and similar actions, as well as parsing, selecting, choosing, establishing, and similar actions, and so on. Relevant definitions of other terms will be given in the description below. Relevant definitions of other terms will be given in the description below. Relevant definitions of other terms will be given in the description below.

[0025] It may be understood that the data involved in the technical solution (including but not limited to the data itself, acquisition or use of the data) shall comply with the provisions of relevant laws and regulations.

[0026] It may be understood that before using the technical solutions disclosed in various embodiments of the present disclosure, the user shall be informed of the type, range of use, use scenarios, etc., of personal information involved in the present disclosure and obtain the authorization of the user in an appropriate manner in accordance with relevant laws and regulations. For example, in response to receiving an active request from a user, prompt information is sent to the user to clearly prompt the user that the requested operation will require access to and use of personal information of the user, so that the user may independently choose whether to provide the personal information to software or hardware, such as an electronic device, an application, a server, or a storage medium, that performs the operations of the technical solutions of the present disclosure based on the prompt information.

[0027] As an optional but non-limiting implementation, in response to receiving the active request from the user, the prompt information may be sent to the user in the form of, for example, a pop-up window, and the

prompt information may be presented in the pop-up window in text. In addition, the pop-up window may also include a selection control for the user to choose whether to "agree" or "disagree" to provide the personal information to the electronic device.

[0028] It may be understood that the images generated by the method provided according to various embodiments of the present disclosure shall be processed in accordance with the provisions of relevant laws and regulations. For example, technical measures shall be taken to add an identification that does not affect the user's use according to the provisions, or a significant identification shall be made in a reasonable position and area according to the provisions to prompt the public with the deep synthesis situation.

[0029] It may be understood that the above process of notifying and obtaining user authorization and the processing of the image are only illustrative, and do not limit the implementations of the present disclosure. Other methods that satisfy relevant laws and regulations may also be applied to the implementations of the present disclosure.

[0030] It should be noted that concepts such as "first" and "second" mentioned in the present disclosure are only used to distinguish different apparatuses, modules, or units, and are not used to limit the order or interdependence of the functions performed by these apparatuses, modules, or units.

[0031] It should be noted that the modifiers of "one" and "multiple" mentioned in the present disclosure are illustrative and not restrictive, and those skilled in the art should understand that, unless clearly indicated in the context, they should be understood as "one or more".

[0032] For the purposes of the present disclosure, the phrase "A and/or B" means (A), (B), or (A and B).

[0033] The names of messages or information exchanged between multiple apparatuses in the implementations of the present disclosure are only used for illustrative purposes, and are not used to limit the scope of these messages or information.

[0034] At present, there is a lack of targeted solutions for discriminating a limb defect of a virtual portrait in content generated by artificial intelligence, and related technologies generally regard it as a classification problem, that is, a classification model is trained by using a sample image and a corresponding manually labeled label indicating whether a limb has a defect, and the classification model is used to discriminate an image. However, as it is difficult to discriminate a limb defect itself, and normal states of a limb corresponding to different viewing angles and different body movements are different, it requires that the classification model be capable of having an in-depth understanding of a state of the limb. However, the solution provided by the related technology to train the model only based on the sample and the manually labeled label easily leads to the classification model being prone to overfitting to other features that are easy to fit in the image (such as an overall complexity of the image) in the training process, so that the trained classification model does not really understand the limb defect. In addition, since there will inevitably be mislabeling in manual labeling of sample data, the existence of such data noise further increases the probability of the model fitting to other features during the training process.

[0035] In view of the above problems, the present disclosure provides an image detection method. By classifying a target object based on a to-be-detected image and structural key point information of the target object extracted from the to-be-detected image, a classification model can be caused to pay more attention to the target object during training and discrimination processes, which can avoid the model from overfitting to other features, thereby enhancing robustness of the model to data labeling noise and improving accuracy of classifying the target object.

[0036] Referring to Fig. 1, which shows a flowchart of an image detection method 100 provided according to an embodiment of the present disclosure. The method 100 includes steps S110 to S130.

[0037] Step S110: determining whether a target object exists in a to-be-detected image;

[0038] Step S120: extracting, in response to the target object existing in the to-be-detected image, structural key point information of the target object in the to-be-detected image;

[0039] Step S130: determining, based on the structural key point information and the to-be-detected image, a classification of the target object in the to-be-detected image.

[0040] Structural key points of the target object are used to determine a structural form of the target object.

In a specific implementation, the target object may include a virtual portrait, and the virtual portrait may include, for example, a portrait included in content generated by artificial intelligence; accordingly, the structural key point information of the target object may include virtual human body key point information, for example, may include information of some or all key points of a nose, eyes, ears, shoulders, elbows, wrists, hips, knees, and ankles.

In some embodiments, step S130 may be used to determine whether a virtual portrait defect exists in the to-be-detected image. A classification scenario of determining whether the virtual portrait has a defect is taken as an example below for description.

[0042] Referring to Fig. 2, which shows a method 200 for determining whether a virtual portrait in an image has a defect, including steps S201 to S204.

[0043] In step 201, a to-be-detected image is acquired. Exemplarily, the to-be-detected image may include content generated by artificial intelligence or another virtual image.

[0044] In step 202, human body detection is performed on the to-be-detected image to determine whether a virtual portrait exists. In some embodiments, whether a virtual portrait exists in the to-be-detected image may be determined by a target detection algorithm. A task of target detection is to find all targets (objects) of interest in the image, and determine their categories and positions. The target detection algorithm may employ a target detection algorithm such as a region-based convolutional neural network, a single shot multibox detector algorithm, a retina net-based target detection algorithm, or the like, but the present disclosure is not limited thereto.

[0045] If it is determined that no virtual portrait exists, it may be directly determined that no virtual portrait

defect exists in the to-be-detected image, and if it is determined that there a virtual portrait exists, step 203 is executed.

[0046] In step 203, virtual human body key point information of the to-be-detected image is extracted. In some embodiments, the virtual human body key point information of the to-be-detected image may be extracted by a human body key point estimation algorithm, and the human body key point estimation algorithm may employ a human body posture estimation method based on, for example, a high-resolution network, a convolutional pose machine, a stacked hourglass network, and a cascaded pyramid network, but the present disclosure is not limited thereto. In a specific implementation, eight coordinates corresponding to both hands, both elbows, both feet, and both legs of the virtual portrait in the to-be-detected image may be extracted as the virtual human body key point information.

[0047] In step 204, the virtual human body key point information and the to-be-detected image may be used as inputs of a trained classification model together, to determine whether the virtual portrait in the to-be-detected image has a defect. In a specific implementation, the classification model is trained to be capable of determining a confidence that the virtual portrait in the image has a defect based on the input image and the virtual human body key point information. Exemplarily, a sample image and virtual human body key point information extracted based on the sample image may be used as inputs of the model, and a manually labeled label corresponding to the sample image (for example, regarding whether the sample image has a virtual portrait defect) may be used as a desired output to train the classification model.

[0048] The manually labeled label may be regarding whether the sample image has a virtual portrait defect, or may refer to another classification type according to another application requirement, which is not limited in the present disclosure. In a specific implementation, a positive label may be 1, and a negative label may be 0.

[0049] It should be noted that the image detection method provided according to the present disclosure may be used for classification scenarios other than determining whether the virtual portrait has a defect, and may also be used to discriminate a classification of a target object with a specific structure, such as an animal, a robot, and an avatar, which is not limited in the present disclosure.

[0050] According to one or more embodiments of the present disclosure, by classifying the target object based on the to-be-detected image and the structural key point information of the target object extracted from the to-be-detected image, the classification model can be caused to pay more attention to the target object during the training and discrimination processes, and the model may be prevented from overfitting to other features, thereby enhancing the robustness of the model to data labeling noise and improving the accuracy of classifying the target object.

[0051] In some embodiments, step S130 includes:

[0052] Step A1: generating a first target feature based on the to-be-detected image, where the first target feature includes a feature associated with the target object;

[0053] Step A2: performing feature conversion on the structural key point information to obtain a second

target feature;

[0054] Step A3: generating a third target feature based on the first target feature and the second target feature;

[0055] Step A4: determining the classification of the target object in the to-be-detected image based on the third target feature.

5 **[0056]** In some embodiments, a number of channels corresponding to the second target feature is the same as a number of channels corresponding to the first target feature, and by making the number of channels corresponding to the second target feature the same as the number of channels corresponding to the first target feature, it can be convenient to subsequently concatenate the first target feature and the second target feature to obtain the third target feature.

10 **[0057]** In some embodiments, before the classification of the target object in the to-be-detected image is determined based on the third target feature, semantic unit fusion processing is performed on the third target feature, to cause a semantic unit in the third target feature to be capable of fusing information of other semantic units in the third target feature, that is, the semantic units are capable of fusing each other. The semantic unit is a unit with semantic information, such as a token (token). Further, subsequently, the classification of the target
15 object in the to-be-detected image may be determined based on the third target feature after the semantic unit fusion processing.

[0058] In the present embodiment, by performing the semantic unit fusion processing on the third target feature, the feature from the to-be-detected image and the feature from the structural key point information may be further fused to a certain extent at the level of the semantic unit, thereby further improving the attention of
20 the classification model to the area of the target object part in the image.

[0059] In some specific implementations, a self-attention mechanism in a transformer may be used to determine a fusion weight based on a correlation between the semantic units to implement the semantic unit fusion, or the semantic unit fusion may be implemented by expanding a size of a convolution kernel in a convolutional neural network, or the semantic unit fusion may be implemented through a full connection
25 between tokens in a multilayer perceptron, but the present disclosure is not limited thereto.

[0060] It should be noted that the semantic unit fusion processing in the present embodiment may be global semantic unit fusion or fusion for some semantic units, which is not limited in the present disclosure.

[0061] In some embodiments, step A4 includes:

[0062] Step a41: inputting the third target feature into a classification network to obtain a confidence that the
30 target object has a defect and/or a confidence that the target object has no defect;

[0063] Step a42: obtaining the category of the target object based on the confidence that the target object has a defect and/or the confidence that the target object has no defect.

[0064] In some embodiments, before the third target feature is input into the classification network, other processing may be performed on the third target feature, such as semantic unit fusion processing, mean

processing for a preset dimension, etc., which is not limited in the present disclosure.

[0065] Referring to Fig. 3, a classification scenario of determining whether a virtual portrait has a defect is taken as an example below for description.

[0066] First step, a normalized tensor T_i may be obtained based on virtual human body key point information of a to-be-detected image, which may be represented as (B, N, C) , where B represents a batch size of training samples, that is, the number of data (samples) delivered to a program at a time for training; N represents a number of semantic units; C represents a number of channels. Exemplarily, assuming that coordinates (x, y) of eight human body key points in two directions x and y are acquired for each to-be-detected image, N may be 8, C may be 2.

[0067] Second step, the tensor T_i is input into a feature extraction network for feature conversion to obtain a tensor T_f , which may be represented as (B, N_1, C_1) . In a specific implementation, the feature extraction network may be a transformer, which is a neural network model based on a self-attention mechanism, and is used to process sequential data.

[0068] Third step, a to-be-detected image (for example, an RGB image) is input into a backbone network for feature extraction to obtain a tensor T_m , which may be represented as (B, N_2, C_1) . A well-trained backbone network is expected to be used to extract features associated with a classification task from the to-be-detected image, and in a classification scenario regarding a virtual portrait, it is used to extract features regarding a structure and a number of virtual portraits. Certainly, in actual applications, the features extracted by the backbone network may also include features that are not expected to be extracted, such as features of buildings or other objects that have nothing to do with the virtual portrait in the image. In a specific implementation, the backbone network may employ a vision transformer or other image feature extraction networks, but the present disclosure is not limited thereto.

[0069] Fourth step, the tensor T_f and the tensor T_m are concatenated in a dimension of the semantic unit to obtain a tensor T_c , which may be represented as (B, N_1+N_2, C_1) . Exemplarily, the concatenating operation may be implemented by a concatenating function (Concat), which may concatenate features in a specified dimension.

[0070] Fifth step, the tensor T_c is input into a feature fusion network to obtain a tensor T_o , which may be represented as (B, N_3, C_2) . The feature fusion network may be used to perform semantic unit fusion processing on the input features, so that each of the semantic units in the features is capable of performing semantic unit fusion based on a correlation between it and other semantic units, thereby improving the degree of information interaction. Exemplarily, the feature fusion network may be a transformer.

[0071] Sixth step, a mean operation is performed on the tensor T_o in the dimension of the semantic unit (that is, " N " in (B, N, C)), and the result is input into a classification head network to obtain a final classification of whether the virtual portrait has a defect, and the classification result may be represented as a confidence of "the virtual portrait has a defect" and/or a confidence of "the virtual portrait has no defect". Exemplarily, the classification head network may be a fully connected layer. In this step, by performing the mean operation on

the tensor T_o in the dimension of the semantic unit, it is easy to process it using the classification head network.

[0072] In the present embodiment, the classification network includes a feature extraction network, a backbone network, a feature fusion network, and a classification head network. Exemplarily, during the training of the classification network, a sample image and corresponding virtual human body key point information may be input into the backbone network and the feature extraction network respectively, and a prediction result is obtained after passing through the feature fusion network and the classification head network in turn, and parameters of the feature extraction network, the backbone network, the feature fusion network, and the classification head network are adjusted based on a loss (for example, cross-entropy loss) between the prediction result and a label corresponding to the sample image, so that a loss function converges. In a specific implementation, the backbone network may be initialized through pre-trained parameters to accelerate convergence.

[0073] Correspondingly, referring to Fig. 4, an image detection apparatus 400 is provided according to an embodiment of the present disclosure, including:

[0074] a determining unit 401, configured to determine whether a target object exists in a to-be-detected image;

[0075] an extracting unit 402, configured to extract, in response to the target object existing in the to-be-detected image, structural key point information of the target object in the to-be-detected image;

[0076] a classifying unit 403, configured to determine, based on the structural key point information and the to-be-detected image, a classification of the target object in the to-be-detected image.

[0077] In some embodiments, the target object includes a virtual portrait, and the structural key point information of the target object includes virtual human body key point information.

[0078] In some embodiments, the classifying unit includes:

[0079] a feature extracting subunit, configured to generate a first target feature based on the to-be-detected image, where the first target feature includes a feature associated with the target object;

[0080] a feature converting subunit, configured to perform feature conversion on the structural key point information to obtain a second target feature;

[0081] a feature generating subunit, configured to generate a third target feature based on the first target feature and the second target feature; and

[0082] a classifying subunit, configured to determine the classification of the target object in the to-be-detected image based on the third target feature.

[0083] In some embodiments, the classifying subunit is configured to input the third target feature into a classification network to obtain a confidence that the target object has a defect and/or a confidence that the target object has no defect, and obtain a category of the target object based on the confidence that the target object has a defect and/or the confidence that the target object has no defect.

[0084] In some embodiments, the image detection apparatus further includes:

[0085] a feature fusion subunit, configured to perform semantic unit fusion processing on the third target feature before the classification of the target object in the to-be-detected image is determined based on the third target feature, to cause a semantic unit in the third target feature to be capable of fusing information of other semantic units in the third target feature; and

[0086] the classifying subunit is configured to determine, based on the third target feature after the semantic unit fusion processing, the classification of the target object in the to-be-detected image.

[0087] In some embodiments, the determining unit is further configured to determine that no defect of the target object exists in the to-be-detected image in response to the target object not existing in the to-be-detected image.

[0088] For the embodiments of the apparatus, since they basically correspond to the method embodiments, reference may be made to the partial description of the method embodiments for relevant parts. The apparatus embodiments described above are only schematic, and the modules described as separate modules may or may not be separated. Some or all of the modules may be selected according to actual needs to achieve the purpose of the solutions of the embodiments. Those of ordinary skill in the art may understand and implement them without any creative effort.

[0089] Correspondingly, according to one or more embodiments of the present disclosure, an electronic device is provided, including:

[0090] at least one memory and at least one processor;

[0091] where the memory is configured to store program code, and the processor is configured to call the program code stored in the memory to cause the electronic device to perform the image detection method provided according to one or more embodiments of the present disclosure.

[0092] Correspondingly, according to one or more embodiments of the present disclosure, a non-transitory computer storage medium is provided, where the non-transitory computer storage medium stores program code, and the program code is executable by a computer device to cause the computer device to perform the image detection method provided according to one or more embodiments of the present disclosure.

[0093] Referring to Fig. 5 below, which shows a schematic structural diagram of an electronic device 800 (such as a terminal device or a server) suitable for implementing the embodiments of the present disclosure. The terminal device in the embodiments of the present disclosure may include, but is not limited to, mobile terminals such as a mobile phone, a notebook computer, a digital broadcast receiver, a PDA (personal digital assistant), a PAD (tablet computer), a PMP (portable multimedia player), an in-vehicle terminal (such as an in-vehicle navigation terminal), etc., and fixed terminals such as a digital TV, a desktop computer, etc. The electronic device shown in Fig. 5 is only an example, and should not impose any limitation on the function and scope of use of the embodiments of the present disclosure.

[0094] As shown in Fig. 5, the electronic device 800 may include a processing apparatus (such as a central processing unit, a graphics processing unit, etc.) 801, which may perform various appropriate actions and processes based on a program stored in a read-only memory (ROM) 802 or a program loaded from a storage apparatus 808 into a random access memory (RAM) 803. The RAM 803 further stores various programs and data required for operations of the electronic device 800. The processing apparatus 801, the ROM 802, and the RAM 803 are interconnected by means of a bus 804. An input/output (I/O) interface 805 is also connected to the bus 804.

[0095] Usually, the following apparatuses may be connected to the I/O interface 805: an input apparatus 806 including, for example, a touch screen, a touchpad, a keyboard, a mouse, a camera, a microphone, an accelerometer, a gyroscope, etc.; an output apparatus 807 including, for example, a liquid crystal display (LCD), a speaker, a vibrator, etc.; the storage apparatus 808 including, for example, a magnetic tape, a hard disk, etc.; and a communication apparatus 809. The communication apparatus 809 may allow the electronic device 800 to communicate wirelessly or wiredly with other devices to exchange data. While Fig. 5 shows the electronic device 800 with various apparatuses, it should be understood that not all of the illustrated apparatuses are necessarily implemented or provided. More or fewer apparatuses may be implemented or provided alternatively.

[0096] In particular, according to the embodiments of the present disclosure, the process described above with reference to the flowchart may be implemented as a computer software program. For example, the embodiments of the present disclosure include a computer program product, which includes a computer program carried on a computer-readable medium, where the computer program contains program code for performing the method shown in the flowchart. In such an embodiment, the computer program may be downloaded and installed from a network via the communication apparatus 809, or installed from the storage apparatus 808, or installed from the ROM 802. When the computer program is executed by the processing apparatus 801, the above-mentioned functions defined in the method of the embodiment of the present disclosure are executed.

[0097] It should be noted that the above-mentioned computer-readable medium in the present disclosure may be a computer-readable signal medium, a computer-readable storage medium, or any combination of the two. The computer-readable storage medium may be, for example, but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, or device, or any combination of the above. More specific examples of the computer-readable storage medium may include, but are not limited to: an electrical connection having one or more wires, a portable computer magnetic disk, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or flash memory), an optical fiber, a portable compact disk read-only memory (CD-ROM), an optical storage device, a magnetic storage device, or any suitable combination of the above. In the present disclosure, the computer-readable storage medium may be any tangible medium that contains or stores a program, and the program may be used by or in combination with an instruction execution system, apparatus, or device. In the present disclosure, the computer-readable signal medium may include a data signal propagated on a baseband or as a part of a carrier wave, and computer-readable program code is carried therein. This propagated data signal may take many forms, including but not limited to an electromagnetic signal, an optical signal, or any suitable combination of the above. The

computer-readable signal medium may also be any computer-readable medium other than the computer-readable storage medium, and the computer-readable signal medium may send, propagate, or transmit the program used by or in combination with the instruction execution system, apparatus, or device. The program code contained on the computer-readable medium may be transmitted by any suitable medium, including but not limited to: a wire, an optical cable, an RF (radio frequency), etc., or any suitable combination of the above.

[0098] In some implementations, the client and the server may communicate using any currently known or future developed network protocol, such as the HTTP (HyperText Transfer Protocol), and may be interconnected with any form or medium of digital data communication (for example, a communication network). Examples of the communication network include a local area network ("LAN"), a wide area network ("WAN"), an internet (for example, the Internet), a peer-to-peer network (for example, an Ad-Hoc network), as well as any network currently known or to be developed in the future.

[0099] The above-mentioned computer-readable medium may be included in the above-mentioned electronic device, or it may exist alone without being assembled into the electronic device.

[00100] The above-mentioned computer-readable medium carries one or more programs, and when the one or more programs are executed by the electronic device, the electronic device is caused to perform the above-mentioned method of the present disclosure.

[00101] The computer program code for performing the operations of the present disclosure may be written in one or more programming languages or a combination thereof, where the programming languages include object-oriented programming languages such as Java, Smalltalk, and C++, and further include conventional procedural programming languages such as "C" language or similar programming languages. The program code may be executed entirely on a computer of a user, executed partly on the computer of the user, executed as an independent software package, executed partly on the computer of the user and partly on a remote computer, or executed entirely on the remote computer or a server. In the case of involving the remote computer, the remote computer may be connected to the computer of the user through any kind of network, including a local area network (LAN) or a wide area network (WAN), or it may be connected to an external computer (for example, connected by using Internet provided by an Internet service provider).

[00102] The flowcharts and block diagrams in the drawings illustrate the possibly implemented architectures, functions, and operations of the system, the method, and the computer program product according to various embodiments of the present disclosure. In this regard, each block in the flowchart or block diagram may represent a module, program segment, or part of code, and the module, program segment, or part of code contains one or more executable instructions for implementing specified logical functions. It should also be noted that, in some alternative implementations, the functions marked in the blocks may also occur in an order different from that marked in the drawings. For example, two blocks shown in succession may actually be performed substantially in parallel, or they may sometimes be performed in the reverse order, depending on the functions involved. It should also be noted that each block in the block diagram and/or the flowchart, and a combination of the blocks in the block diagram and/or the flowchart may be implemented by a dedicated hardware-based system that

executes specified functions or operations, or may be implemented by a combination of dedicated hardware and computer instructions.

[00103] The involved units described in the embodiments of the present disclosure may be implemented by means of software, and may also be implemented by means of hardware. Among them, the name of a unit does not constitute a limitation on the unit itself under certain circumstances.

[00104] The functions described above herein may be performed at least partly by one or more hardware logic components. For example, without limitation, exemplary types of hardware logic components that may be used include: a field programmable gate array (FPGA), an application specific integrated circuit (ASIC), an application specific standard product (ASSP), a system on chip (SOC), a complex programmable logical device (CPLD), etc.

[00105] In the context of the present disclosure, a machine-readable medium may be a tangible medium that may contain or store a program for use by or in combination with an instruction execution system, apparatus or device. The machine-readable medium may be a machine-readable signal medium or a machine-readable storage medium. The machine-readable medium may include, but is not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus or device, or any suitable combination of the above. More specific examples of the machine-readable storage medium will include an electrical connection based on one or more wires, a portable computer magnetic disk, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or flash memory), an optical fiber, a portable compact disk read-only memory (CD-ROM), an optical storage device, a magnetic storage device, or any suitable combination of the above.

[00106] According to one or more embodiments of the present disclosure, an image detection method is provided, including: determining whether a target object exists in a to-be-detected image; extracting, in response to the target object existing in the to-be-detected image, structural key point information of the target object in the to-be-detected image; and determining, based on the structural key point information and the to-be-detected image, a classification of the target object in the to-be-detected image.

[00107] According to one or more embodiments of the present disclosure, the target object includes a virtual portrait, and the structural key point information of the target object includes virtual human body key point information.

[00108] According to one or more embodiments of the present disclosure, determining, based on the structural key point information and the to-be-detected image, the classification of the target object in the to-be-detected image, includes: generating a first target feature based on the to-be-detected image, where the first target feature includes a feature associated with the target object; performing feature conversion on the structural key point information to obtain a second target feature; generating a third target feature based on the first target feature and the second target feature; and determining the classification of the target object in the to-be-detected image based on the third target feature.

5 [00109] According to one or more embodiments of the present disclosure, determining the classification of the target object in the to-be-detected image based on the third target feature includes: inputting the third target feature into a classification network to obtain a confidence that the target object has a defect and/or a confidence that the target object has no defect; and obtaining a category of the target object based on the confidence that the target object has a defect and/or the confidence that the target object has no defect.

10 [00110] According to one or more embodiments of the present disclosure, the method further includes: before the classification of the target object in the to-be-detected image is determined based on the third target feature, performing semantic unit fusion processing on the third target feature, to cause a semantic unit in the third target feature to be capable of fusing information of other semantic units in the third target feature; and determining the classification of the target object in the to-be-detected image based on the third target feature includes: determining, based on the third target feature after the semantic unit fusion processing, the classification of the target object in the to-be-detected image.

15 [00111] The method provided according to one or more embodiments of the present disclosure further includes: in response to the target object not existing in the to-be-detected image, determining that no defect of the target object exists in the to-be-detected image.

20 [00112] According to one or more embodiments of the present disclosure, an image detection apparatus is provided, including: a determining unit, configured to determine whether a target object exists in a to-be-detected image; an extracting unit, configured to extract structural key point information of the target object in the to-be-detected image in response to the target object existing in the to-be-detected image; and a classifying unit, configured to determine a classification of the target object in the to-be-detected image based on the structural key point information and the to-be-detected image.

25 [00113] According to one or more embodiments of the present disclosure, an electronic device is provided, including: at least one memory and at least one processor; where the memory is configured to store program code, and the processor is configured to call the program code stored in the memory to cause the electronic device to perform the image detection method provided according to one or more embodiments of the present disclosure.

30 [00114] According to one or more embodiments of the present disclosure, a non-transitory computer storage medium is provided, the non-transitory computer storage medium stores program code, and the program code, when executed by a computer device, causes the computer device to perform the image detection method provided according to one or more embodiments of the present disclosure.

35 [00115] The above description is only preferred embodiments of the present disclosure and an explanation of the technical principles employed. Those skilled in the art should understand that the scope of disclosure involved in the present disclosure is not limited to the technical solutions formed by the specific combination of the above-mentioned technical features, and should also cover, without departing from the above-mentioned disclosed concept, other technical solutions formed by any combination of the above-mentioned technical features or their equivalent features. For example, a technical solution formed by replacing the above features

with technical features with similar functions disclosed in the present disclosure (but not limited to).

[00116] Additionally, although operations are depicted in a particular order, it should not be understood that these operations are required to be performed in a specific order as illustrated or in a sequential order. Under certain circumstances, multitasking and parallel processing may be advantageous. Likewise, although the above
5 discussion includes several specific implementation details, these should not be interpreted as limitations on the scope of the present disclosure. Certain features that are described in the context of separate embodiments may also be implemented in combination in a single embodiment. Conversely, various features that are described in the context of a single embodiment may also be implemented in multiple embodiments separately or in any suitable sub-combinations.

10 **[00117]** Although the subject matter has been described in language specific to structural features and/or method logical actions, it should be understood that the subject matter defined in the appended claims is not necessarily limited to the specific features or actions described above. Rather, the specific features and actions described above are merely example forms of implementing the claims.

I/We Claim:

1. An image detection method, comprising:
determining whether a target object exists in a to-be-detected image;
extracting, in response to the target object existing in the to-be-detected image, structural key point
5 information of the target object in the to-be-detected image; and
determining, based on the structural key point information and the to-be-detected image, a classification
of the target object in the to-be-detected image.

2. The method of claim 1, wherein the target object comprises a virtual portrait, and the structural key
10 point information of the target object comprises virtual human body key point information.

3. The method of claim 1, wherein determining, based on the structural key point information and the to-
be-detected image, the classification of the target object in the to-be-detected image, comprises:
generating a first target feature based on the to-be-detected image, wherein the first target feature
15 comprises a feature associated with the target object;
performing feature conversion on the structural key point information to obtain a second target feature;
generating a third target feature based on the first target feature and the second target feature; and
determining the classification of the target object in the to-be-detected image based on the third target
feature.

20 4. The method of claim 3, wherein determining the classification of the target object in the to-be-detected
image based on the third target feature comprises:
inputting the third target feature into a classification network to obtain a confidence that the target object
has a defect and/or a confidence that the target object has no defect; and
25 obtaining a category of the target object based on the confidence that the target object has a defect and/or
the confidence that the target object has no defect.

5. The method of claim 3, wherein
the method further comprises: before determining the classification of the target object in the to-be-
30 detected image based on the third target feature, performing semantic unit fusion processing on the third target
feature, to cause a semantic unit in the third target feature to be capable of fusing information of other semantic
units in the third target feature; and
determining the classification of the target object in the to-be-detected image based on the third target
feature comprises: determining the classification of the target object in the to-be-detected image based on the
35 third target feature after the semantic unit fusion processing.

6. The method of claim 1, further comprising:

in response to the target object not existing in the to-be-detected image, determining that a defect of the target object does not exist in the to-be-detected image.

5 7. An image detection apparatus, comprising:

a determining unit, configured to determine whether a target object exists in a to-be-detected image;

an extracting unit, configured to extract, in response to the target object existing in the to-be-detected image, structural key point information of the target object in the to-be-detected image;

10 a classifying unit, configured to determine, based on the structural key point information and the to-be-detected image, a classification of the target object in the to-be-detected image.

8. An electronic device, comprising:

at least one memory and at least one processor;

15 wherein the memory is configured to store program code, and the processor is configured to call the program code stored in the memory to cause the electronic device to perform the method of any of claims 1 to 6.

9. A non-transitory computer storage medium, wherein

20 the non-transitory computer storage medium stores program code, and the program code, when executed by a computer device, causes the computer device to perform the method of any of claims 1 to 6.

10. A computer program product comprising program code that, when executed by a computer device, causes the computer device to perform the method of any of claims 1 to 6.