Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3043230 C 2020/06/23

(11)(21) 3 043 230

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) **Date de dépôt PCT/PCT Filing Date:** 2018/06/07
(87) **Date publication PCT/PCT Publication Date:** 2018/12/13
(45) **Date de délivrance/Issue Date:** 2020/06/23
(85) **Entrée phase nationale/National Entry:** 2019/05/07
(86) **N° demande PCT/PCT Application No.:** US 2018/036505
(87) **N° publication PCT/PCT Publication No.:** 2018/226990
(30) **Priorité/Priority:** 2017/06/07 (CN201710421333.5)

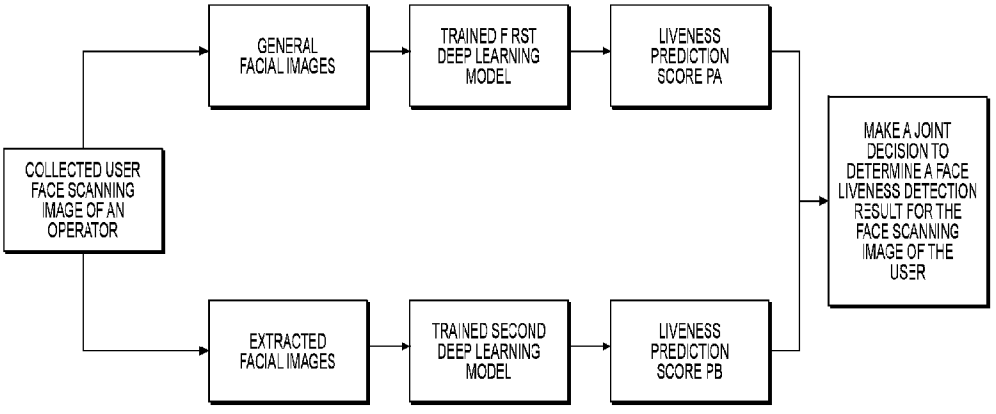
(51) **Cl.Int./Int.Cl.** ***G06K 9/00*** (2006.01)

(72) **Inventeur/Inventor:**
MA, CHENGUANG, CN

(73) **Propriétaire/Owner:**
ALIBABA GROUP HOLDING LIMITED, KY

(74) **Agent:** KIRBY EADES GALE BAKER

(54) **Titre : PROCEDE ET APPAREIL DE DETECTION D'ANIMATION DE VISAGE, ET DISPOSITIF ELECTRONIQUE**
(54) **Title: FACE LIVENESS DETECTION METHOD AND APPARATUS, AND ELECTRONIC DEVICE**



(57) **Abrégé/Abstract:**
A first deep learning model is trained based on general facial images. A second deep learning model is trained based on extracted facial images cropped from the general facial images. Face liveness detection is performed based on the trained first deep learning model to obtain a first prediction score and the trained second deep learning model to obtain a second prediction score. A prediction score result is generated based on the first prediction score and the second prediction score, and the prediction score result is compared with a threshold to determine a face liveness detection result for the extracted facial images.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 December 2018 (13.12.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/226990 A1

(51) International Patent Classification:
G06K 9/00 (2006.01)

(21) International Application Number:
PCT/US2018/036505

(22) International Filing Date:
07 June 2018 (07.06.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
201710421333.5 07 June 2017 (07.06.2017) CN

(71) Applicant: **ALIBABA GROUP HOLDING LIMITED** [—/US]; Fourth Floor, One Capital Place, P.O. Box 847, George Town, Grand Cayman (KY).

(72) Inventor: **MA, Chenguang**; c/o Ants Patent Team, 17f Building B, Huanglong Times Plaza, No. 18 Wantang Road, Hangzhou, 310099 (CN).

(74) Agent: **STALFORD, Terry, J.**; Fish & Richardson P.C., P.O. Box 1022, Minneapolis, MN 55440-1022 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*

WO 2018/226990 A1

(54) Title: FACE LIVENESS DETECTION METHOD AND APPARATUS, AND ELECTRONIC DEVICE

(57) Abstract: A first deep learning model is trained based on general facial images. A second deep learning model is trained based on extracted facial images cropped from the general facial images. Face liveness detection is performed based on the trained first deep learning model to obtain a first prediction score and the trained second deep learning model to obtain a second prediction score. A prediction score result is generated based on the first prediction score and the second prediction score, and the prediction score result is compared with a threshold to determine a face liveness detection result for the extracted facial images.

FACE LIVENESS DETECTION METHOD AND APPARATUS, AND ELECTRONIC DEVICE

TECHNICAL FIELD

[0002] The present application relates to the field of computer software technologies, and in particular, to a face liveness detection method, apparatus, and electronic device.

BACKGROUND

[0003] A face liveness detection technology is used to determine whether the current user is the authentic user by using facial recognition techniques so as to intercept spoofing attacks such as a screen replay attack, a printed photo attack, and a three-dimensional modeling attack.

[0004] Currently, the face liveness detection technology can be classified into an intrusive face liveness detection technology and a non-intrusive face liveness detection technology. In the intrusive face liveness detection technology, a user needs to cooperatively complete some specific live actions such as blinking, head turning, or mouth opening. When performing facial recognition based on the given instructions, the liveness detection module can determine whether an operator accurately completes the live operation and whether the operator is the authentic user. In the non-intrusive face liveness detection technology, a user does not need to cooperatively complete a live action, so that user experience is better, but the technical complexity is higher. In addition, liveness detection is performed mainly depending on information about an input single frame image or information about other device sensors.

[0005] In the described non-intrusive face liveness detection technology in the existing technology, supervised training is usually performed on a single deep learning model by using live and non-live facial images, and then face liveness prediction is performed on the input single frame image by using the trained model.

[0006] However, such a technical solution heavily depends on a spoofing face attack type of training data, and is limited by an objective condition of insufficient training data. It is difficult to fully extract a live face image feature. As a result, this model cannot fully express a live face feature, and accuracy of a face liveness detection result is reduced.

CA 3043230 2019-06-11

SUMMARY

[0007] Embodiments of the present application provide a face liveness detection method, apparatus, and electronic device to resolve the following technical problems in the existing technology. In a technical solution based on a single deep learning model, it is difficult to fully extract a live face image feature. As a result, this model cannot fully express a live face feature, and accuracy of a face liveness detection result is reduced.

[0008] To resolve the described technical problems, the embodiments of the present application are implemented as follows:

[0009] An embodiment of the present application provides a face liveness detection method, including: training a first deep learning model based on the general facial images; training a second deep learning model based on the extracted facial images cropped from the general facial images; and performing face liveness detection based on the trained first deep learning model and the trained second deep learning model.

[0010] An embodiment of the present application provides a face liveness detection apparatus, including: a training module, configured to: train a first deep learning model based on the general facial images; and train a second deep learning model based on the extracted facial images cropped from the general facial images; and a detection module, configured to perform face liveness detection based on the trained first deep learning model and the trained second deep learning model.

[0011] At least one technical solution used in the embodiments of the present application can achieve the following beneficial effects. One such benefit is more live face image features are extracted. Compared with a model in the existing technology, the trained first deep learning model and the trained second deep learning model jointly better express the live face feature, thereby improving the accuracy of the face liveness detection result. Therefore, a part or all of problems in the existing technology can be resolved.

BRIEF DESCRIPTION OF DRAWINGS

[0012] To describe the technical solutions in embodiments of the present application or in the existing technology more clearly, the following briefly introduces the accompanying drawings required for describing the embodiments or the existing technology. Apparently, the accompanying drawings in the following description merely show some embodiments of the present application, and a person of ordinary skill in the art can still derive other drawings from these accompanying drawings without creative efforts.

[0013] FIG. 1 is a schematic flowchart illustrating an example of a model training stage in a solution of the present application;

[0014] FIG. 2 is a schematic flowchart illustrating an example of a liveness detection stage in a solution of the present application;

[0015] FIG. 3 is a schematic flowchart illustrating a face liveness detection method according to an embodiment of the present application;

[0016] FIG. 4 is a schematic diagram illustrating comparison between a general facial image and an extracted facial image according to an embodiment of the present application;

[0017] FIG. 5 is a schematic structural diagram illustrating a face liveness detection apparatus corresponding to FIG. 3 according to an embodiment of the present application; and

[0018] FIG. 6 is a flowchart illustrating an example of a computer-implemented method for determining user authenticity with face liveness detection, according to an implementation of the present disclosure.

DESCRIPTION OF EMBODIMENTS

[0019] Embodiments of the present application provide a face liveness detection method, apparatus, and electronic device.

[0020] To make a person skilled in the art better understand the technical solutions in the present application, the following clearly and completely describes the technical solutions in the embodiments of the present application with reference to the accompanying drawings in the embodiments of the present application. Apparently, the described embodiments are merely a part rather than all of the embodiments of the present application. All other embodiments obtained by a person of ordinary skill in the art based on the embodiments of the present application without creative efforts shall fall within the protection scope of the present application.

[0021] All deep learning models in a solution of the present application are based on a neural network. For ease of description, a core idea of the solution of the present application is first described based on an example and with reference to FIG. 1 and FIG. 2.

[0022] In this example, the solution of the present application can be classified into a model training stage and a liveness detection stage.

[0023] FIG. 1 is a schematic flowchart illustrating an example of a model training stage in a solution of the present application. In a model training stage, two independent deep learning models are trained by using live and non-live samples (belonging to a training data set) in a facial image: a first deep learning model and a second deep learning model. An input image of the first deep learning model is a collected general facial image, and an input image of the second deep learning model can be an extracted facial image cropped from the general facial image. The first deep learning model and the second deep learning model can use different deep learning network structures (i.e. a structure of a neural network that a model is based on). Different network structures are differently sensitive to different image features. Live and non-live training data sets are used to complete training of the first deep learning model and the second deep learning model based on a deep learning method.

[0024] FIG. 2 is a schematic flowchart illustrating an example of a liveness detection stage in a solution of the present application. In a liveness detection stage, a face scanning image of a user is collected as a general facial image of the user, and a first deep learning model is input to obtain a prediction score PA. In addition, face detection is performed on the face scanning image of the user, an extracted facial image is cropped from the face scanning image of the user based on a detection result, and a second deep learning model is input to the

extracted facial image, to obtain a prediction score PB. Afterwards, for example, a prediction score result of (PA+PB) can be compared with a determined threshold (e.g. the threshold can be 1), to make a joint decision to determine a face liveness detection result for the face scanning image of the user.

[0025] Based on the described core idea, the following describes the solution of the present application in detail.

[0026] FIG. 3 is a schematic flowchart illustrating a face liveness detection method according to an embodiment of the present application. From a perspective of a program, the procedure can be executed by a program on a server or a terminal, for example, an identity authentication program or an e-commerce application. From a perspective of a device, the procedure is executed by at least one of the following devices that can be used as a server or a terminal: an access control device, a personal computer, a medium computer, a computer cluster, a mobile phone, a tablet computer, an intelligent wearable device, a car machine, or a point of sale (POS).

[0027] The procedure in FIG. 3 can include the following steps.

[0028] S301. Train a first deep learning model based on general facial images.

[0029] In this embodiment of the present application, the general facial images used to train the first deep learning model can include a plurality of samples. In the plurality of samples, some are live facial images that are collected by shooting a live face and that can be used as positive samples, and some are non-live facial images that are collected by shooting a non-live face such as a face picture or a face model and that can be used as negative samples.

[0030] In this embodiment of the present application, the first deep learning model is a classification model, and the general facial images are used as inputs of the classification model. After model processing, the general facial images can be classified into at least the live facial image category or the non-live facial image category. An objective of training the first deep learning model is to improve classification accuracy of the first deep learning model.

[0031] S302. Train a second deep learning model based on extracted facial images cropped from the general facial images.

[0032] In this embodiment of the present application, in addition to an entire facial region, the general facial image generally includes some unrelated regions, such as a background region and a human body except a face. The extracted facial image can exclude the unrelated regions, and can include at least an extracted facial region, for example, an

entire facial region, an eye region, or a nasal region. There can be one or more second deep learning models, and each second deep learning model can correspond to a type of facial regions.

[0033] FIG. 4 is a schematic diagram illustrating comparison between a general facial image and an extracted facial image according to an embodiment of the present application.

[0034] In FIG. 4, (a) is a general facial image. For ease of understanding, an extracted facial image is marked in (a) by using dashed lines, and (a) can be correspondingly cropped to obtain an extracted facial image shown in (b).

[0035] In addition, when the extracted facial image is an image including only a partial facial region, the general facial image can also be an image including an entire facial region and basically excluding an unrelated region.

[0036] In this embodiment of the present application, the extracted facial image used to train the second deep learning model can also include a variety of samples. In the variety of samples, some are live facial images that can be used as positive samples, and some are non-live facial images that can be used as negative samples.

[0037] In this embodiment of the present application, the second deep learning model is also a classification model, and the extracted facial images are used as input of the classification model. After model processing, the extracted facial images can be classified into at least the live facial image category or the non-live facial image category. An objective of training the second deep learning model is to improve classification accuracy of the second deep learning model.

[0038] In addition to being cropped from the general facial image, the extracted facial image can be obtained by means of special collection without depending on the general facial image.

[0039] In this embodiment of the present application, the first deep learning model and the second deep learning model can be different models or a same model before training.

[0040] An execution sequence of step S301 and step S302 is not limited in the present application, and step S301 and step S302 can be simultaneously or successively performed.

[0041] S303. Perform face liveness detection based on the trained first deep learning model and the trained second deep learning model.

[0042] Each step in FIG. 3 can be performed by a same device or a same program, or can be performed by different devices or different programs. For example, step S301 to step S303 are performed by a device 1. For another example, both step S301 and step S302 are

performed by a device 1, and step S303 is performed by a device 2; etc.

[0043] According to the method in FIG. 3, more live face image features are extracted. Compared with a model in the existing technology, the trained first deep learning model and the trained second deep learning model jointly better express a live face feature, thereby improving accuracy of a face liveness detection result. Therefore, a part or all of problems in the existing technology can be resolved.

[0044] Based on the method in FIG. 3, this embodiment of the present application further provides some specific implementation solutions of the method and an extension solution, which are described below.

[0045] In this embodiment of the present application, to implement a difference between sensitivity of the first deep learning model to an image feature and sensitivity of the second deep learning model to an image feature, the first deep learning model and the second deep learning model can preferably use different deep learning network structures.

[0046] Different network structures of two deep learning models can indicate that the two deep learning models include one or more different network structure parameters. The network structure parameter can include, for example, a quantity of hidden variable layers, a type of a hidden variable layer, a quantity of neuron nodes, a quantity of input layer nodes, or a quantity of output layer nodes.

[0047] Certainly, some specific deep learning models can also include corresponding specific parameters. For example, for a deep learning model based on a convolutional neural network widely used in the image field currently, a size of a convolution kernel of a convolution unit is also a specific network structure parameter of this deep learning model.

[0048] For the solution of the present application, generally, the different deep learning network structures include at least one of the following parameters: a quantity of hidden variable layers, a type of a hidden variable layer, a quantity of neuron nodes, or a size of a convolution kernel of a convolution unit.

[0049] In this embodiment of the present application, to improve model training efficiency and model training reliability, model training can be performed in a supervised training manner.

[0050] For example, in a supervised training manner, for step S301, the general facial image includes a first label, and the first label indicates whether a general facial image corresponding to the first label is a live facial image.

[0051] The training a first deep learning model based on a general facial image can

include: inputting the first deep learning model to the general facial image, where the first deep learning model extracts a feature of the general facial image, and predicts, based on the extracted feature, whether the general facial image is a live facial image; and adjusting the first deep learning model based on a prediction result and the first label of the general facial image. Generally, when the prediction result is inconsistent with the first label, the first deep learning model is adjusted, so that the adjusted first deep learning model can obtain, by means of re-prediction, a prediction result consistent with the first label.

[0052] The feature extracted by the first deep learning model in a training process can preferably include an image structure feature of the general facial image, for example, a screen photo edge or face distortion in the general facial image.

[0053] For another example, similarly, in a supervised training manner, for step S302, the extracted facial image includes a second label, and the second label indicates whether an extracted facial image corresponding to the second label is a live facial image.

[0054] The training a second deep learning model based on the extracted facial images cropped from the general facial images can include: obtaining the extracted facial images cropped from the general facial images; applying the second deep learning model to the obtained extracted facial image, where the second deep learning model extracts a feature of the extracted facial image, and predicts, based on the extracted feature, whether the extracted facial image is a live facial image; and adjusting the second deep learning model based on a prediction result and the second label of the extracted facial image. Generally, when the prediction result is inconsistent with the second label, the second deep learning model is adjusted, so that the adjusted second deep learning model can obtain a prediction result consistent with the second label by means of re-prediction.

[0055] The feature extracted by the second deep learning model in a training process can preferably include an image material feature of the extracted facial image, for example, blurring, texture, or color distortion in the extracted facial image.

[0056] In the two examples described above, the first deep learning model and the second deep learning model are differently sensitive to different image features. The first deep learning model is more sensitive to the image structure feature, and the second deep learning model is more sensitive to the image material feature. For a face image, the image structure feature is relatively a global and generalized feature, and the image material feature is relatively a local and refined feature.

[0057] Therefore, the trained first deep learning model and the trained second deep

learning model can jointly extract a face image feature more hierarchically and abundantly, so as to make a joint decision to obtain a more accurate face liveness detection result.

[0058] In this embodiment of the present application, corresponding training data sets and/or corresponding deep learning network structures are different, so that the first deep learning model and the second deep learning model can be differently sensitive to different image features.

[0059] For example, if the first deep learning model and the second deep learning model are based on a convolutional neural network, a convolution kernel of a convolution unit in a convolutional neural network that the first deep learning model is based on can be relatively large, so that the first deep learning model extracts an image structure feature of the general facial image. Correspondingly, a convolution kernel of a convolution unit in a convolutional neural network that the second deep learning model is based on can be relatively small, so that the second deep learning model extracts an image material feature of the extracted facial image. Therefore, in this example, the convolution kernel of the convolution unit in the convolutional neural network that the first deep learning model is based on is greater than the convolution kernel of the convolution unit in the convolutional neural network that the second deep learning model is based on.

[0060] It should be noted that the size of the convolution kernel is merely an example of a parameter that can affect the sensitivity, and another network structure parameter can also affect the sensitivity.

[0061] In this embodiment of the present application, for step S303, the trained first deep learning model and the trained second deep learning model jointly make a decision to perform the face liveness detection. There are a variety of specific decision manners. For example, a separate decision is made by separately using the first deep learning model and the second deep learning model, and then a final decision result is determined by synthesizing all separate decision results. For another example, a separate decision can be first made by using either of the first deep learning model and the second deep learning model. When a separate decision result satisfies a specific condition, the separate decision result can be directly used as a final decision result; otherwise, a decision is comprehensively made in combination with another remaining model, to obtain a final decision result; etc.

[0062] If a first manner described in the previous paragraph is used, an example is as follows:

[0063] For example, for step S303, the performing face liveness detection based on

the trained first deep learning model and the trained second deep learning model can include: obtaining the general facial image (which is generally a face scanning image of a user) collected for the face liveness detection; inputting the trained first deep learning model to the collected general facial image for processing, to obtain corresponding first prediction data; obtaining the extracted facial image cropped from the collected general facial image, and inputting the trained second deep learning model for processing, to obtain corresponding second prediction data; and making a joint decision based on the first prediction data and the second prediction data, to obtain a face liveness detection result for the face scanning image of the user.

[0064] The first prediction data can be, for example, the described prediction score PA, and the second prediction data can be, for example, the described prediction score PB. Certainly, the prediction score is merely an example of an expression form of the first prediction data and the second prediction data, or there can be another expression form, for example, a probability value or a Boolean value.

[0065] The above is the face liveness detection method provided in this embodiment of the present application. As shown in FIG. 5, based on a same idea of the disclosure, an embodiment of the present application further provides a corresponding apparatus.

[0066] FIG. 5 is a schematic structural diagram illustrating a face liveness detection apparatus corresponding to FIG. 3 according to an embodiment of the present application. The apparatus can be located on an execution body of the procedure in FIG. 3, including: a training module 501, configured to: train a first deep learning model based on the general facial images; and train a second deep learning model based on the extracted facial images cropped from the general facial images; and a detection module 502, configured to perform face liveness detection based on the trained first deep learning model and the trained second deep learning model.

[0067] Optionally, the first deep learning model and the second deep learning model use different deep learning network structures.

[0068] Optionally, the different deep learning network structures include at least one of the following parameters: a quantity of hidden variable layers, a type of a hidden variable layer, a quantity of neuron nodes, or a size of a convolution kernel of a convolution unit.

[0069] Optionally, the general facial image includes a first label, and the first label indicates whether a general facial image corresponding to the first label is a live facial image.

[0070] The training, by the training module 501, a first deep learning model based on

the general facial images includes: inputting, by the training module 501, the first deep learning model to the general facial image, where the first deep learning model predicts, based on an image structure feature of the general facial image, whether the general facial image is a live facial image; and adjusting the first deep learning model based on a prediction result and the first label of the general facial image.

[0071] Optionally, the extracted facial image includes a second label, and the second label indicates whether an extracted facial image corresponding to the second label is a live facial image.

[0072] The training, by the training module 501, a second deep learning model based on the extracted facial images cropped from the general facial images includes: obtaining, by the training module 501, the extracted facial image cropped from the general facial image; and inputting the second deep learning model to the extracted facial image, where the second deep learning model predicts, based on an image material feature of the extracted facial image, whether the extracted facial image is a live facial image; and adjusting the second deep learning model based on a prediction result and the second label of the extracted facial image.

[0073] Optionally, the first deep learning model and the second deep learning model are based on a convolutional neural network.

[0074] A convolution kernel of a convolution unit in a convolutional neural network that the first deep learning model is based on is greater than a convolution kernel of a convolution unit in a convolutional neural network that the second deep learning model is based on, so that the first deep learning model extracts an image structure feature of the general facial image, and the second deep learning model extracts an image material feature of the extracted facial image.

[0075] Optionally, the performing, by the detection module 502, face liveness detection based on the trained first deep learning model and the trained second deep learning model includes: obtaining, by the detection module 502, the general facial image collected for the face liveness detection; inputting the trained first deep learning model to the collected general facial image for processing, to obtain corresponding first prediction data; obtaining extracted facial image cropped from the collected general facial image, and inputting the trained second deep learning model for processing, to obtain corresponding second prediction data; and making a joint decision based on the first prediction data and the second prediction data, to obtain a face liveness detection result for a face scanning image of the user.

[0076] Based on a same idea of the disclosure, an embodiment of the present application further provides a corresponding electronic device, including: at least one processor; and a memory communicatively connected to the at least one processor.

[0077] The memory stores an instruction that can be executed by the at least one processor, and the instruction is executed by the at least one processor, to enable the at least one processor to: train a first deep learning model based on the general facial images; train a second deep learning model based on the extracted facial images cropped from the general facial images; and perform face liveness detection based on the trained first deep learning model and the trained second deep learning model.

[0078] Based on a same idea of the disclosure, an embodiment of the present application further provides a corresponding non-volatile computer storage medium, where the non-volatile computer storage medium stores a computer executable instruction, and the computer executable instruction is set to: train a first deep learning model based on the general facial images; train a second deep learning model based on the extracted facial images cropped from the general facial images; and perform face liveness detection based on the trained first deep learning model and the trained second deep learning model.

[0079] The embodiments in this specification are all described in a progressive manner, for same or similar parts in the embodiments, reference can be made to these embodiments, and each embodiment focuses on a difference from other embodiments. Especially, an apparatus embodiment, an electronic device embodiment, a non-volatile computer storage medium embodiment are basically similar to a method embodiment, and therefore is described briefly; for related parts, reference is made to partial descriptions in the method embodiment.

[0080] The apparatus, the electronic device, and the non-volatile computer storage medium provided in the embodiments of the present application correspond to the method. Therefore, the apparatus, the electronic device, and the non-volatile computer storage medium also have beneficial technical effects similar to a beneficial technical effect of the corresponding method. The beneficial technical effect of the method is described in detail above, so that the beneficial technical effects of the corresponding apparatus, electronic device, and non-volatile computer storage medium are not described here again.

[0081] In the 1990s, whether technology improvement is hardware improvement (for example, improvement of a circuit structure, such as a diode, a transistor, or a switch) or software improvement (improvement of a method procedure) can be obviously distinguished.

However, as technologies develop, improvement of many current method procedures can be considered as direct improvement of a hardware circuit structure. A designer usually programs an improved method procedure to a hardware circuit to obtain a corresponding hardware circuit structure. Therefore, a method procedure can be improved by hardware entity modules. For example, a programmable logic device (PLD) (e.g. a field programmable gate array (FPGA)) is such an integrated circuit, and a logical function of the programmable logic device is determined by a user by means of device programming. The designer performs programming to "integrate" a digital system to a PLD without requesting a chip manufacturer to design and produce an application-specific integrated circuit chip. In addition, the programming is mostly implemented by modifying "logic compiler" software instead of manually making an integrated circuit chip. This is similar to a software compiler used to develop and compose a program. However, original code obtained before compilation is also written in a specific programming language, and this is referred to as hardware description language (Hardware Description Language, HDL). However, there are various HDLs, such as an ABEL (Advanced Boolean Expression Language), an AHDL (Altera Hardware Description Language), Confluence, a CUPL (Cornell University Programming Language), HDCal, a JHDL (Java Hardware Description Language), Lava, Lola, MyHDL, PALASM, and an RHDL (Ruby Hardware Description Language). Currently, a VHDL (Very-High-Speed Integrated Circuit Hardware Description Language) and Verilog are most popular. A person skilled in the art should also understand that, only logic programming needs to be performed on the method procedure by using the described several hardware description languages, and the several hardware description languages are programmed to an integrated circuit, so that a hardware circuit that implements the logical method procedure can be easily obtained.

[0082] A controller can be implemented in any appropriate manner. For example, the controller can use a microprocessor or a processor, and can store forms of a computer readable medium, a logic gate, a switch, an application-specific integrated circuit (ASIC), a programmable logic controller, and an embedded microcontroller that are of computer readable program code (e.g. software or hardware) that can be executed by the (micro) processor. The examples of controller include butare not limited to the following microcontrollers: ARC 625D, Atmel AT91SAM, Microchip PIC18F26K20, or Silicone Labs C8051F320. A memory controller can also be implemented as a part of control logic of the memory. A person skilled in the art also knows that, in addition to implementing the

controller in a pure computer readable program code manner, logic programming can be completely performed by using the method step, so that the controller implements a same function in a form of a logical gate, a switch, an application-specific integrated circuit, a programmable logic controller, an embedded microcontroller, etc. Therefore, the controller can be considered as a hardware component, and an apparatus for implementing various functions in the controller can also be considered as a structure in a hardware component. Alternatively, an apparatus configured to implement various functions can be considered as a software module or a structure in a hardware component that can implement the method.

[0083] The system, apparatus, module, or unit described in the described embodiments can be implemented by a computer chip or an entity, or implemented by a product with a function. A typical implementation device is a computer. Specifically, the computer can be, for example, a personal computer, a laptop computer, a cellular phone, a camera phone, a smartphone, a personal digital assistant, a media player, a navigation device, an email device, a game console, a tablet computer, or a wearable device, or a combination of any of these devices.

[0084] For ease of description, the described apparatus is described by dividing functions into various units. Certainly, when the present application is implemented, the functions of each unit can be implemented in one or more pieces of software and/or hardware.

[0085] A person skilled in the art should understand that the embodiments of the present disclosure can be provided as a method, a system, or a computer program product. Therefore, the present disclosure can use a form of hardware only embodiments, software only embodiments, or embodiments with a combination of software and hardware. In addition, the present disclosure can use a form of a computer program product that is implemented on one or more computer-usable storage media (including but not limited to a disk memory, a CD-ROM, an optical memory, etc.) that include computer-usable program code.

[0086] The present disclosure is described with reference to the flowcharts and/or block diagrams of the method, the device (system), and the computer program product according to the embodiments of the present disclosure. It should be understood that computer program instructions can be used to implement each process and/or each block in the flowcharts and/or the block diagrams and a combination of a process and/or a block in the flowcharts and/or the block diagrams. These computer program instructions can be provided

for a general-purpose computer, a dedicated computer, an embedded processor, or a processor of any other programmable data processing device to generate a machine, so that the instructions executed by a computer or a processor of any other programmable data processing device generate an apparatus for implementing a specific function in one or more processes in the flowcharts or in one or more blocks in the block diagrams.

[0087] These computer program instructions can be stored in a computer readable memory that can instruct the computer or any other programmable data processing device to work in a specific manner, so that the instructions stored in the computer readable memory generate an artifact that includes an instruction apparatus. The instruction apparatus implements a specific function in one or more processes in the flowcharts and/or in one or more blocks in the block diagrams.

[0088] These computer program instructions can be loaded to a computer or another programmable data processing device, so that a series of operations and steps are performed on the computer or another programmable device, thereby generating computer-implemented processing. Therefore, the instructions executed on the computer or another programmable device provide steps for implementing a specific function in one or more processes in the flowcharts or in one or more blocks in the block diagrams.

[0089] In typical configuration, the computing device includes one or more processors (CPU), an input/output interface, a network interface, and a memory.

[0090] The memory can include a form of a volatile memory, a random access memory (RAM) and/or a non-volatile memory, etc. in a computer readable medium, such as a read-only memory (ROM) or a flash memory (flash RAM). The memory is an example of the computer readable medium.

[0091] The computer readable medium includes volatile and non-volatile, removable and non-removable media, and can store information by using any method or technology. The information can be a computer readable instruction, a data structure, a program module, or other data. The examples of computer storage medium include but are not limited to a phase change random access memory (PRAM), a static random access memory (SRAM), a dynamic random access memory (DRAM), a random access memory (RAM) of another type, a read-only memory (ROM), an electrically erasable programmable read-only memory (EEPROM), a flash memory or another memory technology, a compact disc read-only memory (CD-ROM), a digital versatile disc (DVD) or another optical storage, a magnetic tape, a magnetic disk storage, another magnetic storage device, or any other non-transmission

medium. The computer storage medium can be used to store information that can be accessed by the computing device. As described in this specification, the computer readable medium does not include transitory media (transitory media), for example, a modulated data signal and a carrier.

[0092] It should be further noted that, terms "include", "contain", or their any other variant is intended to cover non-exclusive inclusion, so that a process, a method, an article, or a device that includes a series of elements not only includes these very elements, but also includes other elements which are not expressly listed, or further includes elements inherent to such process, method, article, or device. An element preceded by "includes a ..." does not, without more constraints, preclude the existence of additional identical elements in the process, method, article, or device that includes the element.

[0093] The present application can be described in common contexts of computer executable instructions executed by a computer, such as a program module. Generally, the program module includes a routine, a program, an object, a component, a data structure, etc. executing a specific task or implementing a specific abstract data type. The present application can also be practiced in distributed computing environments. In these distributed computing environments, tasks are executed by remote processing devices that are connected by using a communications network. In the distributed computing environments, the program module can be located in local and remote computer storage media that include storage devices.

[0094] The embodiments in this specification are all described in a progressive manner, for same or similar parts in the embodiments, reference can be made to these embodiments, and each embodiment focuses on a difference from other embodiments. Especially, a system embodiment is basically similar to a method embodiment, and therefore is described briefly; for related parts, reference can be made to partial descriptions in the method embodiment.

[0095] The previous descriptions are merely embodiments of the present application, and are not intended to limit the present application. For a person skilled in the art, the present application can have various modifications and changes. Any modifications, equivalent replacements, improvements, etc. made within the spirit and principle of the present application shall fall within the protection scope of the present application.

[0096] FIG. 6 is a flowchart illustrating an example of a computer-implemented method 600 for determining user authenticity with face liveness detection, according to an

implementation of the present disclosure. For clarity of presentation, the description that follows generally describes method 600 in the context of the other figures in this description. However, it will be understood that method 600 can be performed, for example, by any system, environment, software, and hardware, or a combination of systems, environments, software, and hardware, as appropriate. In some implementations, various steps of method 600 can be run in parallel, in combination, in loops, or in any order.

[0097] At 602, a first deep learning model is trained to classify general facial images. The general facial images are classified into at least live facial images and non-live facial images. In some implementations, the live facial images are considered to be positive samples and the non-live facial images are considered to be negative samples. In some implementations, the first deep learning model is a classification model and the general facial images are used as inputs of the first deep learning model. Training the first deep learning model improves classification accuracy with respect to the general facial images.

[0098] In some implementations, a particular general facial image includes a first label indicating whether the particular general facial image corresponding to the first label is a live facial image. In some implementations, the training of the first deep learning model includes: 1) inputting the particular general facial image to the first deep learning model to generate a first prediction result, based on an image structure feature of the particular general facial image, of whether the particular general facial image is a live facial image and 2) adjusting the first deep learning model based on the first prediction result and the first label. From 602, method 600 proceeds to 604.

[0099] At 604, cropped facial images are extracted from the general facial images. In some implementations, a particular cropped facial image includes a second label, and the second label indicates whether the particular cropped facial image corresponding to the second label is a live facial image. In some implementations, the training of the second deep learning model based on the cropped facial image includes: 1) obtaining the particular cropped facial image; 2) inputting the particular cropped facial image to the second deep learning model to generate a second prediction result, based on an image material feature of the particular cropped facial image, of whether particular cropped facial image is a live facial image; and 3) adjusting the second deep learning model based on the second prediction result and the second label. From 604, method 600 proceeds to 606.

[00100] At 606, a second deep learning model is trained based on the cropped facial images. From 606, method 600 proceeds to 608.

[00101] At 608, a face liveness detection is performed based on the trained first deep learning model and the trained second deep learning model. In some implementations, the first deep learning model and the second deep learning model are based on a convolutional neural network, and wherein a convolution kernel of a convolution unit in a convolutional neural network of the first deep learning model is greater than a convolution kernel of a convolution unit in a convolutional neural network of the second deep learning model. After 608, method 600 stops.

[00102] In some implementations, the face liveness detection includes: 1) obtaining a general facial image; 2) inputting the general facial image into the trained first deep learning model to obtain corresponding first prediction data; 3) obtaining a cropped facial image from the general facial image; 4) inputting the cropped facial image into the trained second deep learning model to obtain corresponding second prediction data; and 5) making a joint decision based on the first prediction data and the second prediction data to obtain a face liveness detection result. From 608, method 600 proceeds to 610.

[00103] Implementations of the subject matter described in this specification can be implemented so as to realize particular advantages or technical effects. The described face liveness detection can be used to enhance authentication processes and to ensure data security. For example, the described method can be used to distinguish between images of a live and non-live human face to help avoid fraud and malicious behavior with respect to secured data. The described method can be incorporated into computing devices (such as, mobile computing devices and digital imaging devices).

[00104] The face liveness result can be displayed on a graphical user interface. Based on the face liveness result, a determination of whether to perform subsequent actions (for example, unlocking secured data, operating a software application, storing data, sending data across a network, or displaying data on a graphical user interface).

[00105] The described methodology permits enhancement of various mobile computing device transactions and overall transaction/data security. Participants in transactions using mobile computing devices can be confident that facial images used to unlock a mobile computing device or to authorize a transaction are valid and that they will not be victims of fraud.

[00106] The described methodology can ensure the efficient usage of computer resources (for example, processing cycles, network bandwidth, and memory usage), through the efficient verification of data/transactions. At least these actions can minimize or prevent

waste of available computer resources with respect to multiple parties in a mobile computing transactions by preventing undesired/fraudulent transactions. Instead of users needing to verify data with additional research or transactions, transactions can be depended upon as valid.

[00107] In some implementations, a graphical user interface can be analyzed to ensure that graphical elements used in face liveness detection operations (for example, scanning and verification of the liveness of a human face with a mobile computing device) can be positioned on graphical user interfaces to be least obtrusive for a user (for example, to obscure the least amount of data and to avoid covering any critical or often-used graphical user interface elements).

[00108] Embodiments and the operations described in this specification can be implemented in digital electronic circuitry, or in computer software, firmware, or hardware, including the structures disclosed in this specification or in combinations of one or more of them. The operations can be implemented as operations performed by a data processing apparatus on data stored on one or more computer-readable storage devices or received from other sources. A data processing apparatus, computer, or computing device may encompass apparatus, devices, and machines for processing data, including by way of example a programmable processor, a computer, a system on a chip, or multiple ones, or combinations, of the foregoing. The apparatus can include special purpose logic circuitry, for example, a central processing unit (CPU), a field programmable gate array (FPGA) or an application-specific integrated circuit (ASIC). The apparatus can also include code that creates an execution environment for the computer program in question, for example, code that constitutes processor firmware, a protocol stack, a database management system, an operating system (for example an operating system or a combination of operating systems), a cross-platform runtime environment, a virtual machine, or a combination of one or more of them. The apparatus and execution environment can realize various different computing model infrastructures, such as web services, distributed computing and grid computing infrastructures.

[00109] A computer program (also known, for example, as a program, software, software application, software module, software unit, script, or code) can be written in any form of programming language, including compiled or interpreted languages, declarative or procedural languages, and it can be deployed in any form, including as a stand-alone program or as a module, component, subroutine, object, or other unit suitable for use in a computing

environment. A program can be stored in a portion of a file that holds other programs or data (for example, one or more scripts stored in a markup language document), in a single file dedicated to the program in question, or in multiple coordinated files (for example, files that store one or more modules, sub-programs, or portions of code). A computer program can be executed on one computer or on multiple computers that are located at one site or distributed across multiple sites and interconnected by a communication network.

[00110] Processors for execution of a computer program include, by way of example, both general- and special-purpose microprocessors, and any one or more processors of any kind of digital computer. Generally, a processor will receive instructions and data from a read-only memory or a random-access memory or both. The essential elements of a computer are a processor for performing actions in accordance with instructions and one or more memory devices for storing instructions and data. Generally, a computer will also include, or be operatively coupled to receive data from or transfer data to, or both, one or more mass storage devices for storing data. A computer can be embedded in another device, for example, a mobile device, a personal digital assistant (PDA), a game console, a Global Positioning System (GPS) receiver, or a portable storage device. Devices suitable for storing computer program instructions and data include non-volatile memory, media and memory devices, including, by way of example, semiconductor memory devices, magnetic disks, and magneto-optical disks. The processor and the memory can be supplemented by, or incorporated in, special-purpose logic circuitry.

[00111] Mobile devices can include handsets, user equipment (UE), mobile telephones (for example, smartphones), tablets, wearable devices (for example, smart watches and smart eyeglasses), implanted devices within the human body (for example, biosensors, cochlear implants), or other types of mobile devices. The mobile devices can communicate wirelessly (for example, using radio frequency (RF) signals) to various communication networks (described below). The mobile devices can include sensors for determining characteristics of the mobile device's current environment. The sensors can include cameras, microphones, proximity sensors, GPS sensors, motion sensors, accelerometers, ambient light sensors, moisture sensors, gyroscopes, compasses, barometers, fingerprint sensors, facial recognition systems, RF sensors (for example, Wi-Fi and cellular radios), thermal sensors, or other types of sensors. For example, the cameras can include a forward- or rear-facing camera with movable or fixed lenses, a flash, an image sensor, and an image processor. The camera can be a megapixel camera capable of capturing details for facial and/or iris recognition. The camera

along with a data processor and authentication information stored in memory or accessed remotely can form a facial recognition system. The facial recognition system or one-or-more sensors, for example, microphones, motion sensors, accelerometers, GPS sensors, or RF sensors, can be used for user authentication.

[00112] To provide for interaction with a user, embodiments can be implemented on a computer having a display device and an input device, for example, a liquid crystal display (LCD) or organic light-emitting diode (OLED)/virtual-reality (VR)/augmented-reality (AR) display for displaying information to the user and a touchscreen, keyboard, and a pointing device by which the user can provide input to the computer. Other kinds of devices can be used to provide for interaction with a user as well; for example, feedback provided to the user can be any form of sensory feedback, for example, visual feedback, auditory feedback, or tactile feedback; and input from the user can be received in any form, including acoustic, speech, or tactile input. In addition, a computer can interact with a user by sending documents to and receiving documents from a device that is used by the user; for example, by sending web pages to a web browser on a user's client device in response to requests received from the web browser.

[00113] Embodiments can be implemented using computing devices interconnected by any form or medium of wireline or wireless digital data communication (or combination thereof), for example, a communication network. Examples of interconnected devices are a client and a server generally remote from each other that typically interact through a communication network. A client, for example, a mobile device, can carry out transactions itself, with a server, or through a server, for example, performing buy, sell, pay, give, send, or loan transactions, or authorizing the same. Such transactions may be in real time such that an action and a response are temporally proximate; for example an individual perceives the action and the response occurring substantially simultaneously, the time difference for a response following the individual's action is less than 1 millisecond (ms) or less than 1 second (s), or the response is without intentional delay taking into account processing limitations of the system.

[00114] Examples of communication networks include a local area network (LAN), a radio access network (RAN), a metropolitan area network (MAN), and a wide area network (WAN). The communication network can include all or a portion of the Internet, another communication network, or a combination of communication networks. Information can be transmitted on the communication network according to various protocols and standards,

including Long Term Evolution (LTE), 5G, IEEE 802, Internet Protocol (IP), or other protocols or combinations of protocols. The communication network can transmit voice, video, biometric, or authentication data, or other information between the connected computing devices.

Features described as separate implementations may be implemented, in combination, in a single implementation, while features described as a single implementation may be implemented in multiple implementations, separately, or in any suitable sub-combination. Operations described in a particular order should not be understood as requiring that the particular order, nor that all illustrated operations must be performed (some operations can be optional). As appropriate, multitasking or parallel-processing (or a combination of multitasking and parallel-processing) can be performed.

CLAIMS

1. A face recognition method for determining whether an image that includes a face is a live image or a non-live image, the method comprising:

training a first deep learning model by supervised training on a plurality of general facial images, the general facial images comprising live facial images collected by shooting a live face and labeled as positive samples, and non-live facial images collected by shooting a non-live face that is a face picture or a face model and labeled as negative samples;

training a plurality of second deep learning models by supervised training on a plurality of extracted facial images cropped from the general facial images, the second deep learning models comprising an eye deep learning model and a nose deep learning model corresponding to an eye and a nose type of facial region respectively, the extracted facial images comprising live facial images labeled as positive samples and non-live facial images and labeled as negative samples, wherein the first deep learning model and each of the second deep learning models are classification models and wherein after training the models classify facial images into a live facial image category or a non-live facial image category;

performing face liveness detection on a first general facial image using the trained first deep learning model to obtain a first prediction score and the plurality of trained second deep learning models to obtain second prediction scores, comprising:

obtaining the first general facial image collected for the face liveness detection,

inputting the first general facial image into the trained first deep learning model for processing to obtain the first prediction score,

obtaining a plurality of extracted facial images cropped from the first general facial image, the extracted facial images comprising an eye image region image and a nose image region image, and inputting the extracted facial images into respective trained second deep learning models for processing, the second deep learning models comprising the eye deep learning model and the nose deep learning model, to obtain the second prediction scores,

generating a prediction score result based on the first prediction score and the second prediction scores, and

comparing the prediction score result with a threshold to determine whether the first general facial image is a live image or a non-live image.

2. The method according to claim 1, wherein the first deep learning model and a second deep learning model use different deep learning network structures.

3. The method according to claim 2, wherein the different deep learning network structures comprise at least one of the following parameters: a quantity of hidden variable layers, a type of a hidden variable layer, a quantity of neuron nodes, or a size of a convolution kernel of a convolution unit.

4. The method according to any one of claims 1 to 3, wherein generating a prediction score result based on the first prediction score and the second prediction scores comprises generating the prediction score result as a sum of the first prediction score and the second prediction scores.

5. The method according to any one of claims 1 to 3, wherein the first deep learning model and the second deep learning model are based on a convolutional neural network; and

a convolution kernel of a convolution unit in a convolutional neural network that the first deep learning model is based on is relatively large so that the first deep learning model extracts an image structure feature of a general facial image, and a convolution kernel of a convolution unit in a convolutional neural network that the second deep learning model is based on is relatively small, so that the second deep learning model extracts an image material feature of the extracted facial image.

6. The method according to any one of claims 1 to 5, wherein the prediction scores are all a probability value or a Boolean value.

7. An apparatus for face recognition, comprising a plurality of modules configured to perform the method of any one of claims 1 to 6.

8. An electronic device for face recognition, comprising:

at least one processor; and

a memory communicatively connected to the at least one processor, wherein the memory stores an instruction that can be executed by the at least one processor, and the instruction is executed by the at least one processor, to enable the at least one processor to perform the method of any one of claims 1 to 6.

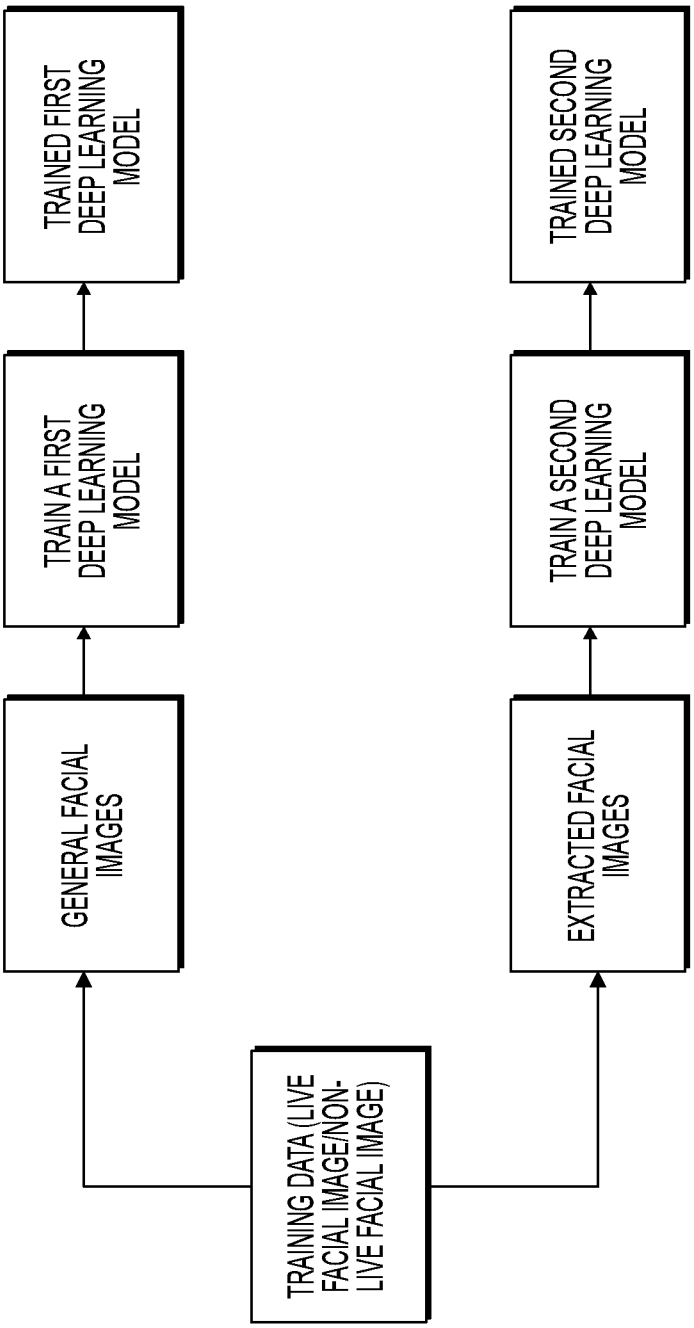

TRAINED FIRST DEEP LEARNING MODEL
TRAIN A FIRST DEEP LEARNING MODEL
GENERAL FACIAL IMAGES
TRAINED SECOND DEEP LEARNING MODEL
TRAIN A SECOND DEEP LEARNING MODEL
EXTRACTED FACIAL IMAGES
TRAINING DATA (LIVE FACIAL IMAGE/NON-LIVE FACIAL IMAGE)

FIG. 1

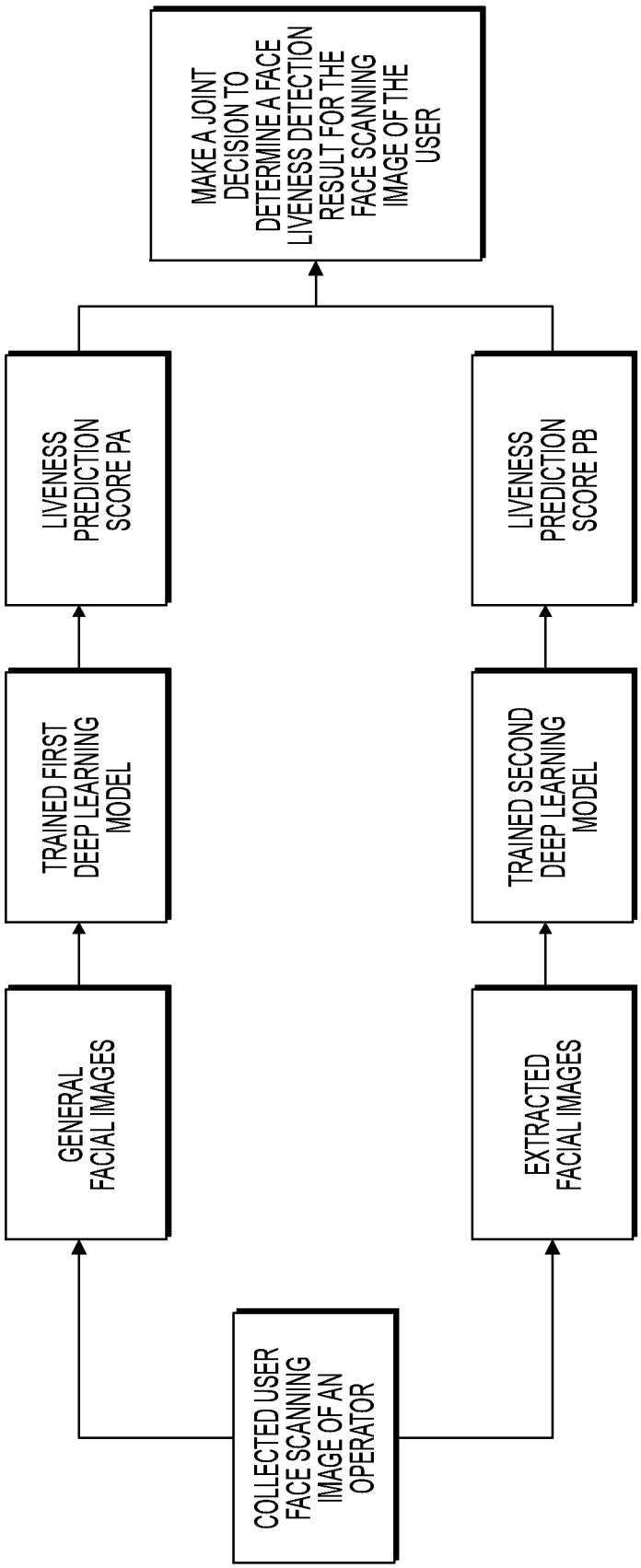

COLLECTED USER FACE SCANNING IMAGE OF AN OPERATOR
GENERAL FACIAL IMAGES
EXTRACTED FACIAL IMAGES
TRAINED FIRST DEEP LEARNING MODEL
TRAINED SECOND DEEP LEARNING MODEL
LIVENESS PREDICTION SCORE PA
LIVENESS PREDICTION SCORE PB
MAKE A JOINT DECISION TO DETERMINE A FACE LIVENESS DETECTION RESULT FOR THE FACE SCANNING IMAGE OF THE USER

FIG. 2

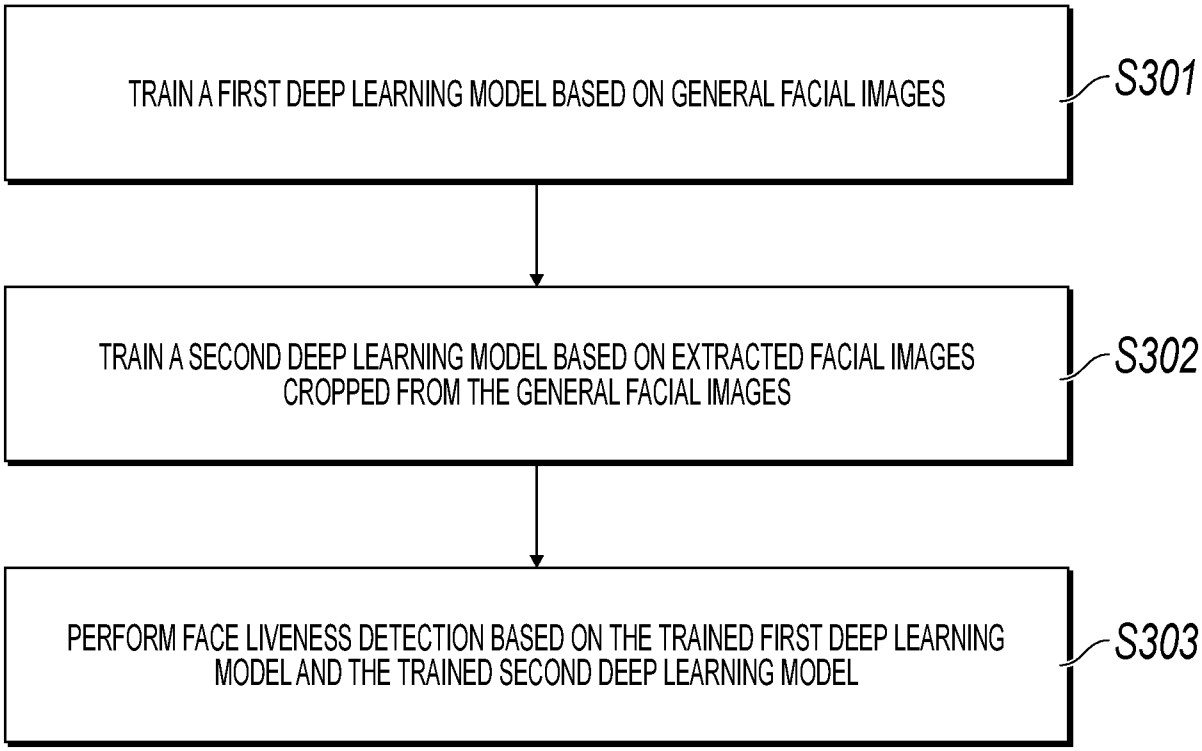

TRAIN A FIRST DEEP LEARNING MODEL BASED ON GENERAL FACIAL IMAGES
S301
TRAIN A SECOND DEEP LEARNING MODEL BASED ON EXTRACTED FACIAL IMAGES CROPPED FROM THE GENERAL FACIAL IMAGES
S302
PERFORM FACE LIVENESS DETECTION BASED ON THE TRAINED FIRST DEEP LEARNING MODEL AND THE TRAINED SECOND DEEP LEARNING MODEL
S303

FIG. 3

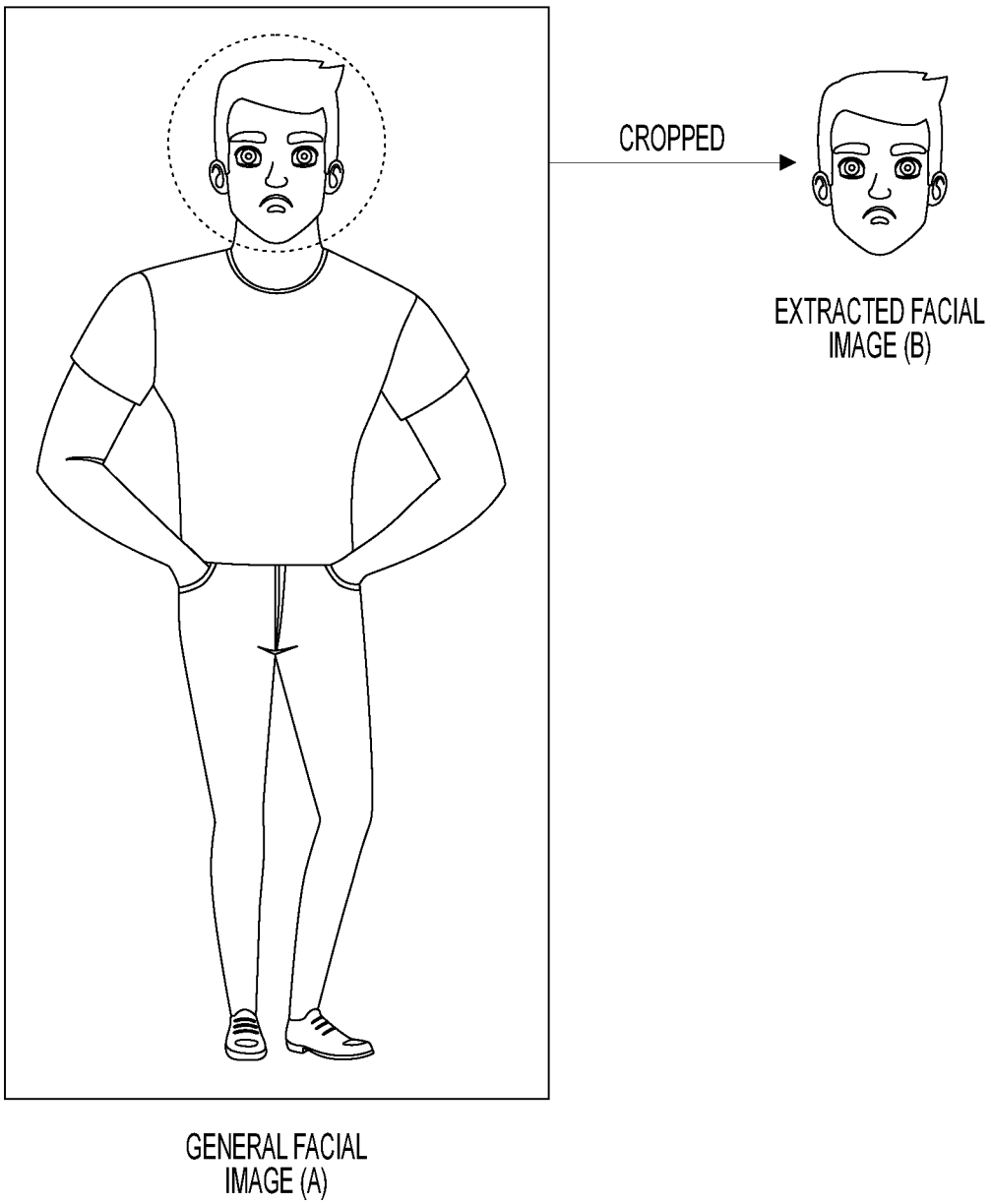


FIG. 4

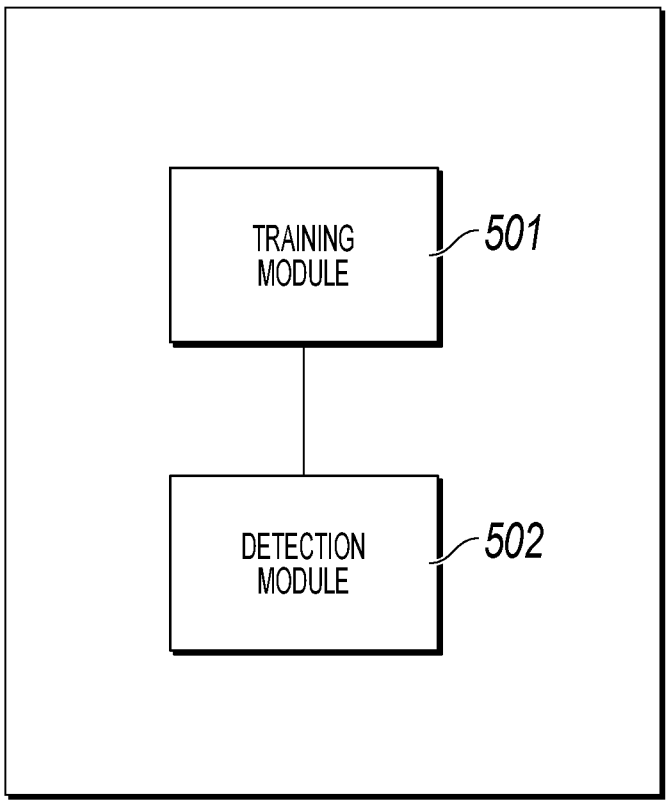


FIG. 5

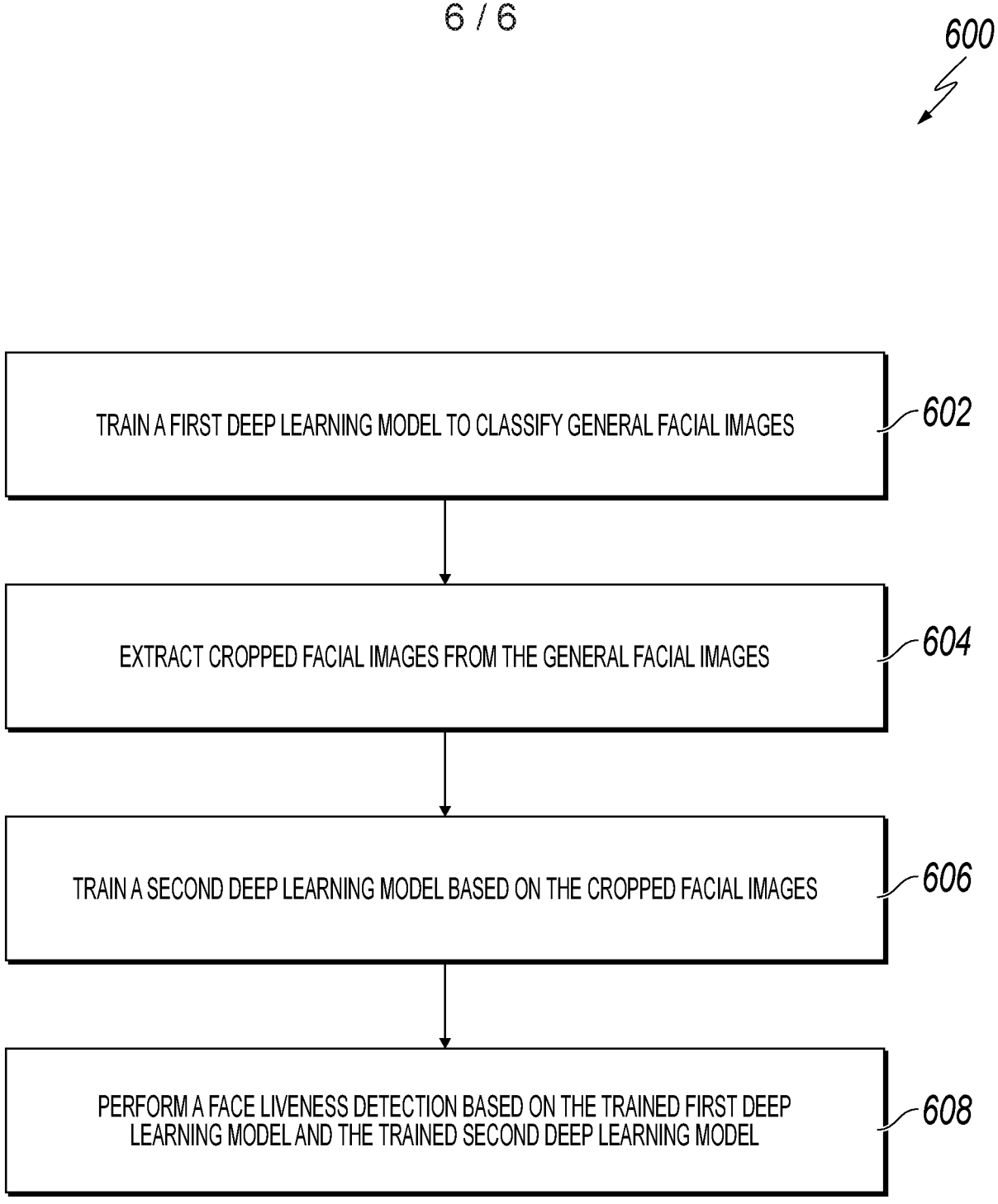
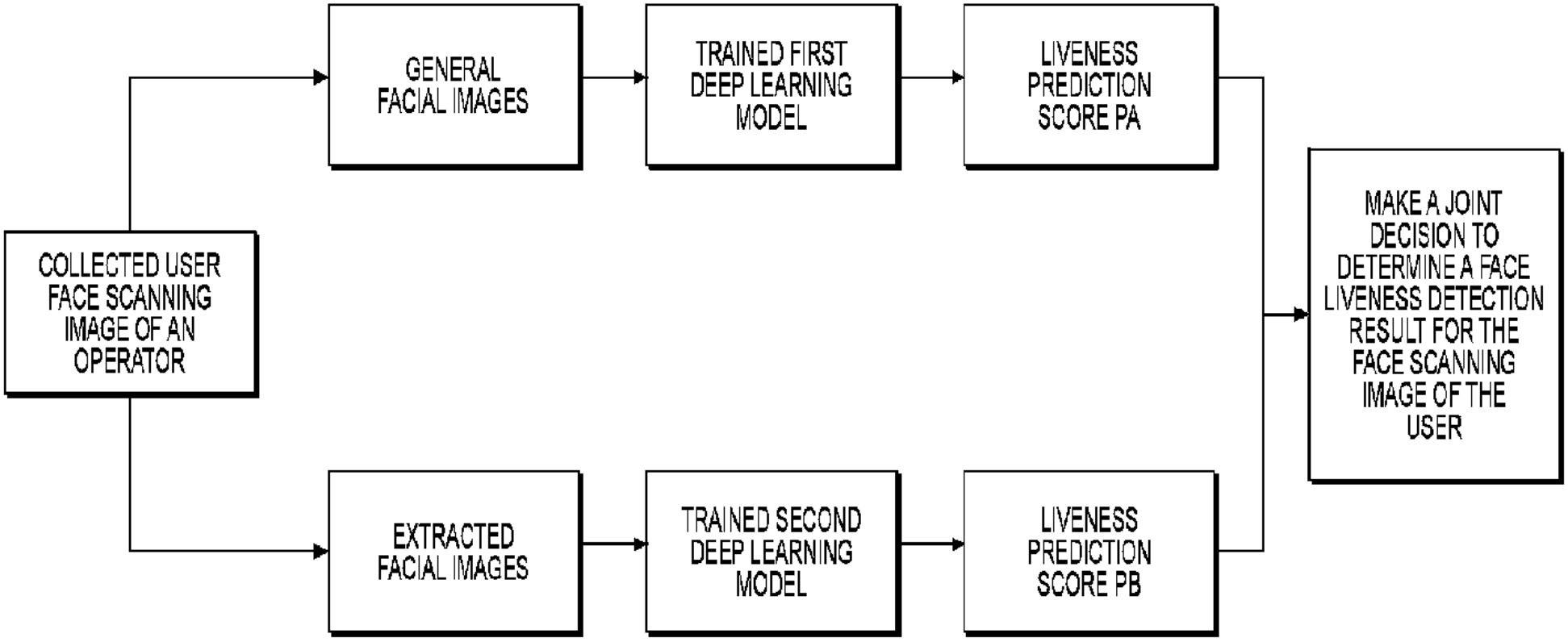

600
TRAIN A FIRST DEEP LEARNING MODEL TO CLASSIFY GENERAL FACIAL IMAGES
602
EXTRACT CROPPED FACIAL IMAGES FROM THE GENERAL FACIAL IMAGES
604
TRAIN A SECOND DEEP LEARNING MODEL BASED ON THE CROPPED FACIAL IMAGES
606
PERFORM A FACE LIVENESS DETECTION BASED ON THE TRAINED FIRST DEEP LEARNING MODEL AND THE TRAINED SECOND DEEP LEARNING MODEL
608

FIG. 6


COLLECTED USER FACE SCANNING IMAGE OF AN OPERATOR
GENERAL FACIAL IMAGES
TRAINED FIRST DEEP LEARNING MODEL
LIVENESS PREDICTION SCORE PA
EXTRACTED FACIAL IMAGES
TRAINED SECOND DEEP LEARNING MODEL
LIVENESS PREDICTION SCORE PB
MAKE A JOINT DECISION TO DETERMINE A FACE LIVENESS DETECTION RESULT FOR THE FACE SCANNING IMAGE OF THE USER