

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 878 374**

51 Int. Cl.:

G06K 9/00

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **07.06.2018 PCT/US2018/036505**

87 Fecha y número de publicación internacional: **13.12.2018 WO18226990**

96 Fecha de presentación y número de la solicitud europea: **07.06.2018 E 18735070 (7)**

97 Fecha y número de publicación de la concesión europea: **14.04.2021 EP 3523754**

54 Título: **Método y aparato de detección de vivacidad facial y dispositivo electrónico**

30 Prioridad:

07.06.2017 CN 201710421333

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

18.11.2021

73 Titular/es:

**ADVANCED NEW TECHNOLOGIES CO., LTD.
(100.0%)**

**Cayman Corporate Centre, 27 Hospital Road
George Town, Grand Cayman KY1-9008, KY**

72 Inventor/es:

MA, CHENGUANG

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 878 374 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Método y aparato de detección de vivacidad facial y dispositivo electrónico

5 Esta solicitud reivindica la prioridad de la Solicitud de Patente China No. 201710421333.5, presentada el 7 de junio de 2017.

CAMPO TÉCNICO

10 La presente patente se refiere al campo de las tecnologías de software informático y, en particular, a un método, aparato y dispositivo electrónico de detección de vivacidad facial.

ANTECEDENTES

15 Se utiliza una tecnología de detección de vivacidad facial para determinar si el usuario actual es el usuario auténtico utilizando técnicas de reconocimiento facial para interceptar ataques de suplantación de identidad, tal como un ataque de reproducción de pantalla, un ataque de foto impresa y un ataque de modelado tridimensional.

20 Actualmente, la tecnología de detección de vivacidad facial se puede clasificar en una tecnología de detección de vivacidad facial intrusiva y una tecnología de detección de vivacidad facial no intrusiva. En la tecnología de detección intrusiva de vivacidad facial, un usuario debe completar cooperativamente algunas acciones en vivo específicas, como parpadear, girar la cabeza o abrir la boca. Al realizar el reconocimiento facial en base a las instrucciones dadas, el módulo de detección de vivacidad puede determinar si un operador completa con precisión la operación en vivo y si el operador es el usuario auténtico. En la tecnología de detección de vivacidad facial no intrusiva, un usuario no necesita completar cooperativamente una acción en vivo, por lo que la experiencia de usuario es mejor, pero la complejidad técnica es mayor. Además, la detección de vivacidad se realiza principalmente en función de la información acerca de una imagen de entrada de un solo cuadro o información acerca de otros sensores del dispositivo.

30 En la tecnología de detección de vivacidad facial no intrusiva descrita en la tecnología existente, el entrenamiento supervisado generalmente se realiza en un solo modelo de aprendizaje profundo utilizando imágenes faciales en vivo y no en vivo, y luego la predicción de vivacidad facial se realiza en la imagen de entrada de un solo cuadro utilizando el modelo entrenado.

35 Sin embargo, una solución técnica de este tipo depende en gran medida de un tipo de ataque facial de suplantación de los datos de entrenamiento, y está limitada por una condición objetiva de datos de entrenamiento insuficientes. Es difícil extraer completamente una característica de imagen facial en vivo. Como resultado, este modelo no puede expresar completamente una característica facial en vivo y se reduce la precisión del resultado de detección de vivacidad facial. Jianwei Yang et al., "Learn Convolutional Neural Network for Face Anti-Spoofing", 24 de agosto de 2014, describe el uso de una red neuronal convolucional profunda (CNN) para aprender características de alta capacidad discriminativa de manera supervisada. Combinado con el procesamiento previo, el rendimiento anti-suplantación facial mejora drásticamente. En los experimentos descritos, se logra una disminución relativa de más del 70% de la mitad de la tasa de error total (HTER) en dos conjuntos de datos de desafío en comparación con las técnicas más avanzadas. Los resultados experimentales de las pruebas entre dos conjuntos de datos indican que la CNN puede obtener características con una mejor capacidad de generalización. Además, las redes entrenadas con datos combinados de dos conjuntos de datos presentan menos sesgos entre dos conjuntos de datos. Menotti, David, et al., "Deep Representations for Iris, Face, and Fingerprint Spoofing Detection", 1 de abril de 2015, describe sistemas de detección de suplantación para modalidades de iris, rostro y huellas dactilares en base a dos enfoques de aprendizaje profundo. El primer enfoque consiste en aprender arquitecturas de red convolucionales adecuadas para cada uno de los dominios. El segundo enfoque se centra en aprender los pesos de la red a través de la propagación hacia atrás. Se consideran nueve puntos de referencia de suplantación biométrica, cada uno de los cuales contiene muestras reales y falsas de una modalidad biométrica y tipo de ataque determinados, y se aprenden representaciones profundas de cada punto de referencia combinando y se contrastan los dos enfoques de aprendizaje.

RESUMEN

55 La invención está definida por las reivindicaciones adjuntas. Las realizaciones de la presente patente proporcionan un método, un aparato y un dispositivo electrónico de detección de vivacidad facial para resolver los siguientes problemas técnicos en la tecnología existente. En una solución técnica basada en un único modelo de aprendizaje profundo, es difícil extraer por completo una característica de imagen facial en vivo. Como resultado, este modelo no puede expresar completamente una característica facial en vivo y se reduce la precisión del resultado de detección de vivacidad facial.

60 Para resolver los problemas técnicos descritos, las realizaciones de la presente patente se implementan de la siguiente manera:

65 Una realización de la presente patente proporciona un método para determinar si una imagen que incluye un rostro es una imagen en vivo o una imagen no en vivo, el método incluye: entrenar un primer modelo de aprendizaje profundo

mediante entrenamiento supervisado en una pluralidad de imágenes faciales generales, las imágenes faciales generales que comprenden imágenes faciales en vivo recopiladas mediante la toma de un rostro en vivo y etiquetadas como muestras positivas, e imágenes faciales no en vivo recopiladas mediante la toma de un rostro que es una imagen facial o un modelo facial y etiquetadas como muestras negativas; entrenar una pluralidad de segundos modelos de aprendizaje profundo mediante entrenamiento supervisado en una pluralidad de imágenes faciales extraídas recortadas de las imágenes faciales generales, los segundos modelos de aprendizaje profundo que comprenden un modelo de aprendizaje profundo de la región facial completa, un modelo de aprendizaje profundo del ojo y un modelo de aprendizaje profundo de la nariz, los dos últimos corresponden a un tipo de región facial de ojo y nariz respectivamente, las imágenes faciales extraídas comprenden imágenes faciales en vivo etiquetadas como muestras positivas e imágenes faciales no en vivo y etiquetadas como muestras negativas, en donde el primer modelo de aprendizaje profundo y cada de los segundos modelos de aprendizaje profundo son modelos de clasificación y en donde, después del entrenamiento, los modelos clasifican las imágenes faciales en una categoría de imagen facial en vivo o una categoría de imagen facial no en vivo; y realizar la detección de vivacidad facial en una primera imagen facial general utilizando el primer modelo de aprendizaje profundo entrenado para obtener una primera puntuación de predicción y la pluralidad de segundos modelos de aprendizaje profundo entrenados para obtener una segunda puntuación de predicción, que comprende: obtener la primera imagen facial general recopilada para el detección de vivacidad facial; introducir la primera imagen facial general en el primer modelo de aprendizaje profundo entrenado para procesar y obtener la primera puntuación de predicción; obtener una pluralidad de imágenes faciales extraídas recortadas de la primera imagen facial general, las imágenes faciales extraídas que comprenden una región de la imagen facial completa, una imagen de región de imagen del ojo y una imagen de región de imagen de la nariz, e ingresar las imágenes faciales extraídas en los respectivos segundos modelos de aprendizaje profundo para el procesamiento, los segundos modelos de aprendizaje profundo que comprenden todo el modelo de aprendizaje profundo de imagen facial, el modelo de aprendizaje profundo del ojo y el modelo de aprendizaje profundo de la nariz, para obtener las segundas puntuaciones de predicción; generar un resultado de puntuación de predicción en base a la primera puntuación de predicción y la segunda puntuación de predicción; y comparar el resultado de puntuación de predicción con un umbral para determinar si la primera imagen facial general es una imagen en vivo o una imagen no en vivo.

Una realización de la presente patente proporciona un aparato que comprende una pluralidad de módulos (501, 502) configurados para realizar el método para determinar si una imagen que incluye un rostro es una imagen en vivo o una imagen no en vivo.. Una realización de la presente patente proporciona un dispositivo electrónico, que incluye al menos un procesador; y una memoria comunicativamente conectada al por lo menos un procesador, en donde la memoria almacena una instrucción que puede ejecutarse por el al menos un procesador, y la instrucción se ejecuta por el al menos un procesador, para permitir que el al menos un procesador realice el método para determinar si una imagen que incluye un rostro es una imagen en vivo o una imagen no en vivo

Al menos una solución técnica utilizada en las realizaciones de la presente patente puede lograr los siguientes efectos beneficiosos. Uno de esos beneficios es que se extraen más características de imagen facial en vivo. En comparación con un modelo en la tecnología existente, el primer modelo de aprendizaje profundo entrenado y el segundo modelo de aprendizaje profundo entrenado expresan mejor en conjunto la característica facial en vivo, mejorando así la precisión del resultado de detección de vivacidad facial. Por lo tanto, se puede resolver una parte o la totalidad de los problemas de la tecnología existente.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Para describir las soluciones técnicas en las realizaciones de la presente patente o en la tecnología existente más claramente, a continuación, se presentan brevemente los dibujos adjuntos necesarios para describir las realizaciones o la tecnología existente. Aparentemente, los dibujos adjuntos en la siguiente descripción simplemente muestran algunas realizaciones de la presente patente, y una persona con experiencia normal en la técnica todavía puede derivar sin esfuerzos creativos otros dibujos a partir de estos dibujos adjuntos.

La FIG. 1 es un diagrama de flujo esquemático que ilustra un ejemplo de una etapa de entrenamiento de modelo;

La FIG. 2 es un diagrama de flujo esquemático que ilustra un ejemplo de una etapa de detección de vivacidad;

La FIG. 3 es un diagrama de flujo esquemático que ilustra un método de detección de vivacidad facial de acuerdo con una realización de la presente patente;

La FIG. 4 es un diagrama esquemático que ilustra la comparación entre una imagen facial general y una imagen facial extraída de acuerdo con una realización de la presente patente;

La FIG. 5 es un diagrama estructural esquemático que ilustra un aparato de detección de vivacidad facial correspondiente a la FIG. 3 de acuerdo con una realización de la presente patente; y

La FIG. 6 es un diagrama de flujo que ilustra un ejemplo de un método implementado por computadora para determinar la autenticidad del usuario con detección de vivacidad facial.

DESCRIPCIÓN DE LAS REALIZACIONES

5 Las realizaciones de la presente patente proporcionan un método, aparato y dispositivo electrónico de detección de vivacidad facial.

10 Para que un experto en la técnica comprenda mejor las soluciones técnicas de la presente patente, a continuación, se describen clara y completamente las soluciones técnicas en las realizaciones de la presente patente con referencia a los dibujos adjuntos en las realizaciones de la presente patente. Aparentemente, las realizaciones descritas son simplemente una parte y no todas las realizaciones de la presente patente.

15 Todos los modelos de aprendizaje profundo se basan en una red neuronal. Para facilitar la descripción, en primer lugar, se describe una idea central basada en un ejemplo y con referencia a la FIG. 1 y FIG. 2.

Este ejemplo se puede clasificar en una etapa de entrenamiento de modelo y una etapa de detección de vivacidad.

20 La FIG. 1 es un diagrama de flujo esquemático que ilustra un ejemplo de una etapa de entrenamiento del modelo. En una etapa de entrenamiento del modelo, se entrenan dos modelos de aprendizaje profundo independientes utilizando muestras en vivo y no en vivo (que pertenecen a un conjunto de datos de entrenamiento) en una imagen facial: un primer modelo de aprendizaje profundo y un segundo modelo de aprendizaje profundo. Una imagen de entrada del primer modelo de aprendizaje profundo es una imagen facial general recopilada, y una imagen de entrada del segundo modelo de aprendizaje profundo puede ser una imagen facial extraída recortada de la imagen facial general. El primer modelo de aprendizaje profundo y el segundo modelo de aprendizaje profundo pueden utilizar diferentes estructuras de red de aprendizaje profundo (es decir, una estructura de una red neuronal en la que se basa un modelo). Las diferentes estructuras de red son sensibles de manera diferente a las diferentes características de imagen. Los conjuntos de datos de entrenamiento en vivo y no en vivo se utilizan para completar el entrenamiento del primer modelo de aprendizaje profundo y el segundo modelo de aprendizaje profundo en base a un método de aprendizaje profundo.

25 La FIG. 2 es un diagrama de flujo esquemático que ilustra un ejemplo de una etapa de detección de vivacidad. En una etapa de detección de vivacidad, se recopila una imagen de escaneo facial de un usuario como una imagen facial general del usuario, y se ingresa un primer modelo de aprendizaje profundo para obtener una puntuación PA de predicción. Además, la detección facial se realiza en la imagen de escaneo facial del usuario, se recorta una imagen facial extraída de la imagen de escaneo facial del usuario en base a un resultado de detección, y se ingresa un segundo modelo de aprendizaje profundo a la imagen facial extraída para obtener una puntuación PB de predicción. Posteriormente, por ejemplo, un resultado de puntuación de predicción de (PA + PB) se puede comparar con un umbral determinado (p. ej., el umbral puede ser 1), para tomar una decisión conjunta para determinar un resultado de detección de vivacidad facial para la imagen de escaneo facial del usuario.

En base a la idea central descrita, a continuación, se describe la solución de la presente patente en detalle.

30 La FIG. 3 es un diagrama de flujo esquemático que ilustra un método de detección de vivacidad facial de acuerdo con una realización de la presente patente. Desde la perspectiva de un programa, el procedimiento puede ejecutarse por un programa en un servidor o un terminal, por ejemplo, un programa de autenticación de identidad o una aplicación de comercio electrónico. Desde la perspectiva de un dispositivo, el procedimiento se ejecuta por al menos uno de los siguientes dispositivos que se pueden utilizar como servidor o terminal: un dispositivo de control de acceso, una computadora personal, una computadora mediana, un grupo de computadoras, un teléfono móvil, una computadora tableta, un dispositivo ponible inteligente, una máquina de automóvil o un punto de venta (POS).

El procedimiento de la FIG. 3 incluye los siguientes pasos.

35 S301. Entrenar un primer modelo de aprendizaje profundo en base a imágenes faciales generales.

40 En esta realización de la presente patente, las imágenes faciales generales utilizadas para entrenar el primer modelo de aprendizaje profundo pueden incluir una pluralidad de muestras. En la pluralidad de muestras, algunas son imágenes faciales en vivo que se recopilan al tomar un rostro en vivo y que se pueden utilizar como muestras positivas, y algunas son imágenes faciales no en vivo que se recopilan al tomar un rostro no en vivo, tal como una imagen facial o un modelo facial y que se pueden utilizar como muestras negativas.

45 En esta realización de la presente patente, el primer modelo de aprendizaje profundo es un modelo de clasificación, y las imágenes faciales generales se utilizan como entradas del modelo de clasificación. Después del procesamiento del modelo, las imágenes faciales generales se pueden clasificar en al menos la categoría de imagen facial en vivo o la categoría de imagen facial no en vivo. Un objetivo del entrenamiento del primer modelo de aprendizaje profundo es mejorar la precisión de clasificación del primer modelo de aprendizaje profundo.

S302. Entrenar un segundo modelo de aprendizaje profundo en base a imágenes faciales extraídas recortadas de las imágenes faciales generales.

5 En esta realización de la presente patente, además de una región facial completa, la imagen facial general incluye generalmente algunas regiones no relacionadas, tales como una región de fondo y un cuerpo humano excepto una cara. La imagen facial extraída puede excluir las regiones no relacionadas e incluye al menos una región facial extraída, por ejemplo, una región facial completa, una región ocular y una región nasal. Hay una pluralidad de segundos modelos de aprendizaje profundo, y cada uno de los segundos modelos de aprendizaje profundo puede corresponder a un tipo de regiones faciales.

La FIG. 4 es un diagrama esquemático que ilustra la comparación entre una imagen facial general y una imagen facial extraída de acuerdo con una realización de la presente patente.

15 En la FIG. 4, (a) es una imagen facial general. Para facilitar la comprensión, una imagen facial extraída se marca en (a) utilizando líneas discontinuas, y (a) se puede recortar correspondientemente para obtener una imagen facial extraída mostrada en (b).

Además, cuando la imagen facial extraída es una imagen que incluye solo una región facial parcial, la imagen facial general también puede ser una imagen que incluye una región facial completa y básicamente excluye una región no relacionada.

En esta realización de la presente patente, la imagen facial extraída utilizada para entrenar los segundos modelos de aprendizaje profundo también puede incluir una variedad de muestras. En la variedad de muestras, algunas son imágenes faciales en vivo que pueden utilizarse como muestras positivas, y algunas son imágenes faciales no en vivo que pueden utilizarse como muestras negativas.

En esta realización de la presente patente, los segundos modelos de aprendizaje profundo también son modelos de clasificación, y las imágenes faciales extraídas se utilizan como entrada del modelo de clasificación. Después del procesamiento del modelo, las imágenes faciales extraídas se clasifican al menos en la categoría de imagen facial en vivo o en la categoría de imagen facial no en vivo. Un objetivo del entrenamiento del segundo modelo de aprendizaje profundo es mejorar la precisión de clasificación del segundo modelo de aprendizaje profundo.

Además de recortarse de la imagen facial general, la imagen facial extraída se puede obtener mediante una colección especial sin depender de la imagen facial general.

En esta realización de la presente patente, el primer modelo de aprendizaje profundo y el segundo modelo de aprendizaje profundo pueden ser modelos diferentes o un mismo modelo antes del entrenamiento.

40 Una secuencia de ejecución del paso S301 y del paso S302 no está limitada en la presente patente, y el paso S301 y el paso S302 pueden realizarse simultánea o sucesivamente.

S303. Realizar la detección de vivacidad facial en base al primer modelo de aprendizaje profundo entrenado y al segundo modelo de aprendizaje profundo entrenado.

45 Cada uno de los pasos en la FIG. 3 puede realizarse mediante un mismo dispositivo o un mismo programa, o puede realizarse mediante diferentes dispositivos o diferentes programas. Por ejemplo, del paso S301 al paso S303 se realizan por un dispositivo 1. Para dar otro ejemplo, tanto el paso S301 como el paso S302 se realizan por un dispositivo 1, y el paso S303 se realiza por un dispositivo 2; etc.

50 De acuerdo con el método en la FIG. 3, se extraen más características de imagen facial en vivo. En comparación con un modelo en la tecnología existente, el primer modelo de aprendizaje profundo entrenado y el segundo modelo de aprendizaje profundo entrenado expresan mejor en conjunto una característica facial en vivo, mejorando así la precisión de un resultado de detección de vivacidad facial. Por lo tanto, se puede resolver una parte o la totalidad de los problemas de la tecnología existente.

En base al método en la FIG. 3, esta realización de la presente patente proporciona además algunas soluciones de implementación específicas del método y una solución de extensión, que se describen a continuación.

60 En esta realización de la presente patente, para implementar una diferencia entre la sensibilidad del primer modelo de aprendizaje profundo a una característica de imagen y la sensibilidad del segundo modelo de aprendizaje profundo a una característica de imagen, el primer modelo de aprendizaje profundo y el segundo modelo de aprendizaje profundo pueden, preferiblemente, utilizar diferentes estructuras de red de aprendizaje profundo.

65 Las diferentes estructuras de red de dos modelos de aprendizaje profundo pueden indicar que los dos modelos de aprendizaje profundo incluyen uno o más parámetros de estructura de red diferentes. El parámetro de estructura de

red puede incluir, por ejemplo, una cantidad de capas variables ocultas, un tipo de capa variable oculta, una cantidad de nodos neuronales, una cantidad de nodos de la capa de entrada o una cantidad de nodos de la capa de salida.

5 Ciertamente, algunos modelos de aprendizaje profundo específicos también pueden incluir parámetros específicos correspondientes. Por ejemplo, para un modelo de aprendizaje profundo basado en una red neuronal convolucional ampliamente utilizada en el campo de la imagen actualmente, el tamaño de un núcleo de convolución de una unidad de convolución también es un parámetro de estructura de red específico de este modelo de aprendizaje profundo.

10 Para la solución de la presente patente, en general, las diferentes estructuras de red de aprendizaje profundo incluyen al menos uno de los siguientes parámetros: una cantidad de capas variables ocultas, un tipo de capa variable oculta, una cantidad de nodos neuronales o un tamaño de un núcleo de convolución de una unidad de convolución.

15 En esta realización de la presente patente, para mejorar la eficiencia del entrenamiento del modelo y la confiabilidad del entrenamiento del modelo, el entrenamiento del modelo se realiza de una manera de entrenamiento supervisada.

Por ejemplo, en una manera de entrenamiento supervisada, para el paso S301, la imagen facial general incluye una primera etiqueta, y la primera etiqueta indica si una imagen facial general correspondiente a la primera etiqueta es una imagen facial en vivo.

20 El entrenamiento de un primer modelo de aprendizaje profundo basado en una imagen facial general puede incluir: ingresar el primer modelo de aprendizaje profundo a la imagen facial general, donde el primer modelo de aprendizaje profundo extrae una característica de la imagen facial general y predice, en base a la característica extraída, si la imagen facial general es una imagen facial en vivo; y ajustar el primer modelo de aprendizaje profundo en base a un resultado de predicción y la primera etiqueta de la imagen facial general. Generalmente, cuando el resultado de la predicción es inconsistente con la primera etiqueta, se ajusta el primer modelo de aprendizaje profundo, de modo que el primer modelo de aprendizaje profundo ajustado puede obtener, mediante una nueva predicción, un resultado de predicción consistente con la primera etiqueta.

30 La característica extraída por el primer modelo de aprendizaje profundo en un proceso de entrenamiento puede incluir preferiblemente una característica de estructura de imagen de la imagen facial general, por ejemplo, un borde de foto de pantalla o distorsión facial en la imagen facial general.

35 Por dar otro ejemplo, de manera similar, en una manera de entrenamiento supervisada, para el paso S302, la imagen facial extraída incluye una segunda etiqueta, y la segunda etiqueta indica si una imagen facial extraída correspondiente a la segunda etiqueta es una imagen facial en vivo.

40 El entrenamiento de los segundos modelos de aprendizaje profundo en base a las imágenes faciales extraídas recortadas de las imágenes faciales generales puede incluir: obtener las imágenes faciales extraídas recortadas de las imágenes faciales generales; aplicar el segundo modelo de aprendizaje profundo a la imagen facial extraída obtenida, donde el segundo modelo de aprendizaje profundo extrae una característica de la imagen facial extraída y predice, en base a la característica extraída, si la imagen facial extraída es una imagen facial en vivo; y ajustar el segundo modelo de aprendizaje profundo en base a un resultado de predicción y la segunda etiqueta de la imagen facial extraída. Generalmente, cuando el resultado de la predicción es inconsistente con la segunda etiqueta, se ajusta el segundo modelo de aprendizaje profundo, de modo que el segundo modelo de aprendizaje profundo ajustado puede obtener un resultado de predicción consistente con la segunda etiqueta mediante una nueva predicción.

50 La característica extraída por el segundo modelo de aprendizaje profundo en un proceso de entrenamiento puede incluir preferiblemente una característica de material de imagen de la imagen facial extraída, por ejemplo, desenfoque, textura o distorsión de color en la imagen facial extraída.

55 En los dos ejemplos descritos anteriormente, el primer modelo de aprendizaje profundo y el segundo modelo de aprendizaje profundo son sensibles de manera diferente a las diferentes características de imagen. El primer modelo de aprendizaje profundo es más sensible a la característica de estructura de imagen, y el segundo modelo de aprendizaje profundo es más sensible a la característica de material de imagen. Para una imagen facial, la característica de estructura de imagen es una característica relativamente global y generalizada, y la característica de material de imagen es una característica relativamente local y refinada.

60 Por lo tanto, el primer modelo de aprendizaje profundo entrenado y el segundo modelo de aprendizaje profundo entrenado pueden extraer conjuntamente una característica de imagen facial de manera más jerárquica y abundante, a fin de tomar una decisión conjunta para obtener un resultado de detección de vivacidad facial más preciso.

65 En esta realización de la presente patente, los conjuntos de datos de entrenamiento correspondientes y/o las estructuras de red de aprendizaje profundo correspondientes son diferentes, de modo que el primer modelo de aprendizaje profundo y el segundo modelo de aprendizaje profundo pueden ser sensibles de manera diferente a diferentes características de imagen.

Por ejemplo, si el primer modelo de aprendizaje profundo y el segundo modelo de aprendizaje profundo se basan en una red neuronal convolucional, un núcleo de convolución de una unidad de convolución en una red neuronal convolucional en la que se basa el primer modelo de aprendizaje profundo puede ser relativamente grande, por lo que el primer modelo de aprendizaje profundo extrae una característica de estructura de imagen de la imagen facial general.

5 En consecuencia, un núcleo de convolución de una unidad de convolución en una red neuronal convolucional en la que se basa el segundo modelo de aprendizaje profundo puede ser relativamente pequeño, de modo que el segundo modelo de aprendizaje profundo extrae una característica de material de imagen de la imagen facial extraída. Por lo tanto, en este ejemplo, el núcleo de convolución de la unidad de convolución en la red neuronal convolucional en la que se basa el primer modelo de aprendizaje profundo es mayor que el núcleo de convolución de la unidad de convolución en la red neuronal convolucional en la que se basa el segundo modelo de aprendizaje profundo.

10

Hay que señalar que el tamaño del núcleo de convolución es simplemente un ejemplo de un parámetro que puede afectar a la sensibilidad, y otro parámetro de la estructura de red también puede afectar a la sensibilidad.

15 En esta realización de la presente patente, para el paso S303, el primer modelo de aprendizaje profundo entrenado y el segundo modelo de aprendizaje profundo entrenado toman conjuntamente una decisión para realizar la detección de vivacidad facial. Hay una variedad de maneras específicas de tomar decisiones. Por ejemplo, se toma una decisión separada utilizando por separado el primer modelo de aprendizaje profundo y el segundo modelo de aprendizaje profundo, y luego se determina un resultado de decisión final sintetizando todos los resultados de decisión separados.

20 Por dar otro ejemplo, se puede tomar una decisión separada primero utilizando cualquiera del primer modelo de aprendizaje profundo y del segundo modelo de aprendizaje profundo. Cuando un resultado de decisión separado satisface una condición específica, el resultado de decisión separado puede utilizarse directamente como resultado de decisión final; de lo contrario, se toma una decisión de manera integral en combinación con otro modelo restante, para obtener un resultado de decisión final; etc.

25 Si se utiliza una primera manera descrita en el párrafo anterior, un ejemplo es el siguiente:

Por ejemplo, para el paso S303, la realización de la detección de vivacidad facial en base al primer modelo de aprendizaje profundo entrenado y al segundo modelo de aprendizaje profundo entrenado incluye: obtener la imagen facial general (que generalmente es una imagen de escaneo facial de un usuario) recopilada para la detección de vivacidad facial; ingresar el primer modelo de aprendizaje profundo entrenado en la imagen facial general recopilada para su procesamiento, para obtener los primeros datos de predicción correspondientes; obtener la imagen facial extraída recortada de la imagen facial general recopilada e ingresar los segundos modelos de aprendizaje profundo entrenados para su procesamiento, para obtener los segundos datos de predicción correspondientes; y tomar una

30

35 decisión conjunta en base a los primeros datos de predicción y los segundos datos de predicción, para obtener un resultado de detección de vivacidad facial para la imagen de escaneo facial del usuario.

Los primeros datos de predicción pueden ser, por ejemplo, la puntuación PA de predicción descrita y los segundos datos de predicción pueden ser, por ejemplo, la puntuación PB de predicción descrita. Ciertamente, la puntuación de predicción es simplemente un ejemplo de una forma de expresión de los primeros datos de predicción y los segundos datos de predicción, o puede haber otra forma de expresión, por ejemplo, un valor de probabilidad o un valor booleano.

40

Lo anterior es el método de detección de vivacidad facial proporcionado en esta realización de la presente patente. Como se muestra en la FIG. 5, en base a una misma idea de la divulgación, una realización de la presente patente proporciona además un aparato correspondiente.

45

La FIG. 5 es un diagrama estructural esquemático que ilustra un aparato de detección de vivacidad facial correspondiente a la FIG. 3 de acuerdo con una realización de la presente patente. El aparato se puede ubicar en un cuerpo de ejecución del procedimiento de la FIG. 3, que incluye: un módulo 501 de entrenamiento, configurado para: entrenar un primer modelo de aprendizaje profundo en base a las imágenes faciales generales; y entrenar un segundo modelo de aprendizaje profundo en base a las imágenes faciales extraídas recortadas de las imágenes faciales generales; y un módulo 502 de detección, configurado para realizar la detección de vivacidad facial en base al primer modelo de aprendizaje profundo entrenado y al segundo modelo de aprendizaje profundo entrenado.

50

55 Opcionalmente, el primer modelo de aprendizaje profundo y el segundo modelo de aprendizaje profundo utilizan diferentes estructuras de red de aprendizaje profundo.

Opcionalmente, las diferentes estructuras de red de aprendizaje profundo incluyen al menos uno de los siguientes parámetros: una cantidad de capas variables ocultas, un tipo de capa variable oculta, una cantidad de nodos neuronales o el tamaño de un núcleo de convolución de una unidad de convolución.

60

Opcionalmente, la imagen facial general incluye una primera etiqueta y la primera etiqueta indica si una imagen facial general correspondiente a la primera etiqueta es una imagen facial en vivo.

65 El entrenamiento, mediante el módulo 501 de entrenamiento, de un primer modelo de aprendizaje profundo en base a las imágenes faciales generales incluye: ingresar, mediante el módulo 501 de entrenamiento, el primer modelo de

aprendizaje profundo a la imagen facial general, donde el primer modelo de aprendizaje profundo predice, en base a una característica de estructura de imagen de la imagen facial general, si la imagen facial general es una imagen facial en vivo; y ajustar el primer modelo de aprendizaje profundo en base a un resultado de predicción y la primera etiqueta de la imagen facial general.

5 La imagen facial extraída incluye una segunda etiqueta, y la segunda etiqueta indica si una imagen facial extraída correspondiente a la segunda etiqueta es una imagen facial en vivo.

10 El entrenamiento, mediante el módulo 501 de entrenamiento, de un segundo modelo de aprendizaje profundo en base a las imágenes faciales extraídas recortadas de las imágenes faciales generales incluye: obtener, mediante el módulo 501 de entrenamiento, la imagen facial extraída recortada de la imagen facial general; e ingresar el segundo modelo de aprendizaje profundo a la imagen facial extraída, donde el segundo modelo de aprendizaje profundo predice, en base a una característica de material de imagen de la imagen facial extraída, si la imagen facial extraída es una imagen facial en vivo; y ajustar el segundo modelo de aprendizaje profundo en base a un resultado de predicción y la segunda etiqueta de la imagen facial extraída.

15 Opcionalmente, el primer modelo de aprendizaje profundo y el segundo modelo de aprendizaje profundo se basan en una red neuronal convolucional.

20 Un núcleo de convolución de una unidad de convolución en una red neuronal convolucional en la que se basa el primer modelo de aprendizaje profundo es mayor que un núcleo de convolución de una unidad de convolución en una red neuronal convolucional en la que se basa el segundo modelo de aprendizaje profundo, de modo que el primer modelo de aprendizaje profundo extrae una característica de estructura de imagen de la imagen facial general, y el segundo modelo de aprendizaje profundo extrae una característica de material de imagen de la imagen facial extraída.

25 Opcionalmente, la realización, mediante el módulo 502 de detección, de la detección de vivacidad facial en base al primer modelo de aprendizaje profundo entrenado y al segundo modelo de aprendizaje profundo entrenado incluye: obtener, mediante el módulo 502 de detección, la imagen facial general recopilada para la detección de vivacidad facial; ingresar el primer modelo de aprendizaje profundo entrenado a la imagen facial general recopilada para su procesamiento, para obtener los primeros datos de predicción correspondientes; obtener una imagen facial extraída recortada de la imagen facial general recopilada e ingresar el segundo modelo de aprendizaje profundo entrenado para su procesamiento, para obtener los segundos datos de predicción correspondientes; y tomar una decisión conjunta en base a los primeros datos de predicción y los segundos datos de predicción, para obtener un resultado de detección de vivacidad facial para una imagen de escaneo facial del usuario.

30 En base a una misma idea de la divulgación, una realización de la presente patente proporciona además un dispositivo electrónico correspondiente, que incluye: al menos un procesador; y una memoria comunicativamente conectada al por lo menos un procesador.

35 La memoria almacena una instrucción que puede ejecutarse por el al menos un procesador, y la instrucción se ejecuta por el al menos un procesador, para permitir que el al menos un procesador: entrene un primer modelo de aprendizaje profundo en base a las imágenes faciales generales; entrene un segundo modelo de aprendizaje profundo en base a las imágenes faciales extraídas recortadas de las imágenes faciales generales; y realizar la detección de vivacidad facial en base al primer modelo de aprendizaje profundo entrenado y al segundo modelo de aprendizaje profundo entrenado.

40 En base a una misma idea de la divulgación, una realización de la presente patente proporciona además un medio de almacenamiento informático no volátil correspondiente, donde el medio de almacenamiento informático no volátil almacena una instrucción ejecutable por computadora, y la instrucción ejecutable por computadora está configurada para: entrenar un primer modelo de aprendizaje profundo en base a las imágenes faciales generales; entrenar un segundo modelo de aprendizaje profundo en base a las imágenes faciales extraídas recortadas de las imágenes faciales generales; y realizar la detección de vivacidad facial en base al primer modelo de aprendizaje profundo entrenado y al segundo modelo de aprendizaje profundo entrenado.

45 Las realizaciones en esta memoria descriptiva se describen todas de una manera progresiva, para partes iguales o similares en las realizaciones, se puede hacer referencia a estas realizaciones, y cada una de las realizaciones se centra en una diferencia con otras realizaciones. Especialmente, una realización de aparato, una realización de dispositivo electrónico, una realización de medio de almacenamiento informático no volátil son básicamente similares a una realización de método y, por lo tanto, se describen brevemente; para las partes relacionadas, se hace referencia a descripciones parciales en la realización del método.

50 El aparato, el dispositivo electrónico y el medio de almacenamiento informático no volátil proporcionados en las realizaciones de la presente patente corresponden al método. Por lo tanto, el aparato, el dispositivo electrónico y el medio de almacenamiento informático no volátil también tienen efectos técnicos beneficiosos similares a un efecto técnico beneficioso del método correspondiente. El efecto técnico beneficioso del método se describe en detalle

65

anteriormente, de modo que los efectos técnicos beneficiosos del correspondiente aparato, dispositivo electrónico y medio de almacenamiento informático no volátil no se describen aquí de nuevo.

En la década de 1990, se puede distinguir obviamente si la mejora de la tecnología es la mejora del hardware (por ejemplo, la mejora de la estructura de un circuito, tal como un diodo, un transistor o un interruptor) o la mejora del software (mejora del procedimiento de un método). Sin embargo, a medida que se desarrollan las tecnologías, la mejora de muchos procedimientos de métodos actuales se puede considerar como una mejora directa de la estructura de un circuito de hardware. Un diseñador suele programar un procedimiento de método mejorado en un circuito de hardware para obtener una estructura de circuito de hardware correspondiente. Por lo tanto, un procedimiento de método puede mejorarse mediante módulos de entidad de hardware. Por ejemplo, un dispositivo lógico programable (PLD) (p. ej., una matriz de puertas programables en campo (FPGA)) es un circuito integrado de este tipo, y un usuario determina una función lógica del dispositivo lógico programable mediante la programación del dispositivo. El diseñador realiza la programación para "integrar" un sistema digital a un PLD sin solicitar al fabricante de chips que diseñe y produzca un chip de circuito integrado de aplicación específica. Además, la programación se implementa principalmente modificando el software del "compilador lógico" en lugar de hacer manualmente un chip de circuito integrado. Esto es similar a un compilador de software utilizado para desarrollar y componer un programa. Sin embargo, el código original obtenido antes de la compilación también se escribe en un lenguaje de programación específico, y esto se conoce como lenguaje de descripción de hardware (Hardware Description Language, HDL). Sin embargo, hay varios HDL, tal como ABEL (lenguaje de expresión booleana avanzado), AHDL (lenguaje de descripción de hardware de Altera), Confluence, CUPL (Lenguaje de programación de la Universidad de Cornell), HDCal, JHDL (lenguaje de descripción de hardware de Java), Lava, Lola, MyHDL, PALASM y un RHDL (lenguaje de descripción de hardware de Ruby). Actualmente, un VHDL (lenguaje de descripción de hardware de circuito integrado de muy alta velocidad) y Verilog son los más populares. Un experto en la técnica también debe comprender que, solo se necesita realizar programación lógica en el procedimiento del método utilizando los varios lenguajes de descripción de hardware descritos, y los diversos lenguajes de descripción de hardware se programan en un circuito integrado, de modo que se puede obtener fácilmente un circuito de hardware que implementa el procedimiento del método lógico.

Un controlador puede implementarse de cualquier manera apropiada. Por ejemplo, el controlador puede utilizar un microprocesador o un procesador, y puede almacenar formas de un medio legible por computadora, una puerta lógica, un interruptor, un circuito integrado de aplicación específica (ASIC), un controlador lógico programable y un microcontrolador integrado que son de código de programa legible por computadora (p. ej., software o hardware) que puede ejecutarse por el (micro) procesador. Los ejemplos de controlador incluyen, pero no se limitan a los siguientes microcontroladores: ARC 625D, Atmel AT91SAM, Microchip PIC18F26K20 o Silicone Labs C8051F320. También se puede implementar un controlador de memoria como una parte de la lógica de control de la memoria. Un experto en la técnica también sabe que, además de implementar el controlador en una forma de código de programa puramente legible por computadora, la programación lógica se puede realizar completamente utilizando el paso del método, de modo que el controlador implementa una misma función en forma de puerta lógica, un interruptor, un circuito integrado de aplicación específica, un controlador lógico programable, un microcontrolador integrado, etc. Por lo tanto, el controlador puede considerarse como un componente de hardware, y también puede considerarse un aparato para implementar diversas funciones en el controlador como estructura en un componente de hardware. Alternativamente, un aparato configurado para implementar diversas funciones puede considerarse como un módulo de software o una estructura en un componente de hardware que puede implementar el método.

El sistema, aparato, módulo o unidad descritos en las realizaciones descritas puede implementarse mediante un chip de computadora o una entidad, o implementarse mediante un producto con una función. Un dispositivo de implementación típico es una computadora. Específicamente, la computadora puede ser, por ejemplo, una computadora personal, una computadora portátil, un teléfono móvil, un teléfono con cámara, un teléfono inteligente, un asistente digital personal, un reproductor multimedia, un dispositivo de navegación, un dispositivo de correo electrónico, una consola de juegos, una computadora tableta, un dispositivo ponible o una combinación de cualquiera de estos dispositivos.

Para facilitar la descripción, el aparato descrito se describe dividiendo las funciones en diversas unidades. Ciertamente, cuando se implementa la presente patente, las funciones de cada unidad se pueden implementar en una o más piezas de software y/o hardware.

Un experto en la técnica debe comprender que las realizaciones de la presente divulgación pueden proporcionarse como un método, un sistema o un producto de programa informático. Por lo tanto, la presente divulgación puede usar una forma de realizaciones de solo hardware, realizaciones de solo software o realizaciones con una combinación de software y hardware. Además, la presente divulgación puede usar una forma de producto de programa informático que se implementa en uno o más medios de almacenamiento utilizables por computadora (que incluyen, entre otros, una memoria de disco, un CD-ROM, una memoria óptica, etc.) que incluyen código de programa utilizable por computadora.

La presente divulgación se describe con referencia a los diagramas de flujo y/o diagramas de bloques del método, el dispositivo (sistema) y el producto de programa informático de acuerdo con las realizaciones de la presente divulgación. Debe entenderse que se pueden utilizar instrucciones de programa informático para implementar cada uno de los

- 5 procesos y/o cada uno de los bloques en los diagramas de flujo y/o los diagramas de bloques y una combinación de un proceso y/o un bloque en los diagramas de flujo y/o los diagramas de bloques. Estas instrucciones de programa informático se pueden proporcionar para una computadora de propósito general, una computadora dedicada, un procesador integrado o un procesador de cualquier otro dispositivo de procesamiento de datos programable para generar una máquina, de modo que las instrucciones ejecutadas por una computadora o un procesador de cualquier otro dispositivo de procesamiento de datos programable genera un aparato para implementar una función específica en uno o más procesos en los diagramas de flujo o en uno o más bloques en los diagramas de bloques.
- 10 Estas instrucciones de programa informático se pueden almacenar en una memoria legible por computadora que puede instruir a la computadora o cualquier otro dispositivo de procesamiento de datos programable para trabajar de una manera específica, de modo que las instrucciones almacenadas en la memoria legible por computadora generan un artefacto que incluye un aparato de instrucción. El aparato de instrucción implementa una función específica en uno o más procesos en los diagramas de flujo y/o en uno o más bloques en los diagramas de bloques.
- 15 Estas instrucciones de programa informático pueden cargarse en una computadora u otro dispositivo de procesamiento de datos programable, de modo que se realicen una serie de operaciones y pasos en la computadora u otro dispositivo programable, generando así un procesamiento implementado por computadora. Por lo tanto, las instrucciones ejecutadas en la computadora u otro dispositivo programable proporcionan pasos para implementar una función específica en uno o más procesos en los diagramas de flujo o en uno o más bloques en los diagramas de bloques.
- 20 En una configuración típica, el dispositivo informático incluye uno o más procesadores (CPU), una interfaz de entrada/salida, una interfaz de red y una memoria.
- 25 La memoria puede incluir una forma de memoria volátil, una memoria de acceso aleatorio (RAM) y/o una memoria no volátil, etc. en un medio legible por computadora, tal como una memoria de solo lectura (ROM) o una memoria flash (flash RAM). La memoria es un ejemplo del medio legible por computadora.
- 30 El medio legible por computadora incluye medios volátiles y no volátiles, extraíbles y no extraíbles, y puede almacenar información utilizando cualquier método o tecnología. La información puede ser una instrucción legible por computadora, una estructura de datos, un módulo de programa u otros datos. Los ejemplos de medio de almacenamiento informático incluyen, entre otros, una memoria de acceso aleatorio de cambio de fase (PRAM), una memoria de acceso aleatorio estática (SRAM), una memoria de acceso aleatorio dinámica (DRAM), una memoria de acceso aleatorio (RAM) de otro tipo, una memoria de solo lectura (ROM), una memoria de solo lectura programable y borrrable eléctricamente (EEPROM), una memoria flash u otra tecnología de memoria, una memoria de disco compacto de solo lectura (CD-ROM), un disco versátil digital (DVD) u otro almacenamiento óptico, una cinta magnética, un almacenamiento en disco magnético, otro dispositivo de almacenamiento magnético o cualquier otro medio que no sea de transmisión. El medio de almacenamiento informático se puede utilizar para almacenar información a la que puede acceder el dispositivo informático. Como se describe en esta memoria descriptiva, el medio legible por computadora no incluye medios transitorios (medios transitorios), por ejemplo, una señal de datos modulada y un portador.
- 35 Hay que señalar además que, los términos "incluye", "contiene", o cualquier otra variante, están destinados a cubrir la inclusión no exclusiva, de modo que un proceso, un método, un artículo o un dispositivo que incluya una serie de elementos no solo incluye estos mismos elementos, sino que también incluye otros elementos que no se enumeran expresamente, o incluye además elementos inherentes a dicho proceso, método, artículo o dispositivo. Un elemento precedido por "incluye un..." no excluye, sin más restricciones, la existencia de elementos idénticos adicionales en el proceso, método, artículo o dispositivo que incluye el elemento.
- 40 La presente patente puede describirse en contextos comunes de instrucciones ejecutables por computadora ejecutadas por una computadora, tal como un módulo de programa. Generalmente, el módulo de programa incluye una rutina, un programa, un objeto, un componente, una estructura de datos, etc. que ejecuta una tarea específica o implementa un tipo de datos abstracto específico. La presente patente también se puede practicar en entornos informáticos distribuidos. En estos entornos informáticos distribuidos, las tareas se ejecutan mediante dispositivos de procesamiento remotos que están conectados mediante una red de comunicaciones. En los entornos informáticos distribuidos, el módulo de programa se puede ubicar en medios de almacenamiento informáticos locales y remotos que incluyen dispositivos de almacenamiento.
- 45 Las realizaciones en esta memoria descriptiva se describen todas de una manera progresiva, para partes iguales o similares en las realizaciones, se puede hacer referencia a estas realizaciones, y cada una de las realizaciones se centra en una diferencia con otras realizaciones. Especialmente, una realización del sistema es básicamente similar a una realización del método y, por lo tanto, se describe brevemente; para las partes relacionadas, se puede hacer referencia a descripciones parciales en la realización del método.
- 50 Las descripciones anteriores son simplemente realizaciones de la presente patente y no pretenden limitar la presente patente. Para un experto en la técnica, la presente patente puede tener diversas modificaciones y cambios.
- 55
- 60
- 65

La FIG. 6 es un diagrama de flujo que ilustra un ejemplo de un método 600 implementado por computadora para determinar la autenticidad del usuario con detección de vivacidad facial. Para mayor claridad de presentación, la descripción que sigue describe generalmente el método 600 en el contexto de las otras figuras en esta descripción. Sin embargo, se entenderá que el método 600 se puede realizar, por ejemplo, por cualquier sistema, entorno, software y hardware, o una combinación de sistemas, entornos, software y hardware, según corresponda. En algunas implementaciones, diversas etapas del método 600 se pueden ejecutar en paralelo, en combinación, en bucles, o en cualquier orden.

En 602, se entrena un primer modelo de aprendizaje profundo para clasificar imágenes faciales generales. Las imágenes faciales generales se clasifican en al menos imágenes faciales en vivo e imágenes faciales no en vivo. En algunas implementaciones, las imágenes faciales en vivo se consideran muestras positivas y las imágenes faciales no en vivo se consideran muestras negativas. En algunas implementaciones, el primer modelo de aprendizaje profundo es un modelo de clasificación y las imágenes faciales generales se utilizan como entradas del primer modelo de aprendizaje profundo. El entrenamiento del primer modelo de aprendizaje profundo mejora la precisión de clasificación con respecto a las imágenes faciales generales.

En algunas implementaciones, una imagen facial general particular incluye una primera etiqueta que indica si la imagen facial general particular correspondiente a la primera etiqueta es una imagen facial en vivo. En algunas implementaciones, el entrenamiento del primer modelo de aprendizaje profundo incluye: 1) ingresar la imagen facial general particular al primer modelo de aprendizaje profundo para generar un primer resultado de predicción, en base a una característica de estructura de imagen de la imagen facial general particular, de si la imagen facial general particular es una imagen facial en vivo y 2) ajustar el primer modelo de aprendizaje profundo en base al primer resultado de predicción y la primera etiqueta. A partir de 602, el método 600 pasa a 604.

En 604, las imágenes faciales recortadas se extraen de las imágenes faciales generales. En algunas implementaciones, una imagen facial recortada particular incluye una segunda etiqueta, y la segunda etiqueta indica si la imagen facial recortada particular correspondiente a la segunda etiqueta es una imagen facial en vivo. En algunas implementaciones, el entrenamiento del segundo modelo de aprendizaje profundo en base a la imagen facial recortada incluye: 1) obtener la imagen facial recortada particular; 2) ingresar la imagen facial recortada particular al segundo modelo de aprendizaje profundo para generar un segundo resultado de predicción, en base a una característica de material de imagen de la imagen facial recortada particular, de si la imagen facial recortada particular es una imagen facial en vivo; y 3) ajustar el segundo modelo de aprendizaje profundo en base al segundo resultado de predicción y la segunda etiqueta. A partir de 604, el método 600 pasa a 606.

En 606, se entrena un segundo modelo de aprendizaje profundo en base a las imágenes faciales recortadas. A partir de 606, el método 600 pasa a 608.

En 608, se realiza una detección de vivacidad facial en base al primer modelo de aprendizaje profundo entrenado y al segundo modelo de aprendizaje profundo entrenado. En algunas implementaciones, el primer modelo de aprendizaje profundo y el segundo modelo de aprendizaje profundo se basan en una red neuronal convolucional, y en donde un núcleo de convolución de una unidad de convolución en una red neuronal convolucional del primer modelo de aprendizaje profundo es mayor que un núcleo de convolución de una unidad de convolución en una red neuronal convolucional del segundo modelo de aprendizaje profundo. Después de 608, el procedimiento 600 finaliza.

En algunas implementaciones, la detección de vivacidad facial incluye: 1) obtener una imagen facial general; 2) ingresar la imagen facial general al primer modelo de aprendizaje profundo entrenado para obtener los primeros datos de predicción correspondientes; 3) obtener una imagen facial recortada de la imagen facial general; 4) ingresar la imagen facial recortada al segundo modelo de aprendizaje profundo entrenado para obtener los segundos datos de predicción correspondientes; y 5) tomar una decisión conjunta en base a los primeros datos de predicción y los segundos datos de predicción para obtener un resultado de detección de vivacidad facial.

Las implementaciones de la materia objeto descrita en esta memoria descriptiva se pueden implementar para obtener ventajas o efectos técnicos particulares. La detección de vivacidad facial descrita se puede utilizar para mejorar los procesos de autenticación y garantizar la seguridad de los datos. Por ejemplo, el método descrito se puede utilizar para distinguir entre imágenes de un rostro humano vivo y no vivo para ayudar a evitar el fraude y el comportamiento malicioso con respecto a los datos protegidos. El método descrito puede incorporarse en dispositivos informáticos (tales como dispositivos informáticos móviles y dispositivos de formación de imágenes digitales).

El resultado de la vivacidad facial se puede mostrar en una interfaz gráfica de usuario. En base al resultado de vivacidad facial, una determinación de si realizar acciones posteriores (por ejemplo, desbloquear datos protegidos, operar una aplicación de software, almacenar datos, enviar datos a través de una red o mostrar datos en una interfaz gráfica de usuario).

La metodología descrita permite la mejora de diversas transacciones de dispositivos informáticos móviles y la seguridad general de transacciones/datos. Los participantes en transacciones que utilizan dispositivos informáticos

móviles pueden estar seguros de que las imágenes faciales utilizadas para desbloquear un dispositivo informático móvil o para autorizar una transacción son válidas y que no serán víctimas de fraude.

La metodología descrita puede garantizar el uso eficiente de los recursos informáticos (por ejemplo, ciclos de procesamiento, ancho de banda de red y uso de memoria), mediante la verificación eficiente de datos/transacciones. Al menos estas acciones pueden minimizar o prevenir el desperdicio de recursos informáticos disponibles con respecto a múltiples partes en una transacción de computación móvil al evitar transacciones no deseadas/fraudulentas. En lugar de que los usuarios necesiten verificar los datos con investigaciones o transacciones adicionales, se puede confiar en que las transacciones son válidas.

En algunas implementaciones, se puede analizar una interfaz gráfica de usuario para garantizar que los elementos gráficos utilizados en las operaciones de detección de vivacidad facial (por ejemplo, escaneo y verificación de la vivacidad de un rostro humano con un dispositivo informático móvil) se puedan colocar en interfaces gráficas de usuario para ser lo menos molestas para un usuario (por ejemplo, para ocultar la menor cantidad de datos y evitar cubrir cualquier elemento crítico o de uso frecuente de la interfaz gráfica de usuario).

Las realizaciones y las operaciones descritas en esta memoria descriptiva pueden implementarse en circuitería electrónica digital, o en software, firmware o hardware informático, incluidas las estructuras descritas en esta memoria descriptiva o en combinaciones de una o más de ellas. Las operaciones se pueden implementar como operaciones realizadas por un aparato de procesamiento de datos sobre datos almacenados en uno o más dispositivos de almacenamiento legibles por ordenador o recibidos de otras fuentes. Un aparato de procesamiento de datos, ordenador o dispositivo informático puede abarcar aparatos, dispositivos y máquinas para el procesamiento de datos, incluyendo a modo de ejemplo un procesador programable, un ordenador, un sistema en un chip, o múltiples o combinaciones de los anteriores. El aparato puede incluir circuitos lógicos de propósito especial, por ejemplo, una unidad central de procesamiento (CPU), una matriz de puertas programables en campo (FPGA) o un circuito integrado de aplicación específica (ASIC). El aparato también puede incluir código que crea un entorno de ejecución para el programa informático en cuestión, por ejemplo, código que constituye el firmware del procesador, una pila de protocolos, un sistema de gestión de bases de datos, un sistema operativo (por ejemplo, un sistema operativo o una combinación de sistemas operativos), un entorno de ejecución multiplataforma, una máquina virtual, o una combinación de uno o más de ellos. El aparato y el entorno de ejecución pueden realizar diversas infraestructuras de modelos de computación diferentes, tales como servicios web, computación distribuida e infraestructuras de computación en red.

Un programa informático (también conocido, por ejemplo, como programa, software, aplicación de software, módulo de software, unidad de software, secuencia de comandos o código) se puede escribir en cualquier forma de lenguaje de programación, incluyendo lenguajes compilados o interpretados, lenguajes declarativos o procedimentales, y se puede implementar en cualquier forma, incluyendo como un programa independiente o como un módulo, componente, subrutina, objeto u otra unidad adecuada para su utilización en un entorno informático. Un programa se puede almacenar en una parte de un archivo que contiene otros programas o datos (por ejemplo, una o más secuencias de comandos almacenadas en un documento de lenguaje de marcas), en un único archivo dedicado al programa en cuestión, o en múltiples archivos coordinados (por ejemplo, archivos que almacenan uno o más módulos, subprogramas o partes de código). Un programa informático se puede ejecutar en un ordenador o en varios ordenadores situados en un mismo sitio o distribuidos en varios sitios e interconectados por una red de comunicación.

Los procesadores para la ejecución de un programa informático incluyen, a modo de ejemplo, tanto los microprocesadores de propósito general como los de propósito especial, y uno o varios procesadores de cualquier clase de ordenador digital. Por lo general, un procesador recibirá instrucciones y datos de una memoria de sólo lectura, de una memoria de acceso aleatorio o de ambas. Los elementos esenciales de un ordenador son un procesador para realizar acciones de acuerdo con las instrucciones y uno o más dispositivos de memoria para almacenar instrucciones y datos. Por lo general, un ordenador también incluirá, o se acoplará de forma operativa para recibir datos desde o transferir datos a, o ambos, uno o más dispositivos de almacenamiento masivo para almacenar datos. Un ordenador se puede integrar en otro dispositivo, por ejemplo, un dispositivo móvil, un asistente digital personal (PDA), una consola de juegos, un receptor del sistema de posicionamiento global (GPS) o un dispositivo de almacenamiento portátil. Los dispositivos adecuados para almacenar las instrucciones y los datos del programa informático incluyen, pero no limitan memorias no volátiles, medios y dispositivos de memoria, incluyendo, a modo de ejemplo, dispositivos de memoria de semiconductores, discos magnéticos y discos magneto-ópticos. El procesador y la memoria se pueden complementar por, o incorporar en, circuitos lógicos de propósito especial.

Los dispositivos móviles pueden incluir teléfonos, equipos de usuario (UE), teléfonos móviles (por ejemplo, teléfonos inteligentes), tabletas, dispositivos para llevar puestos (por ejemplo, relojes inteligentes y gafas inteligentes), dispositivos implantados dentro del cuerpo humano (por ejemplo, biosensores, implantes cocleares), u otros tipos de dispositivos móviles. Los dispositivos móviles se pueden comunicar de forma inalámbrica (por ejemplo, utilizando señales de radiofrecuencia (RF)) con diversas redes de comunicación (descritas a continuación). Los dispositivos móviles pueden incluir sensores para determinar las características del entorno actual del dispositivo móvil. Los sensores pueden incluir cámaras, micrófonos, sensores de proximidad, sensores GPS, sensores de movimiento, acelerómetros, sensores de luz ambiental, sensores de humedad, giroscopios, brújulas, barómetros, sensores de

5 huellas dactilares, sistemas de reconocimiento facial, sensores de RF (por ejemplo, radios Wi-Fi y celulares), sensores térmicos u otros tipos de sensores. Por ejemplo, las cámaras pueden incluir una cámara orientada hacia delante o hacia atrás con lentes móviles o fijas, un flash, un sensor de imagen y un procesador de imágenes. La cámara puede ser una cámara de megapíxeles que puede capturar detalles para el reconocimiento facial y/o del iris. La cámara, junto con un procesador de datos y la información de autenticación almacenada en la memoria o a la que se accede de forma remota, puede formar un sistema de reconocimiento facial. El sistema de reconocimiento facial o uno o más sensores, por ejemplo, micrófonos, sensores de movimiento, acelerómetros, sensores GPS o sensores RF, se pueden utilizar para la autenticación del usuario.

10 Para posibilitar la interacción con un usuario, las formas de realización se pueden implementar en un ordenador que tenga un dispositivo de visualización y un dispositivo de entrada, por ejemplo, una pantalla de cristal líquido (LCD) o un diodo orgánico de emisión de luz (OLED)/realidad virtual (VR)/realidad aumentada (AR) para mostrar información al usuario y una pantalla táctil, un teclado y un dispositivo señalador mediante el cual el usuario pueda proporcionar información al ordenador. También se pueden utilizar otras clases de dispositivos para posibilitar la interacción con el usuario; por ejemplo, la retroalimentación que se proporciona al usuario puede ser cualquier forma de retroalimentación sensorial, por ejemplo, retroalimentación visual, auditiva o táctil; y la entrada del usuario se puede recibir de cualquier forma, incluyendo entrada acústica, verbal o táctil. Además, un ordenador puede interactuar con un usuario enviando documentos a y recibiendo documentos de un dispositivo utilizado por el usuario; por ejemplo, enviando páginas web a un navegador web en un dispositivo cliente del usuario en respuesta a las solicitudes recibidas del navegador web.

25 Las formas de realización se pueden implementar utilizando dispositivos informáticos interconectados por cualquier forma o medio de comunicación de datos digital cableado o inalámbrico (o una combinación de los mismos), por ejemplo, una red de comunicación. Ejemplos de dispositivos interconectados son un cliente y un servidor generalmente remotos entre sí que suelen interactuar a través de una red de comunicación. Un cliente, por ejemplo, un dispositivo móvil, puede realizar transacciones por sí mismo, con un servidor, o a través de un servidor, por ejemplo, realizando transacciones de compra, venta, pago, regalo, envío o préstamo, o autorizando las mismas. Dichas transacciones se pueden realizar en tiempo real, de tal manera que una acción y una respuesta sean temporalmente próximas; por ejemplo, un individuo percibe que la acción y la respuesta se producen, en esencia, de forma simultánea, la diferencia de tiempo para una respuesta que sigue a la acción del individuo es inferior a 1 milisegundo (ms) o inferior a 1 segundo (s), o la respuesta es sin retraso intencionado teniendo en cuenta las limitaciones de procesamiento del sistema.

35 Los ejemplos de redes de comunicación incluyen, pero no limitan una red de área local (LAN), una red de acceso de radio (RAN), una red de área metropolitana (MAN) y una red de área amplia (WAN). La red de comunicación puede incluir la totalidad o una parte de Internet, otra red de comunicación o una combinación de redes de comunicación. La información se puede transmitir en la red de comunicaciones de acuerdo con diversos protocolos y estándares, incluidos evolución a largo plazo (LTE), 5G, IEEE 802, Protocolo de Internet (IP) u otros protocolos o combinaciones de protocolos. La red de comunicación puede transmitir voz, vídeo, datos biométricos o de autenticación, u otra información entre los dispositivos informáticos conectados. Las características descritas como implementaciones separadas se pueden implementar, en combinación, en una única implementación, mientras que las características descritas como una única implementación se pueden implementar en múltiples implementaciones, por separado, o en cualquier subcombinación adecuada. Las operaciones descritas y reivindicadas en un orden particular no se deben entender como que se requiere ese orden particular, ni que se deban realizar todas las operaciones ilustradas (algunas operaciones pueden ser opcionales). Según el caso, se puede realizar la multitarea o el procesamiento en paralelo (o una combinación de multitarea y procesamiento en paralelo).

REIVINDICACIONES

1. Un método para determinar si una imagen que incluye un rostro es una imagen en vivo o una imagen no en vivo, el método que comprende:

5 entrenar (S301) un primer modelo de aprendizaje profundo mediante entrenamiento supervisado en una pluralidad de imágenes faciales generales, las imágenes faciales generales que comprenden imágenes faciales en vivo recopiladas mediante la toma de un rostro en vivo y etiquetadas como muestras positivas, e imágenes faciales no en vivo recopiladas mediante la toma de un rostro no vivo que es una imagen facial o un modelo facial y etiquetado como muestras negativas;

15 entrenar (S302) una pluralidad de segundos modelos de aprendizaje profundo mediante entrenamiento supervisado en una pluralidad de imágenes faciales extraídas recortadas de las imágenes faciales generales, los segundos modelos de aprendizaje profundo comprenden un modelo de aprendizaje profundo de la región facial completa, un modelo de aprendizaje profundo del ojo y un modelo de aprendizaje profundo de la nariz, los dos últimos corresponden a un tipo de región facial de ojo y nariz respectivamente, las imágenes faciales extraídas comprenden imágenes faciales en vivo etiquetadas como muestras positivas e imágenes faciales no en vivo y etiquetadas como muestras negativas, en donde el primer modelo de aprendizaje profundo y cada uno de los segundos modelos de aprendizaje profundo son modelos de clasificación y en donde, después del entrenamiento, los modelos clasifican las imágenes faciales en una categoría de imagen facial en vivo o una categoría de imagen facial no en vivo;

20 realizar (S303) la detección de vivacidad facial en una primera imagen facial general utilizando el primer modelo de aprendizaje profundo entrenado para obtener una primera puntuación de predicción y la pluralidad de segundos modelos de aprendizaje profundo entrenados para obtener una segunda puntuación de predicción, comprende:

25 obtener la primera imagen facial general recopilada para la detección de vivacidad facial;

Ingresar la primera imagen facial general en el primer modelo de aprendizaje profundo entrenado para procesar y obtener la primera puntuación de predicción;

30 obtener una pluralidad de imágenes faciales extraídas recortadas de la primera imagen facial general, las imágenes faciales extraídas que comprenden una región de la imagen facial completa, una imagen de región de imagen del ojo y una imagen de región de imagen de la nariz, e ingresar las imágenes faciales extraídas en los respectivos segundos modelos de aprendizaje profundo entrenados para el procesamiento, los segundos modelos de aprendizaje profundo que comprenden todo el modelo de aprendizaje profundo de imagen facial, el modelo de aprendizaje profundo del ojo y el modelo de aprendizaje profundo de la nariz, para obtener las segundas puntuaciones de predicción;

35 generar un resultado de puntuación de predicción en base a la primera puntuación de predicción y la segunda puntuación de predicción; y

40 comparar el resultado de la puntuación de predicción con un umbral para determinar si la primera imagen facial general es una imagen en vivo o una imagen no en vivo.

45 2. El método de acuerdo con la reivindicación 1, en donde el primer modelo de aprendizaje profundo y un segundo modelo de aprendizaje profundo utilizan diferentes estructuras de red de aprendizaje profundo.

50 3. El método de acuerdo con la reivindicación 2, en donde las diferentes estructuras de red de aprendizaje profundo comprenden al menos uno de los siguientes parámetros: una cantidad de capas variables ocultas, un tipo de capa variable oculta, una cantidad de nodos neuronales o un tamaño de un núcleo de convolución de una unidad de convolución.

55 4. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en donde generar un resultado de la puntuación de predicción en base a la primera puntuación de predicción y las segundas puntuaciones de predicción comprende generar el resultado de la puntuación de predicción como una suma de la primera puntuación de predicción y las segundas puntuaciones de predicción.

60 5. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en donde el primer modelo de aprendizaje profundo y el segundo modelo de aprendizaje profundo se basan en una red neuronal convolucional; y

un núcleo de convolución de una unidad de convolución en una red neuronal convolucional en la que se basa el primer modelo de aprendizaje profundo es mayor que un núcleo de convolución de una unidad de convolución en una red neuronal convolucional en la que se basa el segundo modelo de aprendizaje profundo, de modo que el primer modelo de aprendizaje profundo extrae una característica de estructura de imagen de una imagen facial general y el segundo modelo de aprendizaje profundo extrae una característica de material de imagen de la imagen facial extraída.

65

6. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en donde las puntuaciones de predicción son todas un valor de probabilidad o un valor booleano.

5 7. Un aparato, que comprende una pluralidad de módulos (501, 502) configurados para realizar el método de una cualquiera de las reivindicaciones 1 a 6.

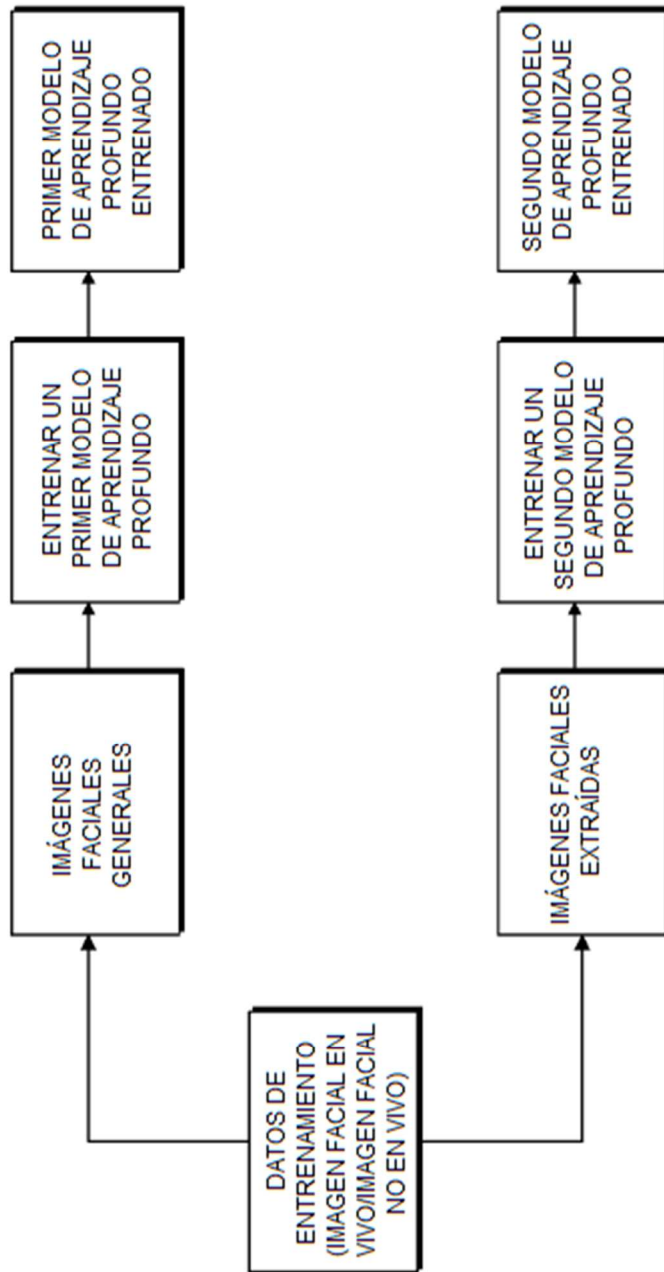


FIG. 1

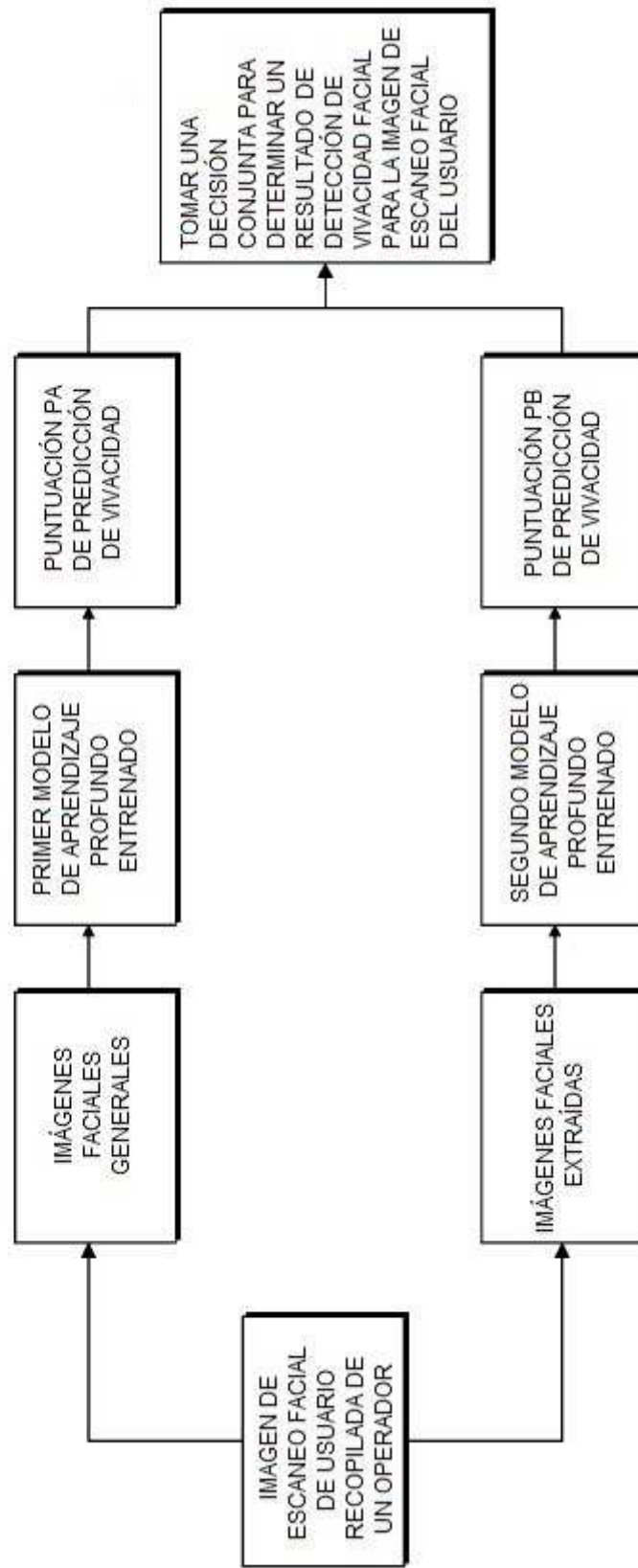


FIG. 2

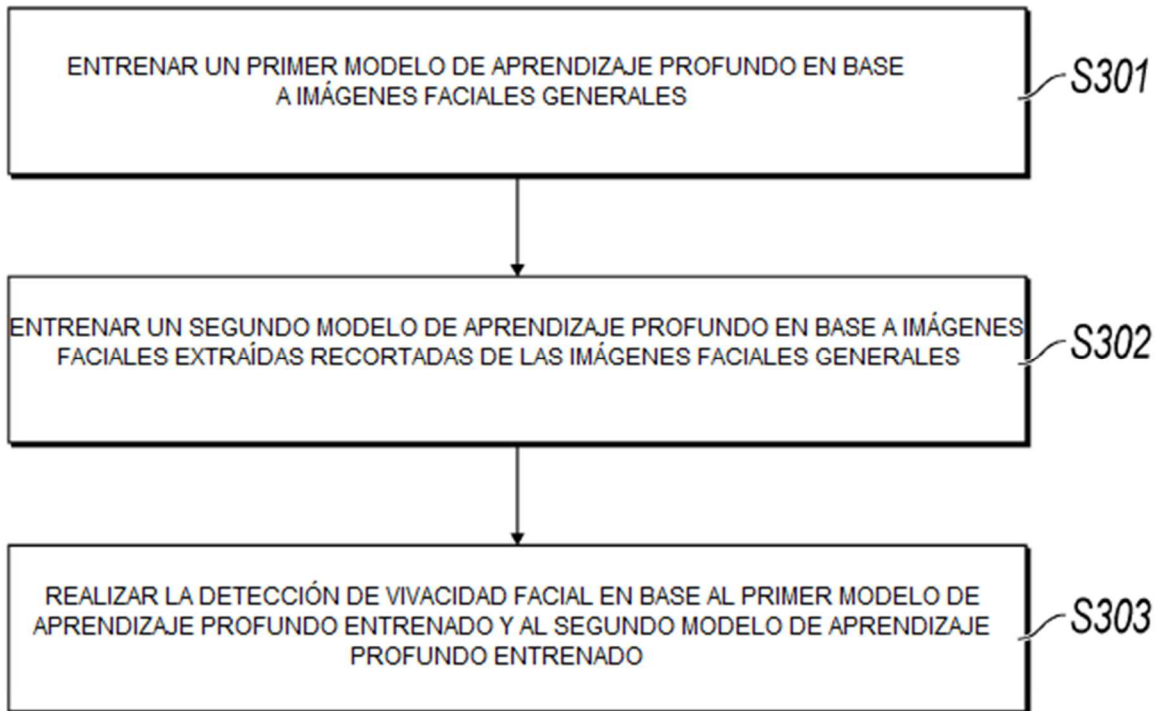


FIG. 3

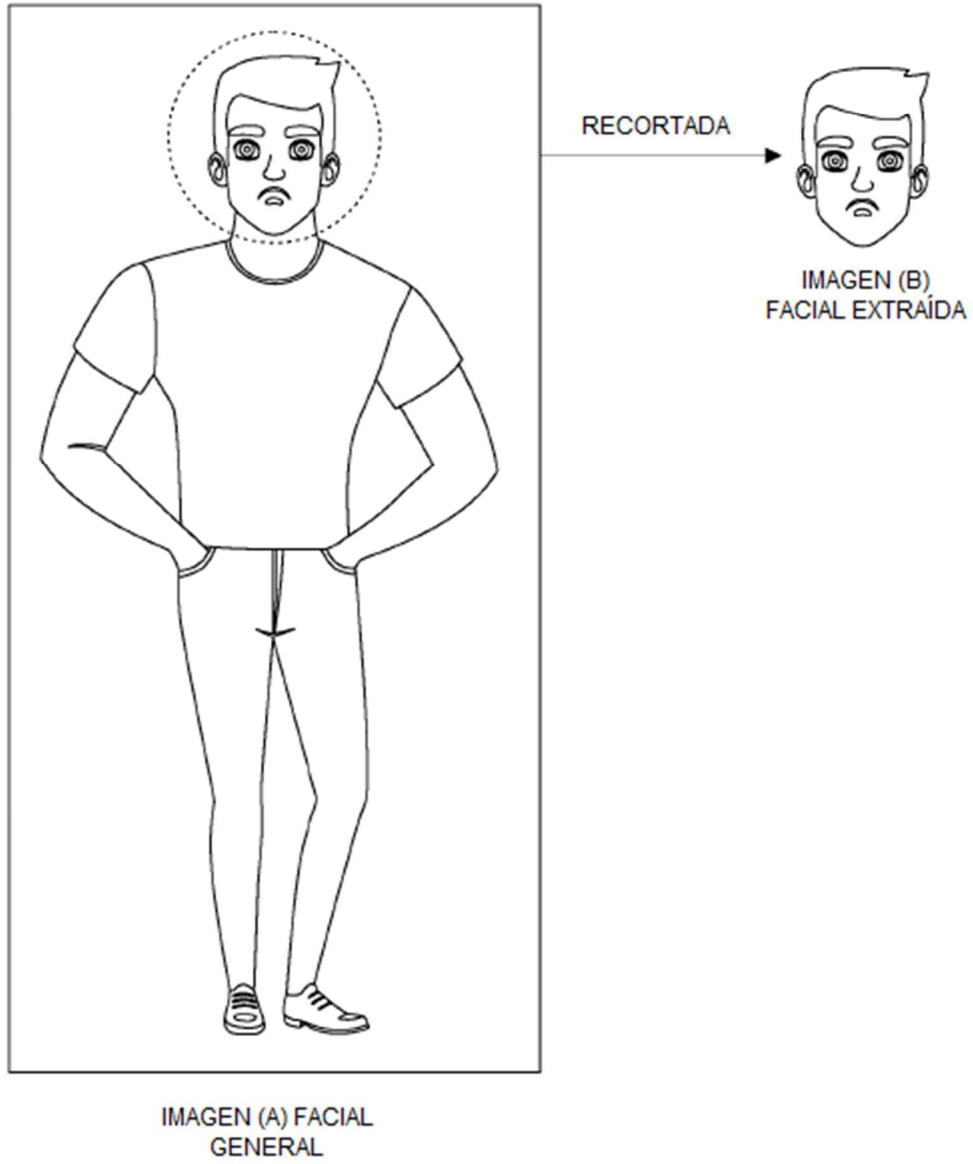


FIG. 4

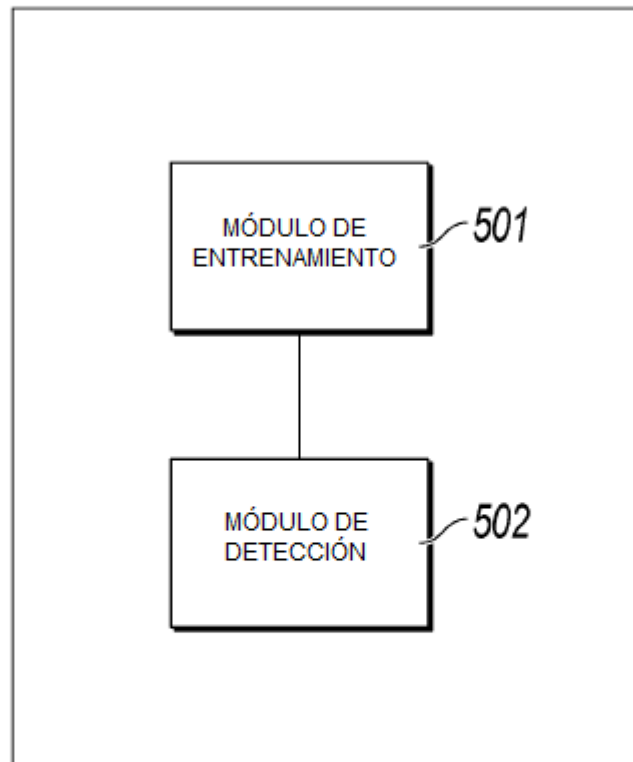


FIG. 5

600
↙

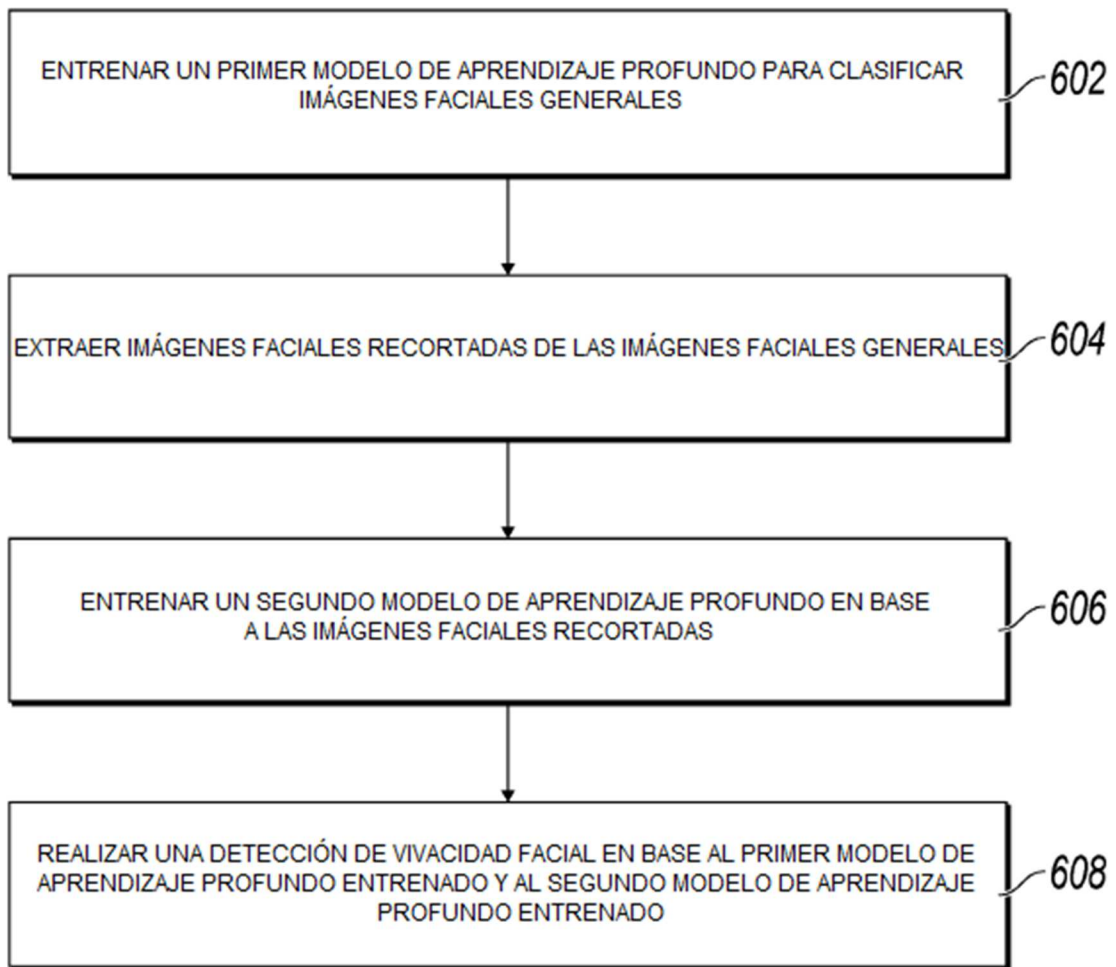


FIG. 6