

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 004 560**

51 Int. Cl.:

G06N 3/045 (2013.01)

G06N 3/047 (2013.01)

G06T 5/77 (2014.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **19.07.2019** **PCT/US2019/042509**

87 Fecha y número de publicación internacional: **03.12.2020** **WO20242508**

96 Fecha de presentación y número de la solicitud europea: **19.07.2019** **E 19749105 (3)**

97 Fecha y número de publicación de la concesión europea: **30.10.2024** **EP 3762873**

54 Título: **Redes neuronales de ampliación de imágenes**

30 Prioridad:

24.05.2019 US 201962852949 P

30.05.2019 US 201962854833 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
12.03.2025

73 Titular/es:

GOOGLE LLC (100.00%)
1600 Amphitheatre Parkway
Mountain View, CA 94043, US

72 Inventor/es:

BONNEVIE, MIKAEL PIERRE;
MASCHINOT, AARON;
SARNA, AARON;
BI, SHUCHAO;
WANG, JINGBIN;
KRAININ, MICHAEL SPENCER;
TONG, WENCHAO;
KRISHNAN, DILIP;
GONG, HAIFENG;
LIU, CE;
TALEBI, HOSSEIN;
SAYAG, RAANAN y
TETERWAK, PIOTR

74 Agente/Representante:

GONZÁLEZ PECES, Gustavo Adolfo

ES 3 004 560 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

Redes neuronales de ampliación de imágenes

5 ANTECEDENTES DE LA INVENCION

Esta memoria descriptiva se refiere al procesamiento de imágenes utilizando modelos de aprendizaje automático.

10 Los modelos de aprendizaje automático reciben una entrada y generan una salida, por ejemplo, una salida predicha, basándose en la entrada recibida. Algunos modelos de aprendizaje automático son modelos paramétricos y generan la salida basándose en la entrada recibida y en los valores de los parámetros del modelo.

15 Algunos modelos de aprendizaje automático son modelos profundos que emplean múltiples capas de modelos para generar una salida para una entrada recibida. Por ejemplo, una red neuronal profunda es un modelo de aprendizaje automático profundo que incluye una capa de salida y una o más capas ocultas que aplican cada una una transformación a una entrada recibida para generar una salida. Steven Cheng-Xian Li Y OTROS: «MisGAN: El aprendizaje a partir de datos incompletos con redes generativas adversarias», ARXIV.ORG, NY 14853, de 25 de febrero de 2019, describe, de acuerdo con su resumen, un marco basado en GAN para el aprendizaje a partir de datos incompletos complejos y de grandes dimensiones. El marco propuesto aprende un generador de datos completo junto con un generador de enmascaramiento que modela la distribución de los datos que faltan.

RESUMEN

25 Esta memoria descriptiva describe un sistema implementado como programas informáticos en uno o más ordenadores en una o más ubicaciones que realiza ampliación de imágenes. La invención a la que se refiere la presente patente europea se define en las reivindicaciones independientes 1, 13 y 14 adjuntas. Las formas de realización preferidas de la invención a la que se refiere la presente patente europea se exponen en las reivindicaciones dependientes adjuntas.

30 De acuerdo con un primer aspecto, se proporciona un método realizado por uno o más aparatos de procesamiento de datos, comprendiendo el método proporcionar una entrada que comprenda la imagen proporcionada a una red neuronal generativa con una pluralidad de parámetros de red neuronal generativa, donde la red neuronal generativa procesa la entrada de conformidad con valores entrenados de la pluralidad de parámetros de red neuronal generativa para generar una imagen ampliada; la imagen ampliada tiene (i) más filas, más columnas, o ambas que la imagen proporcionada, y (ii) se predice que es una ampliación realista de la imagen proporcionada; y la red neuronal generativa se ha entrenado usando una función objetivo de pérdida adversaria.

40 De este modo, las filas y/o columnas adicionales proporcionan una ampliación más allá de uno o más de los bordes originales de la imagen proporcionada y proporcionan una ampliación realista predicha de la imagen proporcionada como, por ejemplo, conservando las características semánticas de alto nivel y las estructuras y las texturas de bajo nivel de la imagen proporcionada.

45 El método puede comprender recibir una solicitud de presentación de una imagen en un bloque; determinar que la imagen proporcionada responde a la solicitud; y, tras generar la imagen ampliada, proporcionar la imagen ampliada en respuesta a la solicitud.

El método puede comprender determinar que la imagen proporcionada es apta para la ampliación de imagen basándose en un tamaño del bloque que difiere de un tamaño de la imagen proporcionada, y proporcionar en respuesta la entrada que comprende la imagen proporcionada a la red neuronal generativa.

50 El tamaño del bloque puede especificar una relación de aspecto del bloque.

La solicitud puede especificar elementos adicionales que han de presentarse con la imagen en el bloque, y proporcionar la imagen ampliada en respuesta a la solicitud puede incluir superponer los elementos adicionales sobre una porción ampliada de la imagen ampliada.

55 El bloque puede presentarse junto a resultados de búsqueda o en una página web de terceros.

La red neuronal generativa puede comprender una pluralidad de capas de red neuronal convolucional.

60 La red neuronal generativa puede comprender una pluralidad de conexiones de salto.

La red neuronal generativa puede comprender una pluralidad de capas de normalización de instancias.

65 La entrada a la red neuronal generativa puede comprender una imagen inicial con un mismo número de filas y columnas que la imagen ampliada, donde la imagen inicial comprenda: (i) una primera porción que corresponda a la imagen proporcionada, y (ii) una segunda porción con valores de píxel predeterminados, y una imagen enmascarada con el mismo

número de filas y columnas que la imagen ampliada, donde la imagen enmascarada identifique la primera porción y la segunda porción de la imagen inicial.

Los píxeles de la imagen enmascarada correspondientes a la primera porción de la imagen inicial pueden tener cada uno un primer valor de píxel, y los píxeles de la imagen enmascarada correspondientes a la segunda porción de la imagen inicial pueden tener cada uno un segundo valor de píxel que sea diferente del primer valor de píxel.

La red neuronal generativa puede entrenarse de manera conjunta con una red neuronal discriminadora con una pluralidad de parámetros de red neuronal discriminadora que esté configurada para procesar una imagen dada para generar una salida discriminadora que caracterice una probabilidad de que la imagen dada se haya generado utilizando la red neuronal generativa.

Entrenar la red neuronal generativa utilizando la función objetivo de pérdida adversaria puede comprender procesar una entrada de entrenamiento que comprenda una imagen de entrenamiento utilizando la red neuronal generativa y, de conformidad con los valores actuales de los parámetros de red neuronal generativa, generar una imagen ampliada de entrenamiento que amplíe la imagen de entrenamiento; generar una entrada de red neuronal discriminadora basándose en la imagen ampliada de entrenamiento; procesar la entrada de red neuronal discriminadora utilizando la red neuronal discriminadora y, de conformidad con los valores actuales de los parámetros de red neuronal discriminadora, generar una salida discriminadora que caracterice una probabilidad de que la entrada de red neuronal discriminadora se haya generado utilizando la red neuronal generativa; y ajustar los valores actuales de los parámetros de red neuronal generativa basándose en la función objetivo de pérdida adversaria, donde la función objetivo de pérdida adversaria dependa de la salida discriminadora que caracterice la probabilidad de que la entrada de red neuronal discriminadora se haya generado utilizando la red neuronal generativa.

Generar una entrada de red neuronal discriminadora basándose en la imagen ampliada de entrenamiento puede incluir sobreescribir una porción de la imagen ampliada de entrenamiento correspondiente a la imagen de entrenamiento con la imagen de entrenamiento.

El método puede incluir ajustar los valores actuales de los parámetros de red neuronal generativa basándose en una función objetivo de pérdida de reconstrucción que caracterice una similitud de la imagen ampliada de entrenamiento con una imagen objetivo, donde la imagen de entrenamiento es una representación recortada de la imagen objetivo.

La red neuronal discriminadora puede estar condicionada a una representación de características semánticas de una imagen objetivo, donde la imagen de entrenamiento sea una representación recortada de la imagen objetivo.

La representación de características semánticas de la imagen objetivo puede determinarse utilizando una salida intermedia de una red neuronal de clasificación mediante el procesamiento de la imagen objetivo.

La salida discriminadora puede basarse en: (i) una salida de una capa final de la red neuronal discriminadora, y (ii) una medida de similitud entre una salida intermedia de la red neuronal discriminadora y la representación de características semánticas de la imagen objetivo.

Según otro aspecto, se proporciona un sistema que comprende uno o más ordenadores y uno o más dispositivos de almacenamiento que almacenan instrucciones que, cuando son ejecutadas por el uno o más ordenadores, hacen que el uno o más ordenadores realicen las operaciones del método respectivo del primer aspecto.

Según otro aspecto, se proporcionan uno o más medios de almacenamiento informático que almacenan instrucciones que, cuando son ejecutadas por uno o más ordenadores, hacen que el uno o más ordenadores realicen las operaciones del método respectivo del primer aspecto.

Las características opcionales de un aspecto pueden combinarse con otro cuando proceda. Las formas de realización concretas del objeto descrito en esta memoria descriptiva se pueden implementar con el fin de poner en práctica una o más de las siguientes ventajas.

El sistema de ampliación de imágenes descrito en esta memoria descriptiva genera imágenes ampliadas (por ejemplo, que extienden las imágenes de entrada más allá de sus bordes originales) utilizando una red neuronal generativa. La red neuronal generativa se entrena de manera conjunta con una red neuronal «discriminadora» configurada para procesar una imagen de entrada para generar una salida que defina una predicción sobre si la imagen de entrada es (i) una imagen «real», o (ii) una imagen ampliada generada por la red neuronal generativa. Para facilitar el entrenamiento conjunto de la red neuronal generativa y la red neuronal discriminadora, el sistema condiciona la red neuronal discriminadora a las características semánticas correspondientes a imágenes de entrada procesadas por la red neuronal discriminadora. Condicionar la red neuronal discriminadora de esta manera puede permitir que el sistema sea entrenado para alcanzar un nivel de rendimiento aceptable en menos iteraciones de entrenamiento, haciendo así posible que el sistema consuma menos recursos computacionales (por ejemplo, memoria y potencia de cálculo) durante el entrenamiento que algunos sistemas convencionales.

Algunos sistemas convencionales de restauración de imágenes (*in-painting*) «rellenan» una porción de una imagen que está rodeada en todas las direcciones por datos de imagen originales. En cambio, el sistema de ampliación de imágenes descrito en esta memoria descriptiva amplía una imagen más allá de sus bordes originales para generar una imagen ampliada que amplía las características semánticas de alto nivel y las estructuras y las texturas de imagen de bajo nivel de la imagen. La aplicación directa de un sistema convencional de restauración de imágenes para realizar tareas de ampliación de imágenes puede, en algunos casos, dar como resultado imágenes ampliadas de menor calidad que las generadas por el sistema de ampliación de imágenes descrito en esta memoria descriptiva. Por ejemplo, los sistemas convencionales de restauración de imágenes pueden generar imágenes ampliadas con píxeles borrosos o repetitivos y semántica incoherente, mientras que las imágenes ampliadas creadas por el sistema de ampliación de imágenes descrito en el presente documento no tendrán las mismas características de píxeles borrosos, repetitivos ni de semántica incoherente.

El sistema de ampliación de imágenes descrito en esta memoria descriptiva genera imágenes ampliadas que pueden utilizarse para cualquiera de entre una variedad de aplicaciones, por ejemplo, aplicaciones de realidad virtual, aplicaciones de fotografía computacional y aplicaciones de distribución de componentes digitales. En las aplicaciones de distribución de componentes digitales, un sistema de distribución de componentes digitales transmite componentes digitales para su presentación en bloques, por ejemplo, junto a resultados de búsqueda o en páginas web de terceros. En algunos casos, una imagen incluida en un componente digital puede tener un tamaño que impida que la imagen llene un bloque definido, por ejemplo, porque la relación de aspecto de la imagen y el bloque son diferentes. En esta situación, el sistema de ampliación de imágenes descrito en esta memoria descriptiva puede utilizarse para redimensionar la imagen para que la imagen pueda llenar el bloque, haciendo así posible un uso más eficiente de los recursos (en particular, un uso más eficiente del espacio disponible en los bloques) y evitando que haya espacios en blanco y/o presentaciones de imagen distorsionadas en una interfaz de usuario. Expresado de otro modo, el sistema descrito en esta solicitud posibilita no solo «rellenar» una imagen, sino ampliarla más allá de sus bordes, lo que permite que una imagen se modifique para encajar con múltiples bloques de diferentes tamaños, incluso cuando solo haya disponible un único tamaño de imagen. Esto reduce el número de imágenes que hay que almacenar, ya que una sola imagen puede modificarse para que encaje en bloques de distintos tamaños conservando a la vez las características de la imagen original. El sistema de ampliación de imágenes puede obviar la necesidad de distorsionar una imagen estirándola o comprimiéndola a lo largo de ciertas dimensiones para cambiar la relación de aspecto de la imagen y hacer que esta llene un bloque. Además, en las aplicaciones de distribución de componentes digitales, algunos proveedores de componentes digitales pueden exigir que las imágenes incluidas en los componentes digitales se presenten sin distorsionar (por ejemplo, estirándolas o comprimiéndolas a lo largo de determinadas dimensiones).

En los dibujos adjuntos y en la descripción que sigue a continuación se exponen de manera detallada una o más formas de realización del objeto de esta memoria descriptiva. Otras características, aspectos y ventajas del objeto de la invención se harán evidentes a partir de la descripción, los dibujos y las reivindicaciones.

DESCRIPCIÓN BREVE DE LOS DIBUJOS

La FIGURA 1 muestra un sistema de ampliación de imágenes a modo de ejemplo.
La FIGURA 2 ilustra un flujo de datos a modo de ejemplo para entrenar una red neuronal generativa de manera conjunta con una red neuronal discriminadora utilizando una función objetivo de pérdida adversaria.
La FIGURA 3 es una ilustración de un ejemplo de una imagen ampliada generada por el sistema de ampliación de imágenes.
La FIGURA 4 es un diagrama de bloques de un entorno a modo de ejemplo en el que un sistema de distribución de componentes digitales utiliza el sistema de ampliación de imágenes.
La FIGURA 5 muestra un componente digital que está siendo visualizado en una pantalla de un dispositivo de usuario, donde la imagen incluida en el componente digital se ha ampliado utilizando el sistema de ampliación de imágenes.
La FIGURA 6 muestra un componente digital que está siendo visualizado en un bloque *banner*, donde la imagen incluida en el componente digital se ha ampliado utilizando el sistema de ampliación de imágenes.
La FIGURA 7 es una ilustración de una interfaz de usuario a modo de ejemplo que puede presentarse a una entidad que proporcione un componente digital para su transmisión por un sistema de distribución de componentes digitales.
La FIGURA 8 es un diagrama de flujo de un proceso a modo de ejemplo para proporcionar una imagen ampliada en respuesta a una solicitud.
La FIGURA 9 es un diagrama de bloques de un sistema informático a modo de ejemplo.

Los números de referencia y las denominaciones similares en los diversos dibujos indican elementos similares.

DESCRIPCIÓN DETALLADA

Esta memoria descriptiva describe tecnología para procesar una imagen con el fin de generar una imagen «ampliada» que amplíe de forma realista la imagen más allá de sus bordes originales, por ejemplo, que amplíe la imagen más allá de sus bordes originales conservando a la vez las características semánticas de alto nivel y las estructuras y las texturas de bajo nivel de la imagen. Tal procesamiento puede utilizarse en la mejora de imágenes. Por ejemplo, si una imagen tiene

dimensiones que no coinciden con un bloque en el que se va a colocar la imagen, las dimensiones de la imagen pueden ampliarse para que coincidan con el bloque, conservándose a la vez las características del contenido de la imagen original. En otro ejemplo, si una imagen está dañada como, por ejemplo, si falta una parte de la imagen, se pueden utilizar aspectos de la presente divulgación para generar de manera automática la parte que falta. Por ejemplo, una imagen capturada incorrectamente o una versión capturada de una imagen dañada que esté incompleta puede regenerarse con las técnicas que se han expuesto en el presente documento.

Para ampliar una imagen, el sistema de ampliación de imágenes descrito en esta memoria descriptiva procesa la imagen mediante una red neuronal generativa que es entrenada usando una función objetivo de pérdida adversaria para generar una imagen ampliada correspondiente. La función objetivo de pérdida adversaria evalúa una imagen ampliada generada mediante la red neuronal generativa basándose en una salida generada por una red neuronal discriminadora mediante el procesamiento de la imagen ampliada. La red neuronal discriminadora se entrena para predecir si una imagen de entrada es (i) una imagen «real» (es decir, natural), o (ii) una imagen ampliada generada mediante la red neuronal generativa. Entrenar de manera conjunta la red neuronal generativa y la red neuronal discriminadora fomenta que el sistema de ampliación de imágenes genere imágenes ampliadas difíciles de distinguir de las imágenes reales.

El sistema de ampliación de imágenes puede ser utilizado para cualquiera de entre una variedad de aplicaciones, por ejemplo, el redimensionamiento de imágenes incluidas en componentes digitales que tengan que presentarse en bloques definidos de varias dimensiones para permitir que las imágenes llenen los bloques sin ser distorsionadas.

Tal y como se utiliza a lo largo del presente documento, la expresión «componente digital» se refiere a una unidad separada de contenido digital o información digital que pueda incluir uno o más de, por ejemplo, imágenes, videoclips, clips de audio, clips multimedia, segmentos de texto o localizadores uniformes de recursos (URL). Un componente digital se puede almacenar electrónicamente en un dispositivo de memoria físico como un único archivo o en una recopilación de archivos, y los componentes digitales pueden adoptar la forma de archivos de vídeo, archivos de audio, archivos multimedia, archivos de imagen o archivos de texto, y pueden incluir la transmisión de vídeos, la transmisión de audio, publicaciones de redes sociales, publicaciones de blogs, y/o información publicitaria, de forma que un anuncio es un tipo de componente digital. Por lo general, un componente digital es definido por (o proporcionado por) un único proveedor o fuente (por ejemplo, un anunciante, un editor u otro proveedor de contenidos), pero un componente digital proporcionado por una fuente podría ser mejorado con datos de otra fuente (por ejemplo, información meteorológica, información de sucesos en tiempo real u otra información obtenida de otra fuente).

Estas características y otras adicionales se describen a continuación de manera más detallada.

La FIGURA 1 muestra un sistema de ampliación de imágenes 100 a modo de ejemplo. El sistema de ampliación de imágenes 100 es un ejemplo de un sistema implementado como programas informáticos en uno o más ordenadores en una o más ubicaciones en los que se implementan los sistemas, los componentes y las técnicas que se describen a continuación.

El sistema de ampliación de imágenes 100 está configurado para procesar una imagen 102 para generar una imagen ampliada 104 que amplíe de forma realista la imagen 102 más allá de sus bordes originales. Más concretamente, una matriz bidimensional (2D) de píxeles que represente la imagen ampliada 104 tiene por lo general más filas, más columnas o ambas que la imagen 102. En el ejemplo representado en la FIGURA 1, la imagen ampliada 104 tiene una o más columnas 106 adicionales en su lado derecho en comparación con la imagen 102 original. En un ejemplo concreto, la imagen 102 puede presentar unas dimensiones de 100×100 , mientras que la imagen ampliada 104 puede presentar unas dimensiones de 100×120 . En otro ejemplo particular, la imagen 102 puede presentar unas dimensiones de 80×120 , mientras que la imagen ampliada puede presentar unas dimensiones de 90×125 . Las unidades de la dimensión pueden ser cualquier unidad adecuada tales como un número de píxeles. Un ejemplo de imagen ampliada generada mediante el sistema 100 se muestra en la FIG. 3.

En algunos casos, antes de generar una imagen ampliada 104, el sistema de ampliación de imágenes 100 puede encoger o expandir de manera uniforme la imagen 102 original (es decir, conservando su relación de aspecto y evitando así distorsiones). En un ejemplo concreto, para generar una imagen de salida con una dimensionalidad de 50×120 a partir de una imagen de entrada con una dimensionalidad de 100×200 , el sistema de ampliación de imágenes puede primero encoger la imagen de entrada de manera uniforme a una dimensionalidad de 50×100 y, a continuación, generar una imagen ampliada con una dimensionalidad de 50×120 a partir de la imagen encogida de dimensionalidad 50×100 .

La imagen 102 (y la imagen ampliada 104) puede representarse en cualquier formato apropiado, por ejemplo, como una matriz 2D de píxeles que estén asociados cada uno con una intensidad de escala de grises correspondiente o con un vector de color que represente el color del píxel en un espacio de color apropiado. El espacio de color puede ser, por ejemplo, un espacio de color rojo-verde-azul (RGB) o un espacio de color CIELAB.

Para generar la imagen ampliada 104, el sistema 100 determina una entrada de red generativa 108 a partir de la imagen 102 y procesa la entrada de red generativa 108 utilizando una red neuronal generativa 110. A continuación, se exponen varios **ejemplos** de entradas de red generativa.

En un ejemplo, la entrada de red generativa 108 incluye una imagen inicial 112 y una imagen enmascarada 114 que tienen la misma dimensionalidad (es decir, número de filas y columnas) que la imagen ampliada que ha de ser generada por la red neuronal generativa 110. La imagen inicial 112 incluye: (i) una primera porción 112-A que corresponde a la imagen 102, y (ii) una segunda porción 112-B que tiene valores de píxel predeterminados (por ejemplo, valores de píxel negros o blancos). Es decir, la imagen base 112 incluye la imagen 102, y puede entenderse que amplía la imagen 102 en una o más filas o columnas de valores de píxel predeterminados. La imagen enmascarada 114 identifica qué porciones de la imagen inicial corresponden a: (i) la imagen 102, y (ii) los valores de píxel predeterminados. Por ejemplo, los píxeles de la imagen enmascarada 114 correspondientes al subconjunto correcto de la imagen inicial 112 que coincide con la imagen 102 pueden tener un primer valor (por ejemplo, el valor 0), y los píxeles restantes de la imagen enmascarada 114 pueden tener un segundo valor (por ejemplo, el valor 1). En este ejemplo, la imagen inicial 112 y la imagen enmascarada 114 pueden concatenarse espacialmente por canales para formar la entrada de red generativa 108.

En otro ejemplo, la entrada de red generativa 108 puede incluir solo la imagen inicial 112, es decir, sin la imagen enmascarada 114.

En otro ejemplo, la entrada de red generativa 108 puede corresponder directamente a la imagen 102, es decir, sin incluir la imagen inicial 112 ni la imagen enmascarada 114.

La red neuronal generativa 110 está configurada para procesar la entrada de red generativa 108 de conformidad con valores entrenados de un conjunto de parámetros de red neuronal generativa para generar la imagen ampliada 104. La red neuronal generativa 110 tiene por lo general una arquitectura de red neuronal convolucional, es decir, una arquitectura de red neuronal que incluye una o más capas de red neuronal convolucional (y, opcionalmente, otros tipos apropiados de capas de red neuronal).

El sistema de entrenamiento adversario 116 está configurado para determinar valores entrenados de los parámetros de red neuronal generativa que permiten a la red neuronal generativa 110 generar imágenes ampliadas 104 realistas. El sistema de entrenamiento adversario 116 entrena de manera conjunta la red neuronal generativa 110 junto con una red neuronal discriminadora utilizando una función objetivo de pérdida adversaria, como se va a describir de manera más detallada haciéndose referencia a la FIGURA 2.

En algunos casos, el sistema 100 puede crear imágenes ampliadas «panorámicas» mediante la ampliación repetida de una imagen original. Por ejemplo, el sistema 100 puede ampliar una imagen original para generar una primera imagen ampliada, ampliar a continuación una porción recortada de la primera imagen ampliada para generar una segunda imagen ampliada, y así sucesivamente. Por último, el sistema puede concatenar las imágenes ampliadas para generar una imagen panorámica que amplíe considerablemente la imagen original, por ejemplo, en un factor de $2 \times$, $3 \times$, $5 \times$, o cualquier otro factor apropiado.

Las imágenes ampliadas 104 generadas por el sistema 100 pueden utilizarse en cualquiera de una gran variedad de aplicaciones. A continuación, se exponen varios **ejemplos**.

En un ejemplo, en aplicaciones de realidad virtual, en ocasiones es necesario simular que una imagen es capturada desde una orientación de cámara diferente a la que se utilizó realmente para capturar una imagen original. El sistema de ampliación de imágenes 100 puede utilizarse para simular una imagen de esta manera rellenando contenido fuera de los límites de la imagen original.

En otro ejemplo, en aplicaciones de cosido panorámico (por ejemplo, donde múltiples imágenes se «cosen» entre sí para generar una única imagen combinada), algunas técnicas convencionales requieren recortar los bordes dentados de las proyecciones cosidas para conseguir una panorámica rectangular. El sistema de ampliación de imágenes 100 puede obviar la necesidad de recortar los bordes dentados generando imágenes ampliadas que en su lugar rellenen los huecos entre los bordes dentados, permitiendo así un cosido panorámico más efectivo.

En otro ejemplo, el sistema de ampliación de imágenes 100 puede utilizarse para permitir que los vídeos que se capturen filmados con una determinada relación de aspecto se visualicen en una pantalla con una relación de aspecto diferente sin recurrir al estiramiento ni al recorte. En particular, el sistema de ampliación de imágenes 100 puede utilizarse para procesar cada fotograma de vídeo de un vídeo para generar un fotograma de vídeo ampliado correspondiente, por ejemplo, que tenga la misma relación de aspecto que la pantalla en la que se va a visualizar el vídeo.

En otro ejemplo, el sistema de ampliación de imágenes 100 puede ser utilizado por un sistema de distribución de componentes digitales para ampliar imágenes que hayan de visualizarse en bloques definidos junto a resultados de búsqueda o en sitios web de terceros para que las imágenes puedan llenar los bloques sin recurrir al estiramiento ni al recorte. El uso del sistema de ampliación de imágenes 100 como parte de un sistema de distribución de componentes digitales se describe de manera más detallada haciéndose referencia a la FIGURA 4.

La FIGURA 2 es un flujo de datos 200 a modo de ejemplo que ilustra operaciones que puede realizar un sistema de entrenamiento adversario 116 para entrenar la red neuronal generativa 110 de manera conjunta con la red neuronal discriminadora 202 utilizando una función objetivo de pérdida adversaria 204. Por lo general, la red neuronal generativa

110 se entrena para generar imágenes ampliadas que amplíen de forma realista imágenes de entrada, y la red neuronal discriminadora 202 se entrena para generar salidas discriminadoras que caractericen la probabilidad de que las imágenes de entrada se hayan generado utilizando la red neuronal generativa 110 (es decir, en contraposición a que sean imágenes naturales). El entrenamiento de la red neuronal generativa 110 se va a describir a continuación de manera detallada, seguido de una descripción del entrenamiento de la red neuronal discriminadora 202.

La red neuronal generativa 110 se entrena a lo largo de múltiples iteraciones de entrenamiento y, en cada iteración de entrenamiento, los valores de parámetro actuales de los parámetros de red neuronal generativa se ajustan basándose en un lote actual (es decir, un conjunto) de **ejemplos** de entrenamiento. Cada ejemplo de formación incluye: (i) una entrada de red generativa de entrenamiento, y (ii) una imagen ampliada objetivo que debe ser generada por la red neuronal generativa 110 mediante el procesamiento de la entrada de red generativa de entrenamiento correspondiente.

Las imágenes ampliadas objetivo 206 incluidas en los **ejemplos** de entrenamiento son por lo general imágenes reales (es decir, naturales). La entrada de red generativa de entrenamiento correspondiente a una imagen ampliada objetivo 206 puede representarse de cualquiera de diversas maneras, como se describe haciéndose referencia a la FIGURA 1. En un ejemplo, la entrada de red generativa de entrenamiento correspondiente a una imagen ampliada objetivo 206 incluye una imagen inicial 208 y una imagen enmascarada 210. La imagen inicial 208 se genera «enmascarando» una porción de la imagen ampliada objetivo 206, por ejemplo, estableciendo los valores de píxel en una porción de la imagen ampliada objetivo 206 en un valor predeterminado (por ejemplo, el valor 0). La imagen enmascarada 210 identifica la porción de la imagen ampliada objetivo 206 que ha sido enmascarada en la imagen inicial. La porción de la imagen ampliada objetivo que está incluida en la entrada de red generativa de entrenamiento puede denominarse en este documento «imagen de entrenamiento».

Por lo general, el lote actual de **ejemplos** de entrenamiento puede incluir múltiples **ejemplos** de entrenamiento. Sin embargo, por conveniencia, la descripción que sigue a continuación va a hacer referencia a un ejemplo de formación concreto del lote actual.

La red neuronal generativa 110 procesa la entrada de red generativa de entrenamiento incluida en el ejemplo de entrenamiento para generar una imagen ampliada de entrenamiento 212 correspondiente. Tras generar la imagen ampliada de entrenamiento 212, el sistema de entrenamiento adversario 116 procesa la imagen ampliada de entrenamiento 212 utilizando la red neuronal discriminadora 202 para generar una salida discriminadora que caracterice la probabilidad de que la imagen ampliada de entrenamiento haya sido generada por la red neuronal generativa 110. Así, la salida discriminadora caracteriza si la imagen ampliada de entrenamiento 212 generada por la red neuronal generativa 110 es una imagen realista que amplía de manera creíble la imagen de entrenamiento. Tal y como se va a describir más detalladamente a continuación, la función objetivo de pérdida adversaria 204 depende de la salida discriminadora generada por la red neuronal discriminadora 202.

En algunos casos, en lugar de procesar directamente la imagen ampliada de entrenamiento 212, el sistema de entrenamiento adversario 116 primero sobrescribe la imagen de entrenamiento sobre la porción correspondiente de la imagen ampliada de entrenamiento 212. Es decir, el sistema de entrenamiento adversario 116 sobrescribe la imagen de entrenamiento sobre la porción de la imagen ampliada de entrenamiento 212 que debe coincidir con la imagen de entrenamiento, sin modificar la porción de la imagen ampliada de entrenamiento 212 que se extiende más allá de los bordes de la imagen de entrenamiento. Modificar la imagen ampliada de entrenamiento 212 de esta manera promueve que la salida discriminadora caracterice no solo si la imagen ampliada de entrenamiento es inherentemente una imagen realista, sino también si la imagen ampliada de entrenamiento es una ampliación realista de la imagen de entrenamiento original.

Antes de proporcionar la imagen ampliada de entrenamiento 212 a la red neuronal discriminadora 202, la imagen enmascarada 210 puede concatenarse espacialmente por canales a la imagen ampliada de entrenamiento 212.

Además de procesar la imagen ampliada de entrenamiento 212 (o una entrada basada en la imagen ampliada de entrenamiento 212), la red neuronal discriminadora 202 puede estar «condicionada» a una representación de características semánticas de la imagen ampliada objetivo 206 del ejemplo de entrenamiento. Es decir, la salida discriminadora generada por la red neuronal discriminadora 202 para la imagen ampliada de entrenamiento 212 puede depender de la representación de características semánticas de la imagen ampliada objetivo 206.

La representación de características semánticas de la imagen ampliada objetivo 206 se refiere a una recopilación ordenada de valores numéricos, por ejemplo, un vector o una matriz de valores numéricos, que, de manera implícita o explícita, caracterice los contenidos de la imagen ampliada objetivo 206. El sistema de entrenamiento adversario 116 determina que la representación de características semánticas de la imagen ampliada objetivo 206 es una salida intermedia generada por una red neuronal de procesamiento de imágenes preentrenada mediante el procesamiento de la imagen ampliada objetivo 206. Una salida intermedia de una red neuronal se refiere a una salida generada por una o más capas intermedias de la red neuronal, es decir, capas que siguen a la capa de entrada pero preceden a la capa de salida.

La red neuronal de procesamiento de imágenes puede tener cualquier arquitectura de red neuronal apropiada (por ejemplo, una arquitectura InceptionV3) y puede ser preentrenada para que realice cualquiera de una variedad de tareas de procesamiento de imágenes, por ejemplo, tareas de clasificación o tareas de regresión. En un ejemplo concreto, la red neuronal de procesamiento de imágenes puede ser preentrenada para realizar una tarea de clasificación mediante la generación de una salida que caracterice las probabilidades respectivas de que una imagen de entrada represente un objeto de cada una de un número predeterminado de clases de objetos (por ejemplo, persona, vehículo, bicicleta, etc.). En otro ejemplo concreto, la red neuronal de procesamiento de imágenes puede ser preentrenada para realizar una tarea de regresión mediante la generación de una salida que caracterice las ubicaciones de los cuadros delimitadores que se prevea que van a rodear objetos representados en la imagen de entrada. El hecho de que la red neuronal de procesamiento de imágenes esté «preentrenada» se refiere a que la red haya sido entrenada previamente para realizar la tarea de procesamiento de imágenes.

El sistema de entrenamiento adversario 116 utiliza un motor de normalización 214 para normalizar la representación semántica de la imagen ampliada objetivo antes de utilizarla para condicionar la red neuronal discriminadora 202. Por ejemplo, el motor de normalización 214 puede determinar la representación semántica normalizada C_{norm} como:

$$C_{norm} = \frac{C - E[C]}{\|C - E[C]\|_2} \quad (1)$$

donde C es la representación semántica no normalizada, $E[C]$ es el valor esperado de las representaciones semánticas de las imágenes (por ejemplo, a partir de un conjunto de imágenes de entrenamiento), y $\|\cdot\|_2$ se refiere a una norma L_2 . Asimismo, el sistema de entrenamiento adversario 116 puede procesar la representación semántica normalizada mediante una o más capas de red neuronal 216 (por ejemplo, capas totalmente conectadas) que se entrenen de manera conjunta con las redes neuronales generativa y discriminadora antes de utilizar la representación semántica para condicionar la red neuronal discriminadora.

La red neuronal discriminadora 202 está condicionada a la representación semántica de la imagen ampliada objetivo de cualquiera de una variedad de formas. En un ejemplo, el sistema de entrenamiento adversario 116 puede proporcionar la representación semántica de la imagen ampliada objetivo como una entrada adicional a la red neuronal discriminadora, por ejemplo, a una capa de entrada o una capa intermedia de la red neuronal discriminadora. En otro ejemplo, el sistema de entrenamiento adversario 116 puede determinar la salida discriminadora basándose al menos en parte en una medida de similitud entre (i) la representación semántica de la imagen ampliada objetivo y (ii) una salida intermedia de la red neuronal discriminadora, por ejemplo, la salida discriminadora D puede venir dada por:

$$D = D_0 + \langle D_N, S \rangle \quad (2)$$

donde D_0 es la salida escalar de una capa final de red neuronal de la red neuronal discriminadora, D_N es una salida intermedia de la red neuronal discriminadora (por ejemplo, generada por la capa intermedia 218), S es la representación semántica de la imagen ampliada objetivo, y $\langle \cdot, \cdot \rangle$ se refiere a la operación de producto escalar.

Condicionar la red neuronal discriminadora 202 a la representación semántica puede estabilizar el entrenamiento conjunto de la red neuronal generativa 110 y la red neuronal discriminadora 202, y también mejorar la calidad de las imágenes ampliadas generadas por la red neuronal generativa 110 entrenada. En concreto, condicionar la red neuronal discriminadora 202 a la representación semántica puede permitir a la red neuronal discriminadora 202 generar una salida discriminadora que dependa de todo el contenido semántico de la imagen ampliada objetivo, incluyendo a partir de la porción enmascarada de la imagen ampliada objetivo. La representación semántica puede proporcionar información condicionante efectiva (es decir, que permita a la red neuronal discriminadora generar salidas discriminadoras más precisas) como resultado de la caracterización del contexto «global» de la imagen ampliada objetivo. En cambio, condicionar la red discriminadora directamente a los valores de píxel de la imagen ampliada objetivo puede ser menos efectivo, por ejemplo, puesto que los valores de píxel son inherentemente «locales» y potencialmente ruidosos. El condicionamiento de la red neuronal discriminadora a la representación semántica también puede permitir que la red neuronal discriminadora se entrene con mayor efectividad (por ejemplo, en menos iteraciones de entrenamiento) al reducir la carga de la red neuronal discriminadora para tener en cuenta el contenido semántico directamente en la salida de su capa final. Por ejemplo, en referencia a la ecuación (2), la salida discriminadora puede tener en cuenta el contenido semántico mediante un producto interno entre una salida intermedia de la red neuronal discriminadora y la representación semántica.

Tras generar la salida discriminadora, el sistema de entrenamiento adversario 116 puede ajustar los valores actuales de los parámetros de red neuronal generativa utilizando gradientes de una función objetivo $\mathcal{L}_G$ que incluya la función objetivo de pérdida adversaria $\mathcal{L}_{adv}$ 204 y una función objetivo de reconstrucción $\mathcal{L}_{rec}$ 220, por ejemplo:

$$\mathcal{L}_G = \mathcal{L}_{rec} + \lambda \cdot \mathcal{L}_{adv} \quad (3)$$

$$\mathcal{L}_{rec} = \|x - \hat{x}\|_1 \quad (4)$$

$$\mathcal{L}_{adv} = -D(\hat{x}) \quad (5)$$

donde λ es un hiperparámetro escalar, x es la imagen ampliada objetivo, $\hat{x}$ es la imagen ampliada de entrenamiento generada por la red neuronal generativa, $\|\cdot\|_1$ es una norma L_1 , y $D(\hat{x})$ es la salida discriminadora generada por la red neuronal discriminadora mediante el procesamiento de la imagen ampliada de entrenamiento (o una entrada basada en la imagen ampliada de entrenamiento).

Las funciones objetivo caracterizadas por las ecuaciones (3) - (5) se proporcionan como **ejemplos** ilustrativos de las funciones objetivo, pero son posibles otras funciones objetivo. En general, la función objetivo de reconstrucción $\mathcal{L}_{rec}$ puede caracterizar la similitud entre la imagen ampliada objetivo x y la imagen ampliada de entrenamiento $\hat{x}$ en cualquiera de una variedad de formas, y la función objetivo de pérdida adversaria $\mathcal{L}_{adv}$ puede depender de la salida discriminadora $D(\hat{x})$ en cualquiera de una variedad de maneras.

El sistema de entrenamiento adversario 116 entrena de manera conjunta la red neuronal discriminadora 202 y la red neuronal generativa 110, por ejemplo, alternando entre el entrenamiento de la red neuronal discriminadora 202 y la red neuronal generativa 110 para números predeterminados de iteraciones de entrenamiento.

En cada iteración de entrenamiento durante el entrenamiento de la red neuronal discriminadora, los valores de parámetros actuales de la red neuronal discriminadora se ajustan basándose en un lote actual de **ejemplos** de entrenamiento. Al igual que con anterioridad, el lote actual de **ejemplos** de entrenamiento puede incluir múltiples **ejemplos** de entrenamiento, pero, por motivos de simplicidad, la descripción que sigue a continuación va a hacer referencia a un ejemplo de entrenamiento en particular.

Para entrenar la red neuronal discriminadora 202, el sistema de entrenamiento adversario 116 procesa la entrada de red generativa de entrenamiento del ejemplo de entrenamiento mediante la red neuronal generativa 110 para generar una imagen ampliada de entrenamiento correspondiente. El sistema de entrenamiento adversario 116 procesa tanto la imagen ampliada de entrenamiento (o una entrada basada en la imagen ampliada de entrenamiento) como la imagen ampliada objetivo correspondiente utilizando la red neuronal discriminadora para generar salidas discriminadoras correspondientes. A continuación, el sistema de entrenamiento adversario 116 ajusta los valores de parámetros actuales de la red neuronal discriminadora 202 basándose en una función objetivo discriminadora que depende de las salidas discriminadoras generadas para la imagen ampliada de entrenamiento y la imagen ampliada objetivo. A modo de ejemplo, la función objetivo discriminadora $\mathcal{L}_{disc}$ puede venir dada por:

$$\mathcal{L}_{disc} = \text{ReLU}(1 - D(x)) + \text{ReLU}(1 + D(\hat{x})) \quad (6)$$

donde $\text{ReLU}(\cdot)$ es una función unitaria lineal rectificadora, $D(x)$ es la salida discriminadora generada por la red neuronal discriminadora para la imagen ampliada objetivo, y $D(\hat{x})$ es la salida discriminadora generada por la red neuronal discriminadora para la imagen ampliada de entrenamiento generada por la red neuronal generativa. La función objetivo discriminadora caracterizada por la ecuación (6) se proporciona únicamente con fines ilustrativos, y son posibles otras funciones objetivo discriminadoras.

La red neuronal discriminadora 202 puede tener cualquier arquitectura de red neuronal apropiada. En un ejemplo, la red neuronal discriminadora 202 tiene seis capas convolucionales con funciones de activación ReLU con fugas, seguidas de una capa totalmente conectada.

Durante el entrenamiento de la red neuronal generativa y la red neuronal discriminadora, el sistema de entrenamiento adversario 116 puede determinar gradientes de las funciones objetivo usando, por ejemplo, técnicas de retropropagación. El sistema de entrenamiento adversario 116 puede utilizar los gradientes de las funciones objetivo para ajustar los valores de parámetros actuales de la red neuronal generativa y la red neuronal discriminadora mediante cualquier procedimiento de optimización de descenso de gradiente apropiado, por ejemplo, Adam o RMSprop.

El sistema de entrenamiento adversario 116 puede continuar entrenando de manera conjunta la red neuronal generativa 110 y la red neuronal discriminadora 202 hasta que se satisfaga un criterio para la finalización del entrenamiento, por ejemplo, hasta que se haya realizado un número predeterminado de iteraciones de entrenamiento para cada red neuronal, o hasta que se satisfaga un criterio relativo al rendimiento.

La FIGURA 3 es una ilustración de un ejemplo de una imagen ampliada 300 generada por el sistema de ampliación de imágenes 100 descrito haciéndose referencia a la FIGURA 1 mediante el procesamiento de la imagen original 302. Se puede observar que la imagen ampliada 300 es una ampliación realista de la imagen original 302 que conserva las características semánticas de alto nivel y las estructuras y las texturas de bajo nivel de la imagen original 302.

La FIGURA 4 es un diagrama de bloques de un entorno 400 a modo de ejemplo en el que un sistema de distribución de componentes digitales 410 utiliza el sistema de ampliación de imágenes 100 mientras se transmiten componentes digitales desde una base de datos de componentes digitales 416 para su presentación con documentos electrónicos. Tal y como se va a describir a continuación de manera más detallada, el sistema de distribución de componentes digitales puede transmitir componentes digitales en respuesta a las solicitudes para que los componentes digitales se presenten con documento electrónico en dispositivos de usuario, por ejemplo, en bloques junto a resultados de búsqueda o en sitios web de terceros.

Después de identificar un componente digital que haya de transmitirse en respuesta a una solicitud, el sistema de distribución puede identificar que la imagen incluida en el componente digital es apta para la ampliación de imágenes. En un ejemplo, el sistema de distribución puede identificar que la imagen es apta para la ampliación de imágenes basándose en el tamaño (por ejemplo, la relación de aspecto) de la imagen que difiere (por ejemplo, en al menos una cantidad umbral) del tamaño del bloque en el que se va a presentar la imagen. En este ejemplo, la solicitud de componente digital puede incluir datos que caractericen el tamaño del bloque donde se va a presentar el componente digital. En otro ejemplo, el sistema de distribución puede identificar que la imagen es apta para la ampliación de imágenes basándose en que un proveedor del componente digital ha especificado que otros elementos visuales del componente digital (por ejemplo, texto, logotipos y elementos interactivos) no deben superponerse a la imagen del componente digital.

En respuesta a la identificación de que la imagen incluida en el componente digital es apta para la ampliación, el sistema de distribución puede procesar la imagen utilizando el sistema de ampliación de imágenes 100 para generar una imagen ampliada. En un ejemplo, el sistema de distribución puede generar una imagen ampliada con una relación de aspecto que coincida con la relación de aspecto del bloque en el que se visualizará la imagen ampliada. En otro ejemplo, el sistema de distribución puede generar una imagen ampliada y, a continuación, superponer los elementos visuales adicionales del componente digital (por ejemplo, texto, logotipos y elementos interactivos) solo en la porción ampliada de la imagen (es decir, sin modificar la imagen original).

Después de ampliar la imagen incluida en el componente digital, el sistema de distribución puede transmitir el componente digital con la imagen ampliada para su presentación en un bloque en un dispositivo de usuario que genere la solicitud del componente.

El entorno 400 a modo de ejemplo y las operaciones realizadas por el sistema de distribución se describen a continuación de manera más detallada.

El sistema 400 a modo de ejemplo incluye una red 402, tal como una red de área local (LAN), una red de área amplia (WAN), Internet o una combinación de estas. La red 402 conecta los servidores de documentos electrónicos 404, los dispositivos cliente 406, los servidores de componentes digitales 408 y el sistema de distribución de componentes digitales 410 (también denominado «sistema de distribución» 410). El entorno 400 a modo de ejemplo puede incluir muchos servidores de documentos electrónicos 404, dispositivos cliente 406 y servidores de componentes digitales 408 diferentes.

Un dispositivo cliente 406 es un dispositivo electrónico que puede solicitar y recibir recursos a través de la red 402. Los **ejemplos** de dispositivos cliente 406 incluyen los ordenadores personales, los dispositivos de comunicación móviles (por ejemplo, los teléfonos móviles) y otros dispositivos que pueden enviar y recibir datos a través de la red 402. Un dispositivo cliente 406 incluye normalmente una aplicación de usuario tal como un navegador web para facilitar el envío y la recepción de datos a través de la red 402, pero las aplicaciones nativas ejecutadas por el dispositivo cliente 406 también pueden facilitar el envío y la recepción de datos a través de la red 402.

Un documento electrónico son datos que presentan un conjunto de contenidos en un dispositivo cliente 406. Los **ejemplos** de documentos electrónicos incluyen las páginas web, los documentos de procesamiento de texto, los documentos en formato de documento portátil (PDF), las imágenes, los vídeos, las páginas de resultados de búsqueda y las fuentes de alimentación. Las aplicaciones nativas (por ejemplo, las «apps») tales como las aplicaciones instaladas en dispositivos informáticos móviles, tabletas o dispositivos informáticos de sobremesa también son **ejemplos** de documentos electrónicos. Los documentos electrónicos pueden ser proporcionados a los dispositivos cliente 406 por los servidores de documentos electrónicos 404 («*Electronic Doc Servers*»). Por ejemplo, los servidores de documentos electrónicos 404 pueden incluir servidores que alojen sitios web de editores. En este ejemplo, el dispositivo cliente 406 puede iniciar una solicitud para la página web de un editor dado, y el servidor electrónico 404 que aloja la página web del editor dado puede responder a la solicitud enviando instrucciones ejecutables por ordenador que inicien la presentación de la página web dada en el dispositivo cliente 406.

En otro ejemplo, los servidores de documentos electrónicos 404 pueden incluir servidores de aplicaciones desde los que los dispositivos cliente 406 puedan descargar aplicaciones. En este ejemplo, el dispositivo cliente 406 puede descargar los archivos necesarios para instalar una aplicación en el dispositivo cliente 406 y, a continuación, ejecutar la aplicación descargada de manera local.

Los documentos electrónicos pueden incluir una variedad de contenidos. Por ejemplo, un documento electrónico puede incluir contenido estático (por ejemplo, texto u otro contenido especificado) que se encuentre dentro del propio documento

electrónico y/o no cambia con el tiempo. Los documentos electrónicos también pueden incluir contenido dinámico que pueda cambiar con el tiempo o en función de las solicitudes. A modo de ejemplo, un editor de un documento electrónico determinado puede mantener una fuente de datos que se utilice para llenar partes del documento electrónico. En este ejemplo, el documento electrónico dado puede incluir una o más etiquetas o *scripts* que hagan que el dispositivo cliente 406 solicite contenido de la fuente de datos cuando el documento electrónico dado es procesado (por ejemplo, renderizado o ejecutado) por un dispositivo cliente 406. El dispositivo cliente 406 integra el contenido obtenido de la fuente de datos en el documento electrónico dado para crear un documento electrónico compuesto que incluya el contenido obtenido de la fuente de datos.

En algunas situaciones, un documento electrónico dado puede incluir una o más etiquetas de componentes digitales o *scripts* de componentes digitales que hagan referencia al sistema de distribución de componentes digitales 410. En estas situaciones, las etiquetas de componentes digitales o los *scripts* de componentes digitales son ejecutados por el dispositivo cliente 406 cuando el documento electrónico dado es procesado por el dispositivo cliente 406. La ejecución de las etiquetas de componentes digitales o los *scripts* de componentes digitales configura el dispositivo cliente 406 para generar una solicitud de uno o más componentes digitales 412 (denominada «solicitud de componente»), que se transmite a través de la red 402 al sistema de distribución de componentes digitales 410. Por ejemplo, una etiqueta de componente digital o un *script* de componente digital pueden permitir al dispositivo cliente 406 generar una solicitud de datos paquetizada que incluya un encabezado y datos de carga útil. La solicitud de componente 412 puede incluir datos de evento que especifiquen características tales como un nombre (o ubicación de red) de un servidor desde el que se solicita el componente digital, un nombre (o ubicación de red) del dispositivo solicitante (por ejemplo, el dispositivo cliente 406), y/o información que el sistema de distribución de componentes digitales 410 puede utilizar para seleccionar uno o más componentes digitales proporcionados en respuesta a la solicitud. La solicitud de componente 412 es transmitida, por el dispositivo cliente 406, a través de la red 402 (por ejemplo, una red de telecomunicaciones) a un servidor del sistema de distribución de componentes digitales 410.

La solicitud de componente 412 puede incluir datos de evento que especifiquen otras características de evento tales como el documento electrónico que se solicita y las características de las ubicaciones del documento electrónico en las que se pueda presentar el componente digital. Por ejemplo, los datos de evento que especifiquen una referencia (por ejemplo, la URL) a un documento electrónico (por ejemplo, página web) en el que se presentará el componente digital, las ubicaciones disponibles de los documentos electrónicos que estén disponibles para presentar componentes digitales, los tamaños de las ubicaciones disponibles, y/o los tipos de medios que sean elegibles para su presentación en las ubicaciones pueden proporcionarse al sistema de distribución de componentes digitales 410. Del mismo modo, los datos de evento que especifiquen palabras clave asociadas al documento electrónico («palabras clave del documento») o entidades (por ejemplo, personas, lugares o cosas) a las que se haga referencia en el documento electrónico también pueden incluirse en la solicitud de componente 412 (por ejemplo, como datos de carga útil) y proporcionarse al sistema de distribución de componentes digitales 410 para facilitar la identificación de componentes digitales que son elegibles para su presentación con el documento electrónico. Los datos de evento también pueden incluir una consulta de búsqueda que se haya enviado desde el dispositivo cliente 406 para obtener una página de resultados de búsqueda, y/o datos que especifiquen los resultados de búsqueda y/o contenido de texto, audible u otro contenido visual que se incluya en los resultados de búsqueda.

Las solicitudes de componente 412 también pueden incluir datos de evento relacionados con otra información, como información que haya proporcionado un usuario del dispositivo cliente, información geográfica que indique un estado o región desde el que se enviara la solicitud de componente u otra información que proporcione contexto para el entorno en el que se va a visualizar el componente digital (por ejemplo, una hora del día de la solicitud de componente, un día de la semana de la solicitud de componente, un tipo de dispositivo en el que se va a visualizar el componente digital tal como un dispositivo móvil o un dispositivo de tableta). Las solicitudes de componente 412 pueden transmitirse, por ejemplo, a través de una red paquetizada, y las propias solicitudes de componente 412 pueden formatearse como datos paquetizados con una cabecera y datos de carga útil. La cabecera puede especificar un destino del paquete y los datos de la carga útil pueden incluir cualquier tipo de información de las expuestas anteriormente.

El sistema de distribución de componentes 410 elige los componentes digitales que se presentarán con el documento electrónico dado en respuesta a la recepción de la solicitud de componente 412 y/o utilizando la información incluida en la solicitud de componente 412. En algunas implementaciones, un componente digital se selecciona (utilizando las técnicas descritas en el presente documento) en menos de un segundo para evitar errores que podrían ser ocasionados por el retraso en la selección del componente digital. Por ejemplo, los retrasos en el suministro de componentes digitales en respuesta a una solicitud de componente 412 pueden dar lugar a errores de carga de página en el dispositivo cliente 406 o provocar que partes del documento electrónico permanezcan sin contenido incluso después de que otras partes del documento electrónico se presenten en el dispositivo cliente 406. Además, a medida que aumenta el retraso en el suministro del componente digital al dispositivo cliente 406, es más probable que el documento electrónico ya no se presente en el dispositivo cliente 406 cuando el componente digital se entregue al dispositivo cliente 406, teniendo así un impacto negativo en la experiencia de un usuario con el documento electrónico. Además, los retrasos en el suministro del componente digital pueden tener como resultado una entrega fallida del componente digital, por ejemplo, si el documento electrónico ya no se presenta en el dispositivo cliente 406 cuando se proporciona el componente digital.

En algunas implementaciones, el sistema de distribución de componentes digitales 410 se implementa en un sistema informático distribuido que incluye, por ejemplo, un servidor y un conjunto de múltiples dispositivos informáticos 414 que están interconectados e identifican y distribuyen componentes digitales en respuesta a las solicitudes 412. El conjunto de múltiples dispositivos informáticos 414 operan juntos para identificar un conjunto de componentes digitales que son elegibles para ser presentados en el documento electrónico a partir de un corpus de millones de componentes digitales disponibles (DC1-x). Los millones de componentes digitales disponibles pueden indexarse, por ejemplo, en una base de datos de componentes digitales 416. Cada entrada de índice de componentes digitales puede hacer referencia al componente digital correspondiente y/o incluir parámetros de distribución (DP1-DPx) que contribuyan (por ejemplo, condicionen o limiten) la distribución-transmisión del componente digital correspondiente. Por ejemplo, los parámetros de distribución pueden contribuir a la transmisión de un componente digital requiriendo que una solicitud de componente incluya al menos un criterio que coincida (por ejemplo, ya sea exactamente o con algún nivel de similitud especificado previamente) con uno de los parámetros de distribución del componente digital.

En algunas implementaciones, los parámetros de distribución para un componente digital particular pueden incluir palabras clave de distribución con las que se tenga que coincidir (por ejemplo, en los documentos electrónicos, las palabras clave de documento, o términos especificados en la solicitud de componente 412) para que el componente digital sea elegible para presentarse. Expresado de otro modo, los parámetros de distribución se utilizan para activar la distribución (por ejemplo, la transmisión) de los componentes digitales a través de la red 402. Los parámetros de distribución también pueden requerir que la solicitud de componente 412 incluya información que especifique una región geográfica concreta (por ejemplo, país o estado) y/o información que especifique que la solicitud de componente 412 se ha originado en un tipo concreto de dispositivo cliente (por ejemplo, dispositivo móvil o dispositivo de tableta) para que el componente digital sea elegible para presentarse.

Los parámetros de distribución también pueden especificar un valor de elegibilidad (por ejemplo, una puntuación de clasificación, una oferta o algún otro valor especificado) que sea utilizado para evaluar la elegibilidad del componente digital para su distribución/transmisión (por ejemplo, entre otros componentes digitales disponibles), por ejemplo, por el proceso de evaluación de componentes. En algunas situaciones, el valor de elegibilidad puede especificar una cantidad máxima de compensación que un proveedor del componente digital está dispuesto a presentar en respuesta a la transmisión del componente digital (por ejemplo, para cada instancia de eventos específicos atribuidos a la presentación del componente digital tal como la interacción del usuario con el componente digital).

La identificación del componente digital elegible puede segmentarse en múltiples tareas 417a-417c que a continuación se asignan entre dispositivos informáticos dentro del conjunto de múltiples dispositivos informáticos 414. Por ejemplo, diferentes dispositivos informáticos del conjunto 414 pueden analizar cada uno una parte diferente de la base de datos de componentes digitales 416 para identificar varios componentes digitales que tengan parámetros de distribución que coincidan con la información incluida en la solicitud de componente 412. En algunas implementaciones, cada dispositivo informático dado en el conjunto 414 puede analizar una dimensión de datos diferente (o conjunto de dimensiones) y pasar (por ejemplo, transmitir) los resultados (Res 1-Res 3) 418a-418c del análisis de regreso al sistema de distribución de componentes digitales 410. A modo de ejemplo, los resultados 418a-418c proporcionados por cada uno de los dispositivos informáticos del conjunto 414 pueden identificar un subconjunto de componentes digitales que sean elegibles para su distribución en respuesta a la solicitud de componente y/o un subconjunto de los componentes digitales que tengan determinados parámetros de distribución. La identificación del subconjunto de componentes digitales puede incluir, por ejemplo, la comparación de los datos de evento con los parámetros de distribución, y la identificación del subconjunto de componentes digitales que tengan parámetros de distribución que coincidan con al menos algunas características de los datos de evento.

El sistema de distribución de componentes digitales 410 combina los resultados 418a-418c recibidos del conjunto de múltiples dispositivos informáticos 414 y utiliza la información asociada a los resultados combinados para: (i) seleccionar uno o más componentes digitales que se proporcionarán en respuesta a la solicitud 412, y (ii) determinar los requisitos de transmisión para el uno o más componentes digitales. Por ejemplo, el sistema de distribución de componentes digitales 410 puede seleccionar un conjunto de componentes digitales ganadores (uno o más componentes digitales) basándose en el resultado de uno o más procesos de evaluación de componentes. A su vez, el sistema de distribución de componentes digitales 410 puede generar y transmitir, a través de la red 402, datos de respuesta 420 (por ejemplo, datos digitales que representen una respuesta) que permitan al dispositivo cliente 406 integrar el conjunto de componentes digitales ganadores en el documento electrónico dado, de tal manera que el conjunto de componentes digitales ganadores y el contenido del documento electrónico se presenten juntos en una pantalla del dispositivo cliente 406.

En algunas implementaciones, el dispositivo cliente 406 ejecuta instrucciones incluidas en los datos de respuesta 420, que configuran y permiten al dispositivo cliente 406 obtener el conjunto de componentes digitales ganadores de uno o más servidores de componentes digitales. Por ejemplo, las instrucciones de los datos de respuesta 420 pueden incluir una ubicación de red [por ejemplo, un Localizador Uniforme de Recursos (URL)] y un *script* que haga que el dispositivo cliente 406 transmita una solicitud de servidor (SR) 421 al servidor de componentes digitales 408 para obtener un componente digital ganador dado del servidor de componentes digitales 408. En respuesta a la solicitud, el servidor de componentes digitales 408 identificará el componente digital ganador dado especificado en la solicitud de servidor 421 (por ejemplo, dentro de una base de datos que almacene múltiples componentes digitales) y transmitirá, al dispositivo

cliente 406, datos de componente digital (Datos de CD) 422 que presenten el componente digital ganador dado en el documento electrónico en el dispositivo cliente 406.

Para facilitar la búsqueda de documentos electrónicos, el entorno 400 puede incluir un sistema de búsqueda 450 que identifique los documentos electrónicos mediante el rastreo y la indexación de los documentos electrónicos (por ejemplo, indexados basándose en el contenido rastreado de los documentos electrónicos). Los datos sobre los documentos electrónicos pueden indexarse basándose en el documento electrónico al que estén asociados los datos. Las copias indexadas y, opcionalmente, almacenadas en caché de los documentos electrónicos se almacenan en un índice de búsqueda 452 (por ejemplo, dispositivo(s) de memoria de hardware). Los datos asociados a un documento electrónico son datos que representan contenido incluido en el documento electrónico y/o metadatos para el documento electrónico.

Los dispositivos cliente 406 pueden enviar consultas de búsqueda al sistema de búsqueda 450 a través de la red 402. En respuesta, el sistema de búsqueda 450 accede al índice de búsqueda 452 para identificar los documentos electrónicos que son relevantes para la consulta de búsqueda. El sistema de búsqueda 450 identifica los documentos electrónicos en forma de resultados de búsqueda y devuelve los resultados de búsqueda al dispositivo de cliente 406 en una página de resultados de búsqueda. Un resultado de búsqueda son datos generados por el sistema de búsqueda 450 que identifican un documento electrónico que responde (por ejemplo, es relevante) a una consulta de búsqueda concreta, e incluyen un enlace activo (por ejemplo, un enlace de hipertexto) que hace que un dispositivo cliente solicite datos desde una ubicación de red especificada (por ejemplo, URL) en respuesta a la interacción del usuario con el resultado de búsqueda. Un ejemplo de resultado de búsqueda puede incluir el título de una página web, un fragmento de texto o una porción de una imagen extraída de la página web, y la URL de la página web. Otro ejemplo de resultado de búsqueda puede incluir un título de una aplicación descargable, un fragmento de texto que describa la aplicación descargable, una imagen que represente una interfaz de usuario de la aplicación descargable, y/o una URL a una ubicación desde la que se pueda descargar la aplicación al dispositivo cliente 406. En algunas situaciones, el sistema de búsqueda 450 puede ser parte de, o interactuar con, una tienda de aplicaciones (o un portal en línea) desde la que se puedan descargar aplicaciones para instalarlas en un dispositivo cliente 406 con el fin de presentar información sobre aplicaciones descargables que sean relevantes para una consulta de búsqueda enviada. Al igual que otros documentos electrónicos, las páginas de resultados de búsqueda pueden incluir uno o más espacios en los que se pueden presentar componentes digitales (por ejemplo, anuncios, videoclips, clips de audio, imágenes u otros componentes digitales).

Para seleccionar un componente digital a transmitir en respuesta a una solicitud de componente, el sistema de distribución 410 puede identificar un conjunto de componentes digitales que sean elegibles para ser transmitidos en respuesta a la solicitud de componente. El sistema de distribución 410 puede entonces seleccionar uno o más de los componentes digitales elegibles para ser transmitidos a través de, por ejemplo, un procedimiento de subasta. En algunas implementaciones, el sistema de distribución 410 realiza un procedimiento de subasta clasificando los componentes digitales elegibles de conformidad con sus valores de elegibilidad respectivos, y seleccionando uno o más componentes digitales de entre los mejor clasificados para que se transmitan en respuesta a la solicitud de componente.

A modo de ejemplo, el sistema de distribución 410 puede identificar los componentes digitales A, B y C como elegibles para que se transmitan en respuesta a una solicitud de componente. En este ejemplo, el componente digital A tiene un valor de elegibilidad de 5 \$, el componente digital B tiene un valor de elegibilidad de 1 \$ y el componente digital C tiene un valor de elegibilidad de 5,5 \$, donde los valores de elegibilidad de los componentes digitales representan ofertas asociadas a los componentes digitales. El sistema de distribución 410 puede clasificar (por ejemplo, en orden descendente) los componentes digitales de conformidad con sus valores de elegibilidad respectivos como C, A, B. Por último, el sistema de distribución 410 puede seleccionar el componente digital C mejor clasificado para su transmisión en respuesta a la solicitud de componente.

Después de seleccionar un componente digital para que sea transmitido en respuesta a una solicitud de componente digital, el sistema de distribución 410 determina un requisito de transmisión para el componente digital seleccionado. Un requisito de transmisión especifica una acción que debe realizar el proveedor de un componente digital en respuesta a una transmisión del componente digital. Por ejemplo, el requisito de transmisión puede especificar que el proveedor del componente digital presente una cantidad de compensación en respuesta a la transmisión del componente digital. En algunos casos, la cantidad de compensación especifica una cantidad que debe presentarse por cada instancia de eventos específicos atribuidos a la presentación del componente digital (por ejemplo, interacciones del usuario con el componente digital).

El sistema de distribución 410 puede determinar el requisito de transmisión del componente digital seleccionado basándose en el valor de elegibilidad del componente digital seleccionado y/o los valores de elegibilidad de los otros componentes digitales que se determinaron como elegibles para transmitirse en respuesta a la solicitud de componente. Por ejemplo, el sistema de distribución 410 puede identificar los componentes digitales A, B y C como elegibles para la transmisión en respuesta a una solicitud de componente digital, donde A, B y C tienen valores de elegibilidad respectivos de 5 \$, 1 \$ y 5,5 \$. El sistema de distribución 410 puede seleccionar el componente digital C para la transmisión (ya que tiene el mayor valor de elegibilidad), y puede determinar que el requisito de transmisión para el componente digital C sea el siguiente valor de elegibilidad más elevado de entre los valores de elegibilidad de los componentes digitales elegibles. En este ejemplo, el siguiente valor de elegibilidad más elevado es de 5 \$ (es decir, el valor de elegibilidad del componente

digital A) y, por lo tanto, el sistema de distribución 410 puede determinar que el requisito de transmisión del componente digital C sea 5 \$.

La FIGURA 5 muestra un componente digital que se está visualizando en una pantalla de un dispositivo de usuario, por ejemplo, un *smartphone*. El componente digital incluye una imagen que llena la pantalla del dispositivo y elementos visuales adicionales que incluyen un logotipo 502, un segmento de texto 504 y un elemento interactivo 506 que se superponen a la imagen. En este ejemplo, la imagen incluye una primera porción 508 que se proporcionó al sistema de distribución de componentes digitales, y una segunda porción 510 que se generó utilizando el sistema de ampliación de imágenes 100. Tras recibir una solicitud para que se proporcione un componente digital al dispositivo de usuario, el sistema de distribución de componentes digitales genera una imagen ampliada que llena la pantalla del dispositivo y superpone los elementos adicionales 502, 504 y 506 a la parte ampliada 510 de la imagen. En este ejemplo, el «bloque» donde se visualiza el componente digital corresponde a toda la pantalla del dispositivo de usuario, pero el bloque podría ocupar menos que toda la pantalla.

La FIGURA 6 muestra un componente digital que se está visualizando en un bloque *banner*, por ejemplo, junto con resultados de búsqueda o en un sitio web de terceros. De manera similar a la FIGURA 6, el componente digital incluye una imagen que llena el bloque y elementos adicionales 602. En este ejemplo, la imagen incluye una primera porción 604 que se proporcionó al sistema de distribución de componentes digitales, y una segunda porción 606 que se generó utilizando el sistema de ampliación de imágenes 100. Tras recibir una solicitud para que un componente digital se presente en un bloque *banner* en un dispositivo de usuario, el sistema de distribución de componentes digitales genera una imagen ampliada que llena el bloque y superpone los elementos adicionales 602 a la parte ampliada 606 de la imagen.

La FIGURA 7 es una ilustración de una interfaz de usuario 700 a modo de ejemplo que puede presentarse a una entidad (por ejemplo, un anunciante, un editor u otro proveedor de contenidos) que proporciona un componente digital para su transmisión por un sistema de distribución de componentes digitales (por ejemplo, como se ha descrito haciendo referencia a la FIGURA 4). La interfaz de usuario 700 presenta una imagen 702 incluida en el componente digital proporcionado y múltiples imágenes ampliadas 704-A-C que amplían la imagen 702. Por lo general, cada una de las imágenes ampliadas 704-A-C son diferentes como resultado de ser generadas mediante diferentes procedimientos de ampliación de imágenes. Por ejemplo, una de las imágenes ampliadas puede haberse generado utilizando el sistema de ampliación de imágenes descrito haciéndose referencia a la FIGURA 1, mientras que otra de las imágenes ampliadas puede haberse generado utilizando un procedimiento de ampliación de imágenes basado en las ecuaciones diferenciales parciales (EDP), que se va a describir de manera más detallada más adelante.

La interfaz de usuario 700 requiere al proveedor del componente digital que seleccione una preferida de las imágenes ampliadas (por ejemplo, la que sea más atractiva visualmente para el proveedor del componente digital). El proveedor del componente digital puede seleccionar una imagen ampliada en particular, por ejemplo, haciendo clic en una imagen ampliada concreta mediante un ratón 706 (o de cualquiera de una variedad de otras maneras). Si el sistema de distribución determina posteriormente que la imagen incluida en el componente digital proporcionado es apta para ser ampliada, la distribución genera la ampliación de la imagen utilizando el procedimiento de ampliación de imágenes correspondiente a la imagen ampliada concreta seleccionada por el usuario.

Las imágenes pueden ampliarse de diversas maneras, por ejemplo, mediante el sistema de ampliación de imágenes descrito haciéndose referencia a la FIGURA 1, o mediante un procedimiento de ampliación de imágenes basado en las EDP. En un método de ampliación mediante EDP, los valores de píxel en la porción ampliada de una imagen se identifican como una solución (aproximada o exacta) a una EDP sujeta a condiciones límite especificadas por los valores de píxel en una porción del contorno de la imagen original. La EDP puede ser, por ejemplo, una EDP de difusión, o cualquier otra EDP apropiada. La solución a la EDP sujeta a las condiciones límite puede obtenerse utilizando cualquier técnica numérica apropiada, por ejemplo, una técnica de elementos finitos.

La FIGURA 8 es un diagrama de flujo de un proceso 800 a modo de ejemplo para proporcionar una imagen ampliada en respuesta a una solicitud. Por conveniencia, el proceso 800 se va a describir siendo realizado por un sistema de uno o más ordenadores situados en una o más ubicaciones. A modo de ejemplo, un sistema de distribución, por ejemplo, el sistema de distribución 100 de la FIGURA 1, programado adecuadamente de conformidad con esta memoria descriptiva, puede realizar el proceso 800.

El sistema determina que una imagen proporcionada es apta para la ampliación de imágenes (802). En un ejemplo, el sistema puede recibir una solicitud para que un componente digital se presente en un bloque, por ejemplo, junto a resultados de búsqueda o en una página web de terceros, y determinar a continuación que un componente digital que incluye la imagen proporcionada responde a la solicitud. En este ejemplo, el sistema puede determinar que la imagen proporcionada es apta para la ampliación basándose en el tamaño de la imagen que difiere del tamaño del bloque, por ejemplo, si la imagen y el bloque tienen relaciones de aspecto diferentes.

El sistema proporciona una entrada que incluye la imagen a una red neuronal generativa que procesa la entrada de conformidad con valores entrenados de los parámetros de red neuronal generativa para generar una imagen ampliada (804). La imagen ampliada tiene (i) más filas, más columnas o ambas que la imagen proporcionada, y (ii) se predice que es una ampliación realista de la imagen proporcionada.

La entrada procesada por la red neuronal generativa puede incluir una imagen inicial y una imagen enmascarada, ambas con el mismo número de filas y columnas que la imagen ampliada. La imagen inicial incluye: (i) una primera porción que corresponde a la imagen proporcionada, y (ii) una segunda porción con valores de píxel predeterminados. La imagen enmascarada identifica la primera porción y la segunda porción de la imagen inicial. En un ejemplo, los píxeles de la imagen enmascarada correspondiente a la primera porción de la imagen inicial tienen cada uno un primer valor de píxel, y los píxeles de la imagen enmascarada correspondiente a la segunda porción de la imagen inicial tienen cada uno un segundo valor de píxel diferente.

La red neuronal generativa incluye múltiples capas de red neuronal convolucional y puede tener cualquier arquitectura de red neuronal apropiada. Por ejemplo, la red neuronal generativa puede incluir una o más conexiones de salto, una o más capas de normalización de instancias, o ambas.

La red neuronal generativa ha sido entrenada de manera conjunta con una red neuronal discriminadora utilizando una función objetivo de pérdida adversaria. La red neuronal discriminadora está configurada para procesar una imagen dada y generar una salida discriminadora que caracteriza una probabilidad de que la imagen dada se haya generado utilizando la red neuronal generativa.

El sistema entrena la red neuronal generativa mediante la función objetivo de pérdida adversaria a lo largo de múltiples iteraciones de entrenamiento. En cada iteración de entrenamiento, el sistema procesa una entrada de entrenamiento que incluye una imagen de entrenamiento utilizando la red neuronal generativa para generar una imagen ampliada de entrenamiento que amplía la imagen de entrenamiento y, a continuación, genera una entrada de red neuronal discriminadora basándose en la imagen ampliada de entrenamiento. En un ejemplo, el sistema genera la entrada de red neuronal discriminadora utilizando la imagen de entrenamiento para sobrescribir la porción de la imagen ampliada de entrenamiento correspondiente a la imagen de entrenamiento. El sistema procesa la entrada de red neuronal discriminadora utilizando la red neuronal discriminadora para generar una salida discriminadora que caracteriza una probabilidad de que la entrada de red neuronal discriminadora se haya generado utilizando la red neuronal generativa. A continuación, el sistema ajusta los valores actuales de los parámetros de red neuronal generativa basándose en la función objetivo de pérdida adversaria, donde la función objetivo de pérdida adversaria depende de la salida discriminadora generada por la red neuronal discriminadora.

La imagen de entrenamiento es una representación recortada de una imagen objetivo, y la red neuronal discriminadora está condicionada a una representación de características semánticas de la imagen objetivo. El sistema puede determinar que la representación semántica de la imagen objetivo es una salida intermedia generada por una red neuronal de clasificación mediante el procesamiento de la imagen objetivo. Condicionar la red neuronal discriminadora a la representación de características semánticas de la imagen objetivo puede incluir, por ejemplo, determinar la salida discriminadora basándose en una medida de similitud entre una salida intermedia de la red neuronal discriminadora y la representación de características semánticas de la imagen objetivo. La salida discriminadora puede determinarse de manera adicional basándose en la salida de la capa final de la red neuronal discriminadora.

Además de entrenar la red neuronal generativa mediante la función objetivo de pérdida adversaria, el sistema puede entrenar adicionalmente la red neuronal generativa usando una función objetivo de pérdida de reconstrucción. Más concretamente, el sistema puede ajustar los valores actuales de los parámetros de red neuronal generativa basándose en una función objetivo de pérdida de reconstrucción que caracterice una similitud de la imagen ampliada de entrenamiento con la imagen objetivo.

El sistema proporciona la imagen ampliada en respuesta a una solicitud (806). En algunos casos, la solicitud especifica elementos adicionales que se han de presentar con la imagen en el bloque, y el sistema superpone los elementos adicionales a la parte ampliada de la imagen ampliada. Después de que el sistema proporcione la imagen ampliada en respuesta a la solicitud, la imagen ampliada puede presentarse, por ejemplo, junto a resultados de búsqueda o en una página web de terceros.

La FIGURA 9 es un diagrama de bloques de un sistema informático 900 a modo de ejemplo que puede utilizarse para realizar las operaciones descritas anteriormente. El sistema 900 incluye un procesador 910, una memoria 920, un dispositivo de almacenamiento 930, y un dispositivo de entrada/salida 940. Cada uno de los componentes 910, 920, 930 y 940 puede interconectarse, por ejemplo, mediante un bus de sistema 950. El procesador 910 es capaz de procesar instrucciones para su ejecución dentro del sistema 900. En una implementación, el procesador 910 es un procesador de un solo hilo. En otra implementación, el procesador 910 es un procesador multihilo. El procesador 910 puede procesar instrucciones almacenadas en la memoria 920 o en el dispositivo de almacenamiento 930.

La memoria 920 almacena información dentro del sistema 900. En una implementación, la memoria 920 es un medio legible por ordenador. En una implementación, la memoria 920 es una unidad de memoria volátil. En otra implementación, la memoria 920 es una unidad de memoria no volátil.

El dispositivo de almacenamiento 930 puede proporcionar almacenamiento masivo para el sistema 900. En una implementación, el dispositivo de almacenamiento 930 es un medio legible por ordenador. En diversas implementaciones

diferentes, el dispositivo de almacenamiento 930 puede incluir, por ejemplo, un dispositivo de disco duro, un dispositivo de disco óptico, un dispositivo de almacenamiento que se comparte a través de una red por múltiples dispositivos informáticos (por ejemplo, un dispositivo de almacenamiento en la nube), o algún otro dispositivo de almacenamiento de gran capacidad.

El dispositivo de entrada/salida 940 proporciona operaciones de entrada/salida para el sistema 900. En una implementación, el dispositivo de entrada/salida 940 puede incluir uno o más dispositivos de interfaz de red, por ejemplo, una tarjeta Ethernet, un dispositivo de comunicación en serie, por ejemplo, un puerto RS-232, y/o un dispositivo de interfaz inalámbrico, por ejemplo, una tarjeta 802.11. En otra implementación, el dispositivo de entrada/salida puede incluir dispositivos controladores configurados para recibir datos de entrada y enviar datos de salida a otros dispositivos de entrada/salida, por ejemplo, teclado, impresora y dispositivos de visualización 960. Sin embargo, también se pueden utilizar otras implementaciones, tales como dispositivos informáticos móviles, dispositivos de comunicación móviles, dispositivos cliente de televisión con decodificador, etc.

Aunque se ha descrito un ejemplo de sistema de procesamiento en la FIGURA 9, las implementaciones de la materia y las operaciones funcionales descritas en la presente memoria descriptiva se pueden implementar en otros tipos de circuitos electrónicos digitales, o en software, firmware, o hardware de ordenador, que incluyen las estructuras divulgadas en la presente memoria descriptiva y sus equivalentes estructurales, o en combinaciones de uno o más de ellas.

Esta memoria descriptiva utiliza el término «configurado» en relación con sistemas y componentes de programas informáticos. Que un sistema de uno o más ordenadores esté configurado para realizar determinadas operaciones o acciones significa que el sistema tiene instalado software, firmware, hardware o una combinación de ellos que, en funcionamiento, hacen que el sistema realice las operaciones o acciones. Que uno o más programas informáticos estén configurados para realizar determinadas operaciones o acciones significa que el uno o más programas incluyen instrucciones que, cuando son ejecutadas por aparatos de procesamiento de datos, hacen que el aparato realice las operaciones o acciones.

Las formas de realización del objeto de la invención y las operaciones funcionales descritas en la presente memoria descriptiva se pueden implementar en circuitos electrónicos digitales, en software o firmware informático materializado de manera tangible, en hardware informático, que incluyen las estructuras divulgadas en la presente memoria descriptiva y sus equivalentes estructurales, o en combinaciones de uno o más de ellos. Las formas de realización del objeto de la invención descrito en la presente memoria descriptiva se pueden implementar como uno o más programas informáticos, es decir, uno o más módulos de instrucciones de programa informático, codificados en un medio de almacenamiento no transitorio tangible para su ejecución por, o para controlar el funcionamiento de, un aparato de procesamiento de datos. Un medio de almacenamiento informático puede ser un dispositivo de almacenamiento legible por máquina, un sustrato de almacenamiento legible por máquina, un dispositivo de memoria de acceso aleatorio o en serie, o una combinación de uno o más de ellos. De manera alternativa o adicional, las instrucciones del programa se pueden codificar en una señal propagada generada artificialmente, por ejemplo, una señal eléctrica, óptica o electromagnética generada por una máquina, que se genera para codificar información para su transmisión a un aparato receptor adecuado para su ejecución mediante un aparato de procesamiento de datos.

El término «aparato de procesamiento de datos» se refiere a hardware para el procesamiento de datos y abarca todo tipo de aparatos, dispositivos y máquinas para el procesamiento de datos, que incluyen a modo de ejemplo un procesador programable, un ordenador o múltiples procesadores u ordenadores. El aparato puede ser también, o incluir además, circuitos lógicos de propósito especial, por ejemplo, un FPGA (matriz de puertas programable in situ) o un ASIC (circuito integrado específico de una sola aplicación). De manera opcional, el aparato también puede incluir, además del hardware, un código que cree un entorno de ejecución para programas informáticos, por ejemplo, código que constituya el firmware del procesador, una pila de protocolos, un sistema de gestión de base de datos, un sistema operativo, o una combinación de uno o más de ellos.

Un programa informático, que también puede denominarse o describirse como un programa, software, una aplicación de software, una app, un módulo, un módulo de software, un *script* o código, puede escribirse en cualquier forma de lenguaje de programación, incluidos los lenguajes compilados o interpretados, o lenguajes declarativos o de procedimiento, y puede implementarse en cualquier forma, incluido como un programa independiente o como un módulo, componente, subrutina u otra unidad adecuada para su uso en un entorno informático. Un programa informático puede, pero no tiene que, corresponder a un archivo en un sistema de archivos. Un programa puede estar almacenado en una parte de un archivo que soporte otros programas o datos, por ejemplo, uno o más *scripts* almacenados en un documento de lenguaje de marcado, en un solo archivo destinado al programa en cuestión, o en múltiples archivos coordinados, por ejemplo, archivos que almacenen uno o más módulos, subprogramas, o partes de código. Un programa informático puede implementarse para ser ejecutado en un ordenador o en múltiples ordenadores que estén ubicados en un sitio o distribuidos a través de múltiples sitios e interconectados por una red de comunicación.

En esta memoria descriptiva, el término «motor» se utiliza en sentido amplio para referirse a un sistema, subsistema o proceso basado en software que esté programado para realizar una o más funciones específicas. Por lo general, un motor se implementará como uno o más módulos o componentes de software, instalados en uno o más ordenadores en una o

más ubicaciones. En algunos casos, uno o más ordenadores estarán destinados a un motor en particular; en otros casos, varios motores pueden estar instalados y funcionando en el mismo ordenador u ordenadores.

Los procesos y flujos lógicos descritos en la presente memoria descriptiva se pueden realizar mediante uno o más ordenadores programables que ejecuten uno o más programas informáticos para realizar funciones operando sobre los datos de entrada y generando salida. Los procesos y flujos lógicos también pueden ser realizados por circuitos lógicos de propósito especial, por ejemplo, un FPGA o un ASIC, o por una combinación de circuitos lógicos de propósito especial y uno o más ordenadores programados.

Los ordenadores adecuados para la ejecución de un programa informático pueden basarse en microprocesadores de propósito general o especial o ambos, o cualquier otro tipo de unidad central de procesamiento. Por lo general, una unidad central de procesamiento recibirá instrucciones y datos de una memoria de solo lectura o de una memoria de acceso aleatorio o de ambas. Los elementos esenciales de un ordenador son una unidad central de procesamiento para realizar o ejecutar instrucciones y uno o más dispositivos de memoria para almacenar instrucciones y datos. La unidad central de procesamiento y la memoria pueden complementarse mediante, o incorporarse en, circuitos lógicos de propósito especial. Generalmente, un ordenador también incluirá, o se acoplará operativamente para recibir datos desde o transferir datos a, o ambos, uno o más dispositivos de almacenamiento masivo para almacenar datos, por ejemplo, discos magnéticos, magneto-ópticos, o discos ópticos. Sin embargo, no es necesario que un ordenador tenga tales dispositivos. Además, un ordenador puede ser incorporado en otro dispositivo, por ejemplo, un teléfono móvil, un asistente digital personal (PDA), un reproductor de audio o vídeo móvil, una consola de juegos, un receptor del Sistema de Posicionamiento Global (GPS), o un dispositivo de almacenamiento portátil, por ejemplo, una unidad flash de bus serie universal (USB), por nombrar solo algunos.

Los medios legibles por ordenador adecuados para almacenar instrucciones y datos de programas informáticos incluyen todas las formas de memoria no volátil, medios y dispositivos de memoria, que incluyen, a modo de ejemplo, dispositivos de memoria de semiconductores, por ejemplo, EPROM, EEPROM, y dispositivos de memoria flash; discos magnéticos, por ejemplo, discos duros internos o discos extraíbles; discos magneto-ópticos; y discos CD ROM y DVD-ROM.

Para proporcionar interacción con un usuario, las formas de realización del objeto descrito en la presente memoria descriptiva se pueden implementar en un ordenador que tenga un dispositivo de visualización, por ejemplo, un monitor CRT (tubo de rayos catódicos) o LCD (pantalla de cristal líquido), para visualizar información al usuario y un teclado y un dispositivo señalador, por ejemplo, un ratón o una rueda de desplazamiento, por medio del cual el usuario pueda proporcionar entrada al ordenador. También se pueden utilizar otros tipos de dispositivos para proporcionar la interacción con un usuario; por ejemplo, la retroalimentación proporcionada 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, que incluye entrada acústica, de voz, o táctil. Además, un ordenador puede interactuar con un usuario mediante el envío de documentos a y la recepción de documentos de un dispositivo que utilice el usuario; por ejemplo, enviando páginas web a un navegador web en un dispositivo del usuario en respuesta a las solicitudes recibidas del navegador web. Asimismo, un ordenador puede interactuar con un usuario enviando mensajes de texto u otras formas de mensaje a un dispositivo personal, por ejemplo, un *smartphone* que esté ejecutando una aplicación de mensajería, y recibiendo a cambio mensajes de respuesta del usuario.

Los aparatos de procesamiento de datos para implementar modelos de aprendizaje automático también pueden incluir, por ejemplo, unidades aceleradoras de hardware de propósito especial para procesar partes comunes y de computación intensiva de las cargas de trabajo de entrenamiento o producción de aprendizaje automático, es decir, de inferencia.

Los modelos de aprendizaje automático pueden implementarse y utilizarse mediante un marco de aprendizaje automático, por ejemplo, un marco TensorFlow, un marco Microsoft Cognitive Toolkit, un marco Apache Singa o un marco Apache MXNet.

Las formas de realización del objeto descrito en la presente memoria descriptiva se pueden implementar en un sistema informático que incluya un componente de *back-end*, por ejemplo, como un servidor de datos, o que incluya un componente de *middleware*, por ejemplo, un servidor de aplicación, o que incluya un componente de *front-end*, por ejemplo, un ordenador cliente que tenga una interfaz gráfica de usuario, un navegador web, o una app a través de la cual un usuario pueda interactuar con una implementación del objeto descrita en la presente memoria descriptiva, o cualquier combinación de uno o más de tales componentes de *back-end*, *middleware* o *front-end*. Los componentes del sistema pueden estar interconectados mediante cualquier forma o medio de comunicación de datos digitales, por ejemplo, una red de comunicación. Los **ejemplos** de redes de comunicación incluyen una red de área local (LAN) y una red de área amplia (WAN), por ejemplo, Internet.

El sistema informático puede incluir clientes y servidores. Un cliente y un servidor están en general alejados el uno del otro e interactúan típicamente a través de una red de comunicación. La relación de cliente y servidor surge en virtud de que los programas informáticos se ejecutan en los respectivos ordenadores y tienen una relación cliente-servidor entre sí. En algunas formas de realización, un servidor transmite datos, por ejemplo, una página HTML, a un dispositivo de usuario, por ejemplo, con el fin de visualizar datos a y recibir entradas de usuario de un usuario que interactúa con el dispositivo,

que actúa de cliente. Los datos generados en el dispositivo de usuario, por ejemplo, un resultado de la interacción del usuario, pueden ser recibidos en el servidor desde el dispositivo.

- 5 Aunque esta memoria descriptiva contiene numerosos detalles de implementación específicos, estos no se deben interpretar como limitaciones en el alcance de ninguna invención ni en el alcance de lo que se puede reivindicar, sino más bien como descripciones de características que pueden ser específicas de formas de realización concretas de invenciones concretas. Ciertas características que se describen en esta memoria descriptiva en el contexto de formas de realización separadas también se pueden implementar en combinación en una sola forma de realización.
- 10 De manera similar, aunque las operaciones están representadas en los dibujos en un orden concreto, no se debe entender que por ello se requiera que tales operaciones se lleven a cabo en el orden concreto que se muestra ni en orden secuencial, ni que todas las operaciones ilustradas se lleven a cabo, para lograr resultados deseables. En determinadas circunstancias, la multitarea y el procesamiento en paralelo pueden resultar ventajosos. Además, no ha de entenderse
- 15 que la separación de varios módulos y componentes del sistema en las formas de realización descritas anteriormente requiere que dicha separación se dé en todas las formas de realización, y ha de entenderse que los componentes de programa y los sistemas descritos se pueden integrar por lo general juntos en un único producto de software o empaquetar en múltiples productos de software.
- 20 Se han descrito formas de realización concretas del objeto de la invención. Otras formas de realización quedan dentro del ámbito de las siguientes reivindicaciones. A modo de ejemplo, los procesos que se representan en las figuras adjuntas no requieren necesariamente el orden particular mostrado ni un orden secuencial para lograr resultados deseables. En algunos casos, la multitarea y el procesamiento en paralelo pueden resultar ventajosos.

REIVINDICACIONES

1. Método realizado por uno o más aparatos de procesamiento de datos, comprendiendo el método:

proporcionar una entrada que comprenda una imagen proporcionada a una red neuronal generativa con una pluralidad de parámetros de red neuronal generativa, donde:

la red neuronal generativa procesa la entrada de conformidad con valores entrenados de la pluralidad de parámetros de red neuronal generativa para generar una imagen ampliada;

la imagen ampliada tiene (i) más filas, más columnas o ambas que la imagen proporcionada, y (ii) se predice que es una ampliación de la imagen proporcionada;

la red neuronal generativa se entrena de manera conjunta con una red neuronal discriminadora con una pluralidad de parámetros de red neuronal discriminadora que está configurada para procesar una imagen dada para generar una salida discriminadora que caracteriza una probabilidad de que la imagen dada se haya generado utilizando la red neuronal generativa;

la red neuronal generativa se ha entrenado utilizando una función objetivo de pérdida adversaria que depende de la salida discriminadora;

entrenar la red neuronal generativa utilizando la función objetivo de pérdida adversaria comprende procesar una entrada de entrenamiento que comprende una imagen de entrenamiento utilizando la red neuronal generativa; y la red neuronal discriminadora está condicionada a una representación de características semánticas de una imagen objetivo, siendo la imagen de entrenamiento una representación recortada de la imagen objetivo, donde la red neuronal discriminadora está condicionada a la representación de características semánticas de la imagen objetivo por:

proporcionar la representación de características semánticas de la imagen objetivo como entrada adicional a la red neuronal discriminadora; o

determinar la salida discriminadora basándose al menos en parte en una medida de similitud entre la representación semántica de una imagen ampliada objetivo y una salida intermedia de la red neuronal discriminadora, y

donde la representación semántica de la imagen ampliada objetivo se normaliza mediante un motor de normalización antes de utilizarla para condicionar la red neuronal discriminadora.

2. Método según la reivindicación 1, que comprende además:

recibir una solicitud de presentación de una imagen en un bloque; determinar que la imagen proporcionada responde a la solicitud; y, tras generar la imagen ampliada, proporcionar la imagen ampliada en respuesta a la solicitud.

3. Método según la reivindicación 2, que comprende además:

determinar que la imagen proporcionada es apta para la ampliación de imágenes basándose en un tamaño del bloque que difiere de un tamaño de la imagen proporcionada, y proporcionar en respuesta la entrada que comprende la imagen proporcionada a la red neuronal generativa.

4. Método según la reivindicación 3, donde el tamaño del bloque especifica una relación de aspecto del bloque.

5. Método según cualquiera de las reivindicaciones 2 a 4, donde:

la solicitud especifica elementos adicionales que han de presentarse con la imagen en el bloque; y proporcionar la imagen ampliada en respuesta a la solicitud comprende superponer los elementos adicionales sobre una porción ampliada de la imagen ampliada.

6. Método según cualquiera de las reivindicaciones 2 a 5, donde el bloque ha de presentarse junto a resultados de búsqueda o en una página web de terceros.

7. Método según cualquier reivindicación anterior, donde la red neuronal generativa comprende una pluralidad de capas de red neuronal convolucional, y opcionalmente una pluralidad de conexiones de salto, y opcionalmente una pluralidad de capas de normalización de instancias.

8. Método según cualquier reivindicación anterior, donde la entrada a la red neuronal generativa comprende:

una imagen inicial con un mismo número de filas y columnas que la imagen ampliada, donde la imagen inicial comprende: (i) una primera porción que corresponde a la imagen proporcionada, y (ii) una segunda porción con valores de píxel predeterminados; y

una imagen enmascarada con el mismo número de filas y columnas que la imagen ampliada, donde la imagen enmascarada identifica la primera porción y la segunda porción de la imagen inicial.

9. Método según la reivindicación 8, donde: (i) los píxeles de la imagen enmascarada correspondiente a la primera porción de la imagen inicial tienen cada uno un primer valor de píxel, y (ii) los píxeles de la imagen enmascarada correspondiente a la segunda porción de la imagen inicial tienen cada uno un segundo valor de píxel que es diferente del primer valor de píxel.

10. Método según la reivindicación 1, donde entrenar la red neuronal generativa utilizando la función objetivo de pérdida adversaria comprende además:

procesar la entrada de entrenamiento de conformidad con los valores actuales de los parámetros de red neuronal generativa para generar una imagen ampliada de entrenamiento que amplíe la imagen de entrenamiento;
generar una entrada de red neuronal discriminadora basándose en la imagen ampliada de entrenamiento;
procesar la entrada de red neuronal discriminadora utilizando la red neuronal discriminadora y, de conformidad con los valores actuales de los parámetros de red neuronal discriminadora, generar una salida discriminadora que caracterice una probabilidad de que la entrada de red neuronal discriminadora se haya generado utilizando la red neuronal generativa; y
ajustar los valores actuales de los parámetros de red neuronal generativa basándose en la función objetivo de pérdida adversaria, donde la función objetivo de pérdida adversaria depende de la salida discriminadora que caracteriza la probabilidad de que la entrada de red neuronal discriminadora se haya generado utilizando la red neuronal generativa.

11. Método según la reivindicación 10, donde generar una entrada de red neuronal discriminadora basándose en la imagen ampliada de entrenamiento comprende:
sobrescribir una porción de la imagen ampliada de entrenamiento correspondiente a la imagen de entrenamiento con la imagen de entrenamiento.

12. Método según cualquiera de las reivindicaciones 10 a 11, que comprende además
ajustar los valores actuales de los parámetros de red neuronal generativa basándose en una función objetivo de pérdida de reconstrucción que caracteriza una similitud de la imagen ampliada de entrenamiento con la imagen objetivo, donde la imagen de entrenamiento es una representación recortada de la imagen objetivo.

13. Sistema que comprende uno o más ordenadores y uno o más dispositivos de almacenamiento que almacenan instrucciones que, cuando son ejecutadas por el uno o más ordenadores, hacen que el o uno o más ordenadores realicen las operaciones del método respectivo según cualquiera de las reivindicaciones 1 a 12.

14. Uno o más medios de almacenamiento informático que almacenan instrucciones que, cuando son ejecutadas por uno o más ordenadores, hacen que el uno o más ordenadores realicen las operaciones del método respectivo según cualquiera de las reivindicaciones 1 a 12.

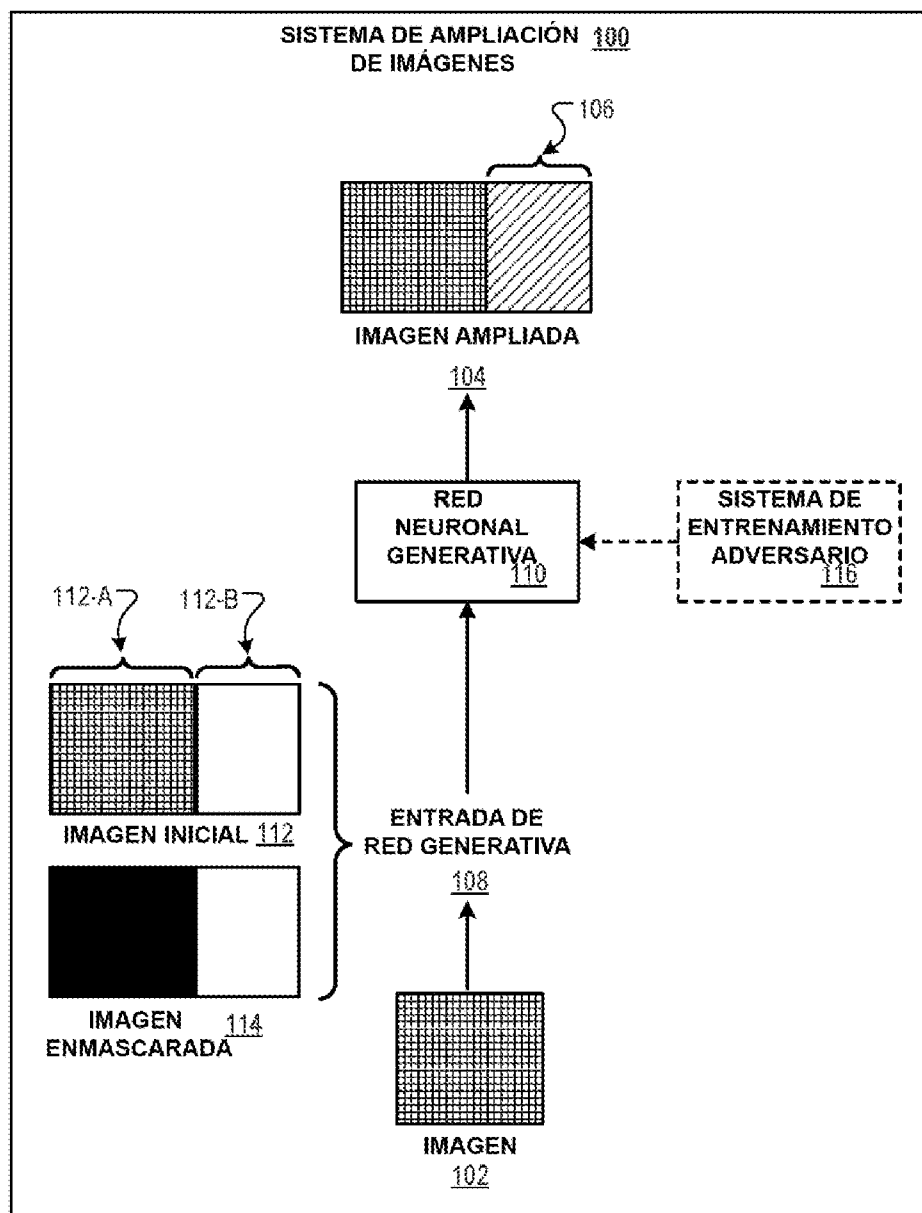


FIG. 1

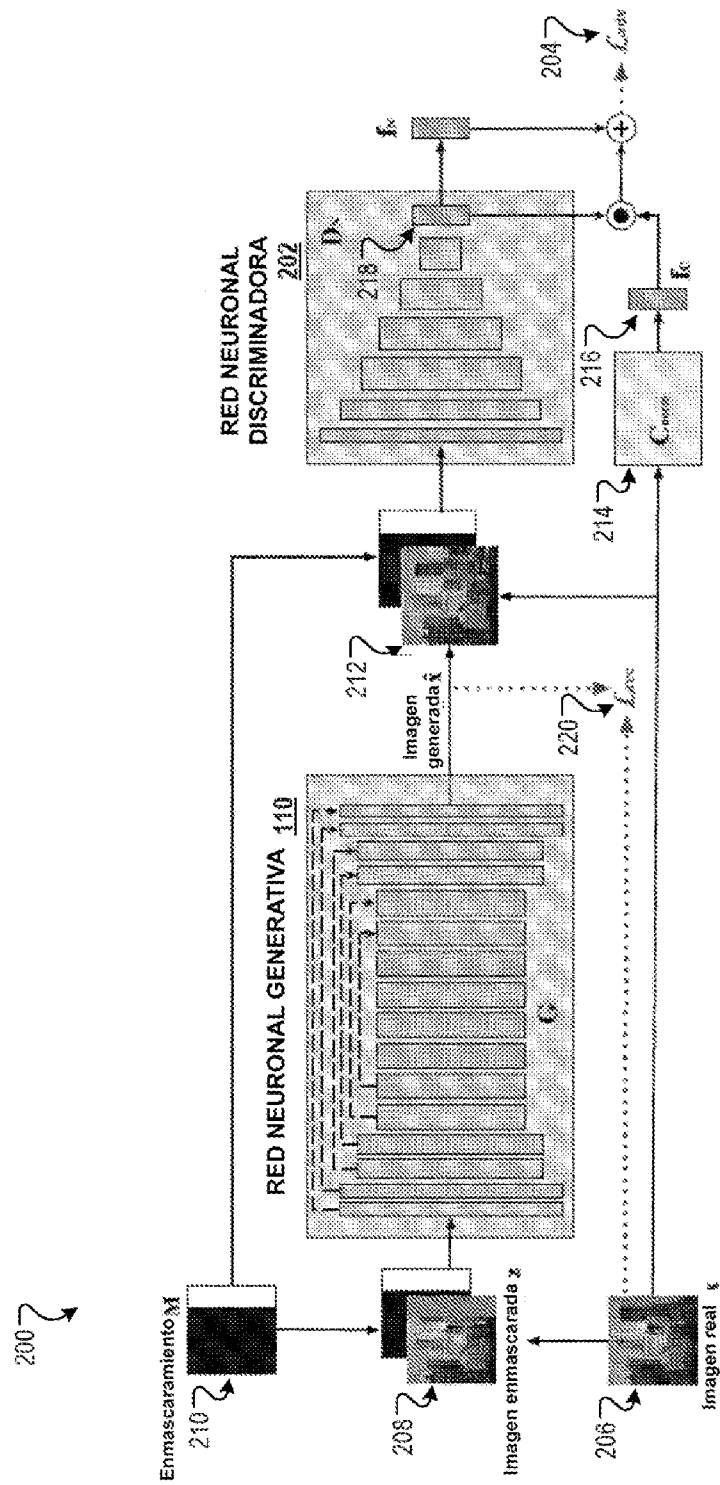


FIG. 2

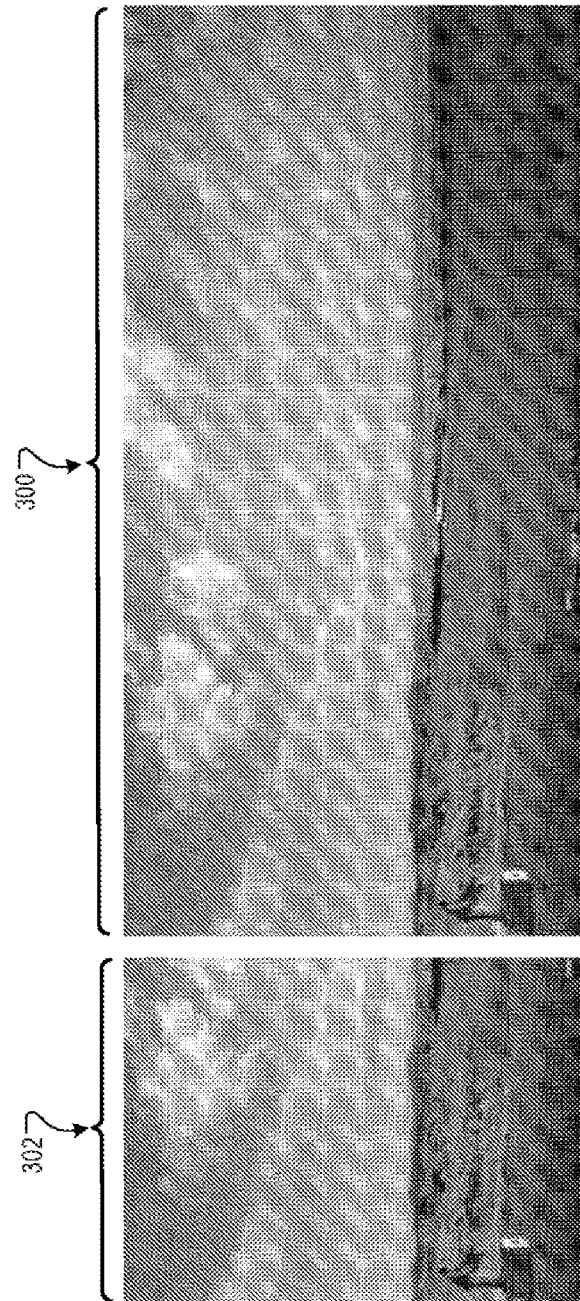


FIG. 3

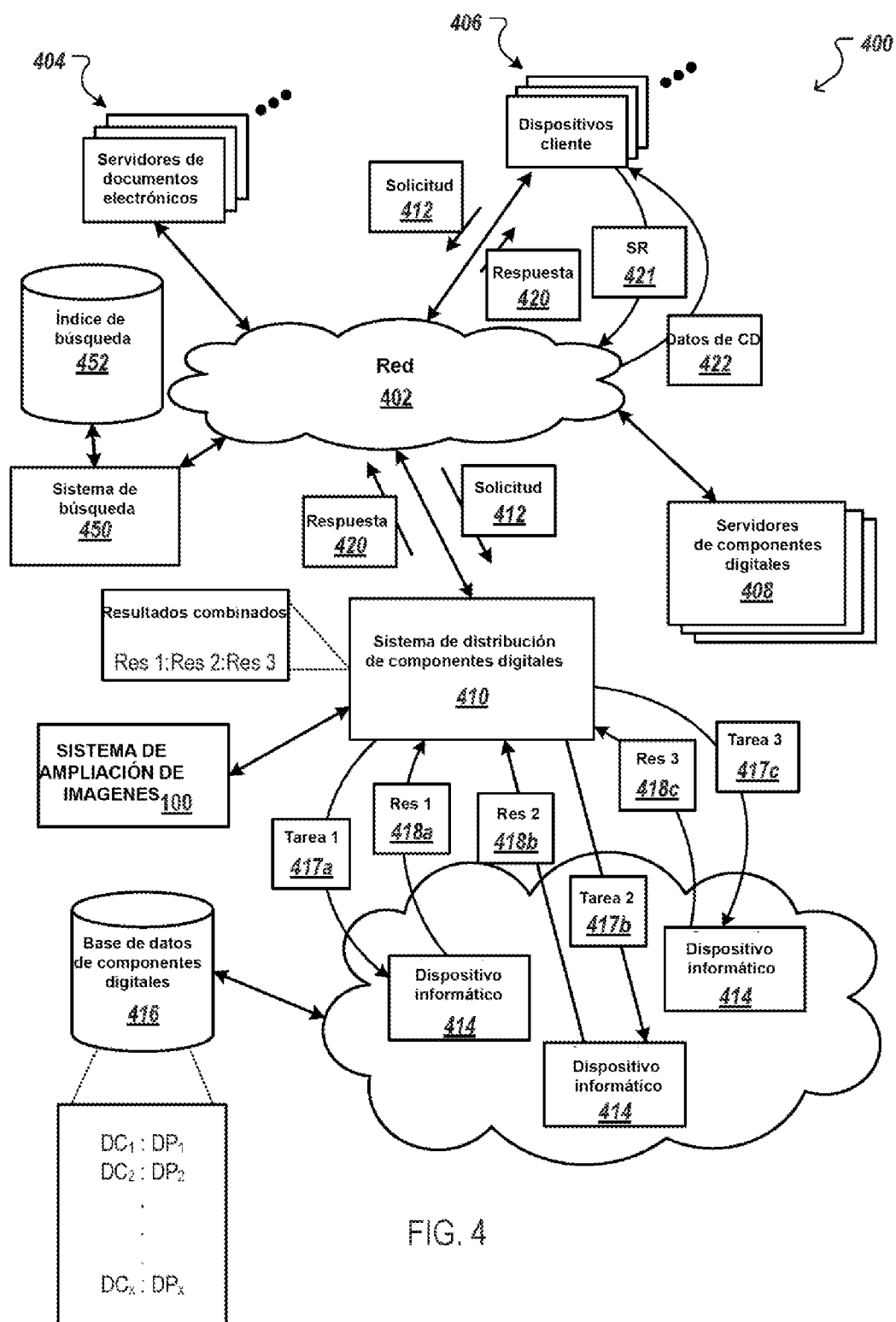


FIG. 4

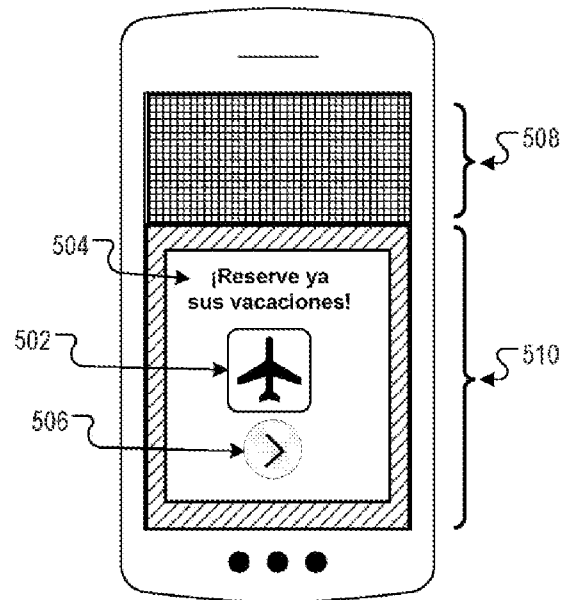


FIG. 5



FIG. 6

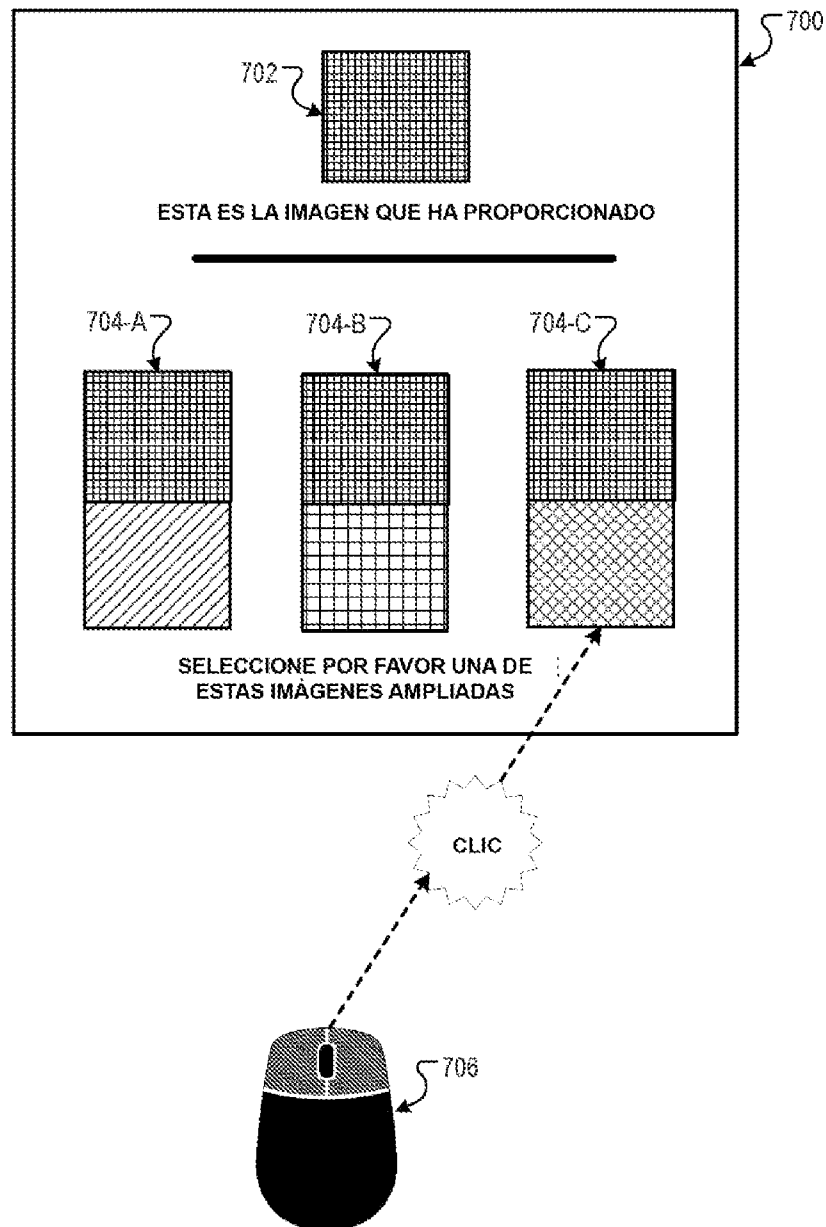


FIG. 7

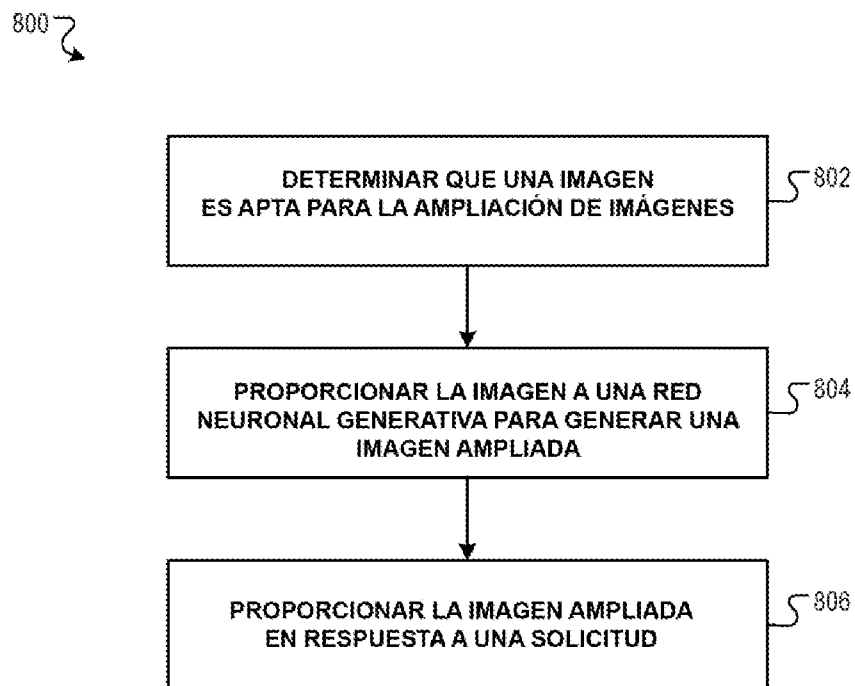


FIG. 8

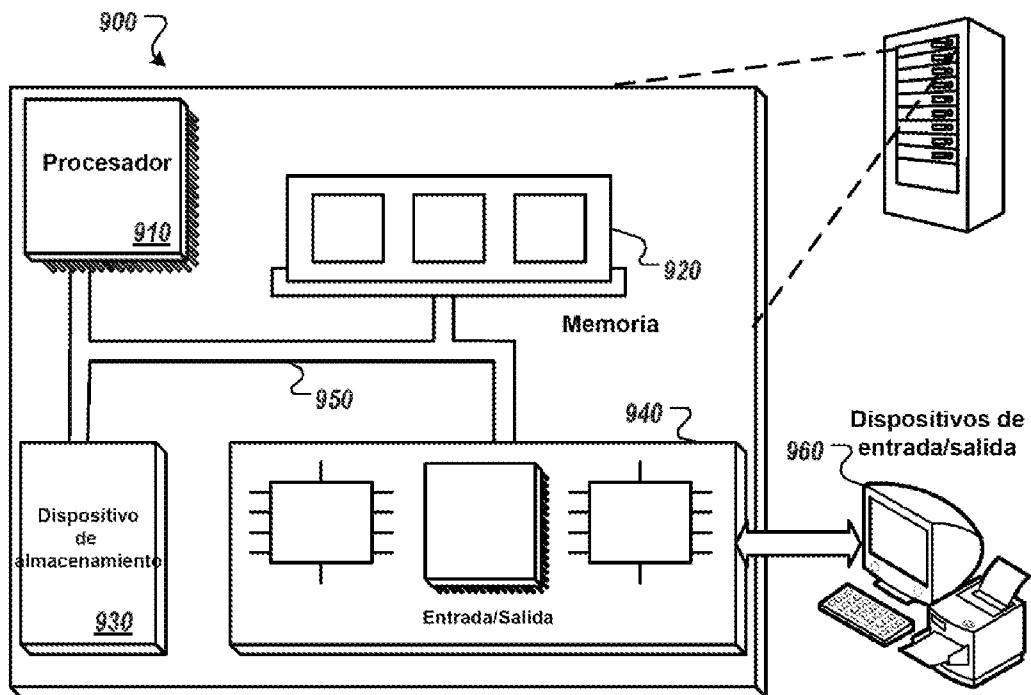


FIG. 9