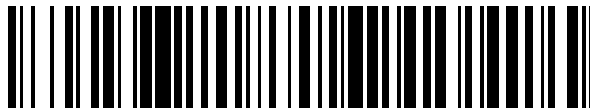


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 807 263**

51 Int. Cl.:

**G06K 9/46**

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.07.2018 PCT/EP2018/067850**

87 Fecha y número de publicación internacional: **03.01.2019 WO19002631**

96 Fecha de presentación y número de la solicitud europea: **02.07.2018 E 18733658 (1)**

97 Fecha y número de publicación de la concesión europea: **29.04.2020 EP 3449421**

54 Título: **Clasificación y modelado 3D de estructuras dentomaxilofaciales 3D usando métodos de aprendizaje profundo**

30 Prioridad:

**30.06.2017 EP 17179185**

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

**22.02.2021**

73 Titular/es:

**PROMATON HOLDING B.V. (100.0%)  
Overschiestraat 182 unit 5.3  
1062 XK Amsterdam, NL**

72 Inventor/es:

**CLAESSEN, FRANK THEODORUS CATHARINA;  
VERHEIJ, BAS ALEXANDER y  
ANSSARI MOIN, DAVID**

74 Agente/Representante:

**ARIAS SANZ, Juan**

**ES 2 807 263 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

Clasificación y modelado 3D de estructuras dentomaxilofaciales 3D usando métodos de aprendizaje profundo

5 **Campo de la invención**

La invención se refiere a clasificación y modelado 3D de estructuras dentomaxilofaciales 3D usando redes neuronales de aprendizaje profundo, y, en particular, aunque no exclusivamente, a sistemas y métodos de clasificación y modelado 3D de estructuras dentomaxilofaciales 3D usando redes neuronales de aprendizaje profundo, un método de entrenamiento de tales redes neuronales de aprendizaje profundo, un método de procesamiento previo de datos de imágenes dentomaxilofaciales 3D y un método de procesamiento posterior de datos de vóxeles clasificados de estructuras dentomaxilofaciales y un producto de programa informático para usar tal método.

15 **Antecedentes de la invención**

En el análisis de imágenes de estructuras dentomaxilofaciales, es fundamental la visualización y reconstrucción de imágenes 3D de partes o tejidos específicos para permitir un diagnóstico y tratamientos precisos. Antes de la reconstrucción de imágenes 3D, se aplica un proceso de clasificación y segmentación a los datos de imágenes 3D, por ejemplo vóxeles, para formar un modelo 3D de diferentes partes (por ejemplo, dientes y mandíbula) de la estructura dentomaxilofacial como se representa en una pila de datos de imágenes 3D. La tarea de segmentación puede definirse como la identificación del conjunto de píxeles o vóxeles que componen el contorno o el interior de un objeto de interés. Sin embargo, el proceso de segmentación de estructuras dentomaxilofaciales tales como dientes, mandíbula y nervio alveolar inferior de exploraciones de TC 3D es desafiante. Los métodos de segmentación manual consumen mucho tiempo e incluyen una aproximación general mediante la selección manual de umbrales y correcciones manuales. Los resultados de segmentaciones manuales tienen baja reproducibilidad y dependen de la interpretación humana de exploraciones de TC.

Se han usado diferentes metodologías de obtención de imágenes para generar modelos 3D de dientes y mandíbula sobre la base de datos de imágenes de exploraciones de TC. Inicialmente, se usó la aplicación secuencial de procesamiento de píxeles de bajo nivel y modelado matemático para segmentar estructuras dentomaxilofaciales. Un ejemplo se describe en el artículo de Pavaloiu *et al.*, "Automatic segmentation for 3D dental reconstruction", IEEE 6<sup>o</sup> ICCNT, 13-15 de julio de 2015. Estas técnicas incluyen métodos activos de seguimiento de contornos, transformación divisora, crecimiento de regiones y modelado de conjuntos de niveles con forma e intensidad previa. Actualmente, en la obtención de imágenes médicas se usan técnicas más avanzadas, como técnicas de aprendizaje profundo, para segmentar objetos de interés en imágenes médicas.

Estas redes neuronales están entrenadas para aprender las características que representan de manera óptima los datos. Tales algoritmos de aprendizaje profundo incluyen una red neuronal profunda multicapa que transforma datos de entrada (por ejemplo, imágenes) en salidas (por ejemplo, presencia/ausencia de enfermedad) mientras aprenden características de nivel cada vez más alto. Un modelo de red neuronal exitoso para el análisis de imágenes es la denominada red neuronal convolucional (CNN). Las CNN contienen muchas capas que transforman su entrada usando núcleos, también conocidos como filtros de convolución, que consisten en una matriz de tamaño relativamente pequeño. Una visión general del uso de las CNN para la obtención de imágenes médicas puede encontrarse en el artículo de Litjens *et al.*, *A Survey on Deep Learning in Medical Image Analysis*, publicado el 21 de febrero de 2017 arXiv (presentado en *Computer Vision and Pattern Recognition*). Sin embargo, el modelado 3D de estructuras dentomaxilofaciales usando CNN 3D, es difícil debido a la complejidad de las estructuras dentomaxilofaciales. Pavaloiu *et al.* descrito en su artículo "*Neural network based edge detection for CBCT segmentation*", 5<sup>o</sup> IEEE EHB, 19-21 de noviembre de 2015, el uso de una red neuronal muy simple en la detección de bordes en las imágenes de CBCT 2D. Sin embargo, hasta ahora no se ha informado sobre la segmentación 3D exacta automática de datos de imágenes de CBCT 3D basándose en el aprendizaje profundo.

Un problema en la clasificación 3D y el modelado 3D de estructuras dentomaxilofaciales es que las imágenes dentomaxilofaciales se generan usando tomografía computarizada de haz cónico (CBCT). La CBCT es una técnica de obtención de imágenes médicas usando tomografía computarizada de rayos X en la que la radiación de rayos X se forma en un haz divergente de baja dosis. La radiodensidad, medida en unidades Hounsfield (HU), no es fiable en las exploraciones de CBCT porque las diferentes áreas en la exploración aparecen con diferentes valores de escala de grises dependiendo de sus posiciones relativas en el órgano que está explorándose. Las HU medidas desde el mismo área anatómica tanto con escáneres de CBCT y TC de grado médico no son idénticas y no son, por tanto, fiables para la determinación de la densidad ósea específica del sitio, identificada de manera radiográfica.

Además, los sistemas de CBCT para la exploración de estructuras dentomaxilofaciales no emplean un sistema estandarizado para escalar los niveles de grises que representan los valores de densidad reconstruida. Estos valores son como tales arbitrarios y no permiten evaluar la calidad ósea. En ausencia de tal estandarización, es difícil interpretar los niveles de grises o imposible comparar los valores resultantes de diferentes máquinas. Además, las raíces de los dientes y la estructura ósea de la mandíbula tienen densidades similares por lo que es difícil para un ordenador distinguir entre vóxeles que pertenecen a los dientes y vóxeles que pertenecen a una mandíbula.

Adicionalmente, los sistemas de CBCT, como, por ejemplo, los dados a conocer en el documento US2014/227655, son muy sensibles al denominado endurecimiento de haz, que produce rayas oscuras entre dos objetos de alta atenuación (como metal o hueso), con rayas brillantes circundantes. Los problemas mencionados anteriormente hacen que la segmentación automática de estructuras dentomaxilofaciales sea particularmente difícil.

Por tanto, existe una necesidad en la técnica de sistemas informáticos adaptados para segmentar de manera precisa datos de imágenes de TC 3D de estructuras dentomaxilofaciales en un modelo 3D. En particular, existe una necesidad en la técnica de sistemas informáticos que puedan segmentar de manera precisa datos de imágenes de TC 3D de estructuras dentomaxilofaciales provenientes de diferentes sistemas de CBCT en un modelo 3D.

## Sumario de la invención

Las características esenciales de la invención que permiten la segmentación precisa de los datos de imágenes de TC 3D de estructuras dentomaxilofaciales en un modelo 3D, se definen en las reivindicaciones independientes. Realizaciones particulares se definen en las reivindicaciones dependientes.

## Breve descripción de los dibujos

La figura 1 representa de manera esquemática un sistema informático para la clasificación y segmentación de estructuras dentomaxilofaciales 3D según una realización de la invención;

la figura 2 representa un diagrama de flujo de entrenamiento de una red neuronal profunda para clasificar datos de imágenes dentomaxilofaciales 3D según una realización de la invención;

las figuras 3A y 3B representan ejemplos de datos de imágenes de TC 3D y datos de exploración óptica 3D respectivamente;

las figuras 4A y 4B representan ejemplos de arquitecturas de redes neuronales profundas para clasificar datos de imágenes 3D dentomaxilofaciales;

las figuras 5A y 5B ilustran métodos de determinación de características posicionales 3D según diversas realizaciones de la invención;

la figura 6 proporciona una visualización que contiene los valores de vóxel sumados de una pila de imágenes 3D y una curva ajustada a vóxeles que representa un arco dentomaxilofacial;

las figuras 7A-7E representan ejemplos de características posicionales 3D según diversas realizaciones de la invención;

las figuras 8A-8D representan ejemplos de la salida de una red neuronal de aprendizaje profundo entrenada según una realización de la invención;

la figura 9 representa un diagrama de flujo de procesamiento posterior de vóxeles clasificados de estructuras dentomaxilofaciales 3D según una realización de la invención;

la figura 10 representa una arquitectura de red neuronal profunda para el procesamiento posterior de vóxeles clasificados de estructuras dentomaxilofaciales 3D según una realización de la invención;

las figuras 11A-11B representan un proceso de reconstrucción de superficie de vóxeles clasificados según una realización de la invención;

la figura 12 es un diagrama de bloques que ilustra un sistema informático de datos a modo de ejemplo que puede usarse para ejecutar métodos y productos de software descritos en esta divulgación.

## Descripción detallada

En esta divulgación se describen realizaciones de sistemas informáticos y métodos implementados por ordenador que usan redes neuronales profundas para clasificar, segmentar y modelar en 3D estructuras dentomaxilofaciales basándose en datos de imágenes 3D, por ejemplo, datos de imágenes 3D definidos por una secuencia de imágenes que forman una pila de datos de imágenes de TC, en particular una pila de datos de imágenes de TC de haz cónico (CBCT). Los datos de imágenes 3D pueden incluir vóxeles que forman un espacio de imagen 3D de una estructura dentomaxilofacial. Un sistema informático según la invención puede comprender al menos una red neuronal profunda que está entrenada para clasificar una pila de datos de imágenes 3D de una estructura dentomaxilofacial en vóxeles de diferentes clases, en el que cada clase puede asociarse con una parte distinta (por ejemplo, dientes, mandíbula, nervio) de la estructura. El sistema informático puede configurarse para ejecutar un proceso de entrenamiento que, iterativamente, entrena (optimiza) una o más redes neuronales profundas basándose en uno o más conjuntos de

entrenamiento que pueden incluir modelos 3D precisos de estructuras dentomaxilofaciales. Estos modelos 3D pueden incluir estructuras dentomaxilofaciales exploradas ópticamente (dientes y/o hueso de mandíbula).

5 Una vez entrenada, la red neuronal profunda puede recibir una pila de datos de imágenes 3D de una estructura dentomaxilofacial y clasificar los vóxeles de la pila de datos de imágenes 3D. Antes de que se presenten los datos a la red neuronal profunda entrenada, los datos pueden procesarse previamente de modo que la red neuronal pueda clasificar vóxeles de manera eficiente y precisa. La salida de la red neuronal puede incluir diferentes recopilaciones de datos de vóxel, en la que cada recopilación puede representar una parte distinta, por ejemplo, dientes o hueso de mandíbula de los datos de imágenes 3D. Los vóxeles clasificados pueden procesarse posteriormente para reconstruir un modelo 3D preciso de la estructura dentomaxilofacial.

15 El sistema informático que comprende una red neuronal entrenada para clasificar de manera automática vóxeles de estructuras dentomaxilofaciales, el entrenamiento de la red, el procesamiento previo de los datos de imágenes 3D antes de que se alimente a la red neuronal, así como el procesamiento posterior de vóxeles que se clasifican por la red neuronal se describen a continuación con más detalle.

20 La figura 1 representa de manera esquemática un sistema informático para la clasificación y segmentación de estructuras dentomaxilofaciales 3D según una realización de la invención. En particular, el sistema informático 102 puede configurarse para recibir una pila de datos de imágenes 3D 104 de una estructura dentomaxilofacial. La estructura puede incluir estructuras de mandíbula, dientes y nervio. Los datos de imágenes 3D pueden comprender vóxeles, es decir, elementos de espacio 3D asociados con un valor de vóxel, por ejemplo, un valor de escala de grises o un valor de color, que representa un valor de intensidad o densidad de radiación. Preferiblemente, la pila de datos de imágenes 3D puede incluir datos de imágenes de CBCT según un formato predeterminado, por ejemplo, el formato DICOM o un derivado del mismo.

25 El sistema informático puede comprender un preprocesador 106 para el procesamiento previo de los datos de imágenes 3D antes de alimentarse a la entrada de una primera red neuronal de aprendizaje profundo 3D 112, que está entrenado para producir un conjunto 3D de vóxeles clasificados como salida 114. Como se describirá a continuación con más detalle, la red neuronal de aprendizaje profundo 3D puede entrenarse según un esquema de entrenamiento predeterminado de modo que la red neuronal entrenada sea capaz de clasificar vóxeles manera precisa en la pila de datos de imágenes 3D en vóxeles de diferentes clases (por ejemplo, vóxeles asociados con dientes, hueso de mandíbula y/o tejido nervioso). La red neuronal de aprendizaje profundo 3D puede comprender una pluralidad de capas de red neuronal convolucional 3D conectadas (CNN 3D).

35 El sistema informático puede además comprender un postprocesador 116 para reconstruir de manera precisa modelos 3D de diferentes partes de la estructura dentomaxilofacial (por ejemplo, diente, mandíbula y nervio) usando los vóxeles clasificados por la red neuronal de aprendizaje profundo 3D. Como se describirá a continuación en mayor detalle, parte de los vóxeles clasificados, por ejemplo, vóxeles que se clasifican como que pertenece a una estructura de diente o una estructura de mandíbula se introducen a una segunda red neuronal de aprendizaje profundo 3D adicional 120, que está entrenada para reconstruir volúmenes 3D para las estructuras dentomaxilofaciales, por ejemplo, la forma de la mandíbula 124 y la forma de los dientes 126, basándose en los vóxeles que se clasificaron como pertenecientes a tales estructuras. Otras partes de los vóxeles clasificados, por ejemplo, vóxeles que se clasificaron por la red neuronal profunda 3D como pertenecientes a los nervios pueden procesarse posteriormente usando una función de interpolación 118 y almacenarse como datos de nervios 3D 122. La tarea de determinar el volumen que representa un nervio a partir de los vóxeles clasificados es de una naturaleza que actualmente está más allá de la capacidad de (la potencia de procesamiento disponible para) una red neuronal profunda. Además, los vóxeles clasificados presentados pueden no contener la información que sería adecuada para una red neuronal para resolver este problema en particular. Por tanto, con el fin de procesar posteriormente de manera precisa y eficiente los vóxeles de nervios clasificados se usa una interpolación de los vóxeles clasificados. Después del procesamiento posterior de los datos 3D de las diversas partes de la estructura dentomaxilofacial, los datos de nervio, mandíbula y diente 122-126 pueden combinarse y formatearse en modelos 3D separados 128 que representan de manera precisa las estructuras dentomaxilofaciales en los datos de imágenes 3D que se alimentaron a la entrada de la sistema informático.

55 En las exploraciones CBCT, la radiodensidad (medida en unidades Hounsfield (HU)) es inexacta debido a que las diferentes áreas de la exploración aparecen con diferentes valores de escala de grises dependiendo de sus posiciones relativas en el órgano que está explorándose. Las HU medidas desde el mismo área anatómica con ambos escáneres de CBCT y TC de grado médico no son idénticas y no son, por tanto, fiables para la determinación de la densidad ósea específica del sitio, identificada de manera radiográfica.

60 Además, los sistemas de CBCT dentales no emplean un sistema estandarizado para escalar los niveles de grises que representan los valores de densidad reconstruida. Estos valores son como tales arbitrarios y no permiten evaluar la calidad ósea. En ausencia de una estandarización de este tipo, es difícil interpretar los niveles de grises o imposible comparar los valores resultantes de diferentes máquinas.

65 Los dientes y la estructura ósea de mandíbula tienen una densidad similar de modo que es difícil para un ordenador distinguir entre vóxeles que pertenecen a dientes y vóxel que pertenece a una mandíbula. Adicionalmente, los sistemas

de CBCT son muy sensibles al llamado endurecimiento del haz, que produce rayas oscuras entre dos objetos de alta atenuación (como metal o hueso), con rayas brillantes circundantes.

Con el fin de hacer que la red neuronal de aprendizaje profundo 3D sea robusta frente a los problemas antes mencionados, la red neuronal 3D puede entrenarse usando un módulo 138 para hacer uso de modelos 3D de partes de la estructura dentomaxilofacial como se representa por los datos de imágenes 3D. Los datos de entrenamiento 3D 130 pueden alinearse de manera correcta con respecto a una imagen de CBCT presentada en 104 para la que se conoce la salida objetivo asociada (por ejemplo, datos de imágenes de TC 3D de una estructura dentomaxilofacial y una representación segmentada 3D asociada de la estructura dentomaxilofacial). Los datos de entrenamiento 3D convencionales pueden obtenerse segmentando de manera manual los datos de entrada, lo que puede representar una cantidad significativa de trabajo. Adicionalmente, la segmentación manual da como resultado una baja reproducibilidad y consistencia de los datos de entrada que van a usarse.

Con el fin de contrarrestar este problema, en una realización, se produjeron de manera óptica datos de entrenamiento 130, es decir, pueden usarse modelos 3D precisos de (partes de) estructura dentomaxilofacial en lugar de o al menos además de datos de entrenamiento segmentados de manera manual. Las estructuras dentomaxilofaciales que se usan para producir los datos de entrenamiento pueden explorarse usando un escáner óptico 3D. Tales escáneres ópticos 3D se conocen en la técnica y pueden usarse para producir datos de superficie de diente y mandíbula 3D de alta calidad. Los datos de superficie 3D pueden incluir mallas de superficie 3D 132 que pueden llenarse (determinar qué vóxeles específicos forman parte del volumen abarcado por la malla) y usarse por un clasificador de vóxel 134. De esta manera, el clasificador de vóxel es capaz de generar vóxeles clasificados de alta calidad para el entrenamiento 136. Adicionalmente, como se mencionó anteriormente, pueden usarse vóxeles de entrenamiento clasificados manualmente por el módulo de entrenamiento para entrenar a la red también. El módulo de entrenamiento puede usar los vóxeles de entrenamiento clasificados como objetivo y los datos de entrenamiento de TC asociados como entrada.

Adicionalmente, durante el proceso de entrenamiento, los datos de entrenamiento de TC pueden procesarse previamente por un extractor de características 108, que puede configurarse para determinar características posicionales 3D. Una característica dentomaxilofacial puede codificar al menos información espacial asociada con una o más partes de la estructura dentomaxilofacial de la que se obtienen imágenes (el conjunto de datos 3D recibido). Por ejemplo, en una realización, una característica posicional 3D diseñada manualmente puede incluir una curva 3D que representa (parte de) el hueso de mandíbula, en particular el arco dental, en el volumen 3D que contiene los vóxeles. Pueden asignarse uno o más parámetros de peso a los puntos a lo largo de la curva 3D. El valor de un valor de peso puede usarse para codificar una traslación en el espacio 3D de vóxel a vóxel. En lugar de incorporar, por ejemplo, una versión codificada del espacio original en el que se recibe la pila de imágenes, el espacio codificado es específico de las estructuras dentomaxilofaciales tal como se detectan en la entrada. El extractor de características puede determinar una o más curvas que se aproximan a una de más curvas de la mandíbula y/o dientes (por ejemplo, el arco dental) examinando los valores de vóxel que representan los valores de intensidad o densidad de radiación y ajustando una o más curvas (por ejemplo, un polinomio) a través de determinados vóxeles. Pueden almacenarse derivados de (partes de) curvas de arco dental de una pila de datos de imágenes de TC 3D como un mapeo de características posicionales 110.

En otra realización, tales características posicionales 3D pueden, por ejemplo, determinarse mediante un método de aprendizaje automático (entrenado), como una red neuronal profunda 3D diseñada para derivar información relevante a partir de todo el conjunto de datos 3D recibido.

La figura 2 representa un diagrama de flujo de entrenamiento de una red neuronal profunda para clasificar datos de imágenes dentomaxilofaciales 3D según una realización de la invención. Los datos de entrenamiento se usan para entrenar una red neuronal de aprendizaje profundo 3D para que sea capaz de clasificar vóxeles de manera automática de una exploración de TC 3D de una estructura dentomaxilofacial. Como se muestra en esta figura, puede proporcionarse una representación de un complejo dentomaxilofacial 202 al sistema informático. Los datos de entrenamiento pueden incluir una pila de datos de imágenes de TC 204 de una estructura dentomaxilofacial y un modelo 3D asociado, por ejemplo, datos 3D 206 a partir de exploración óptica de la misma estructura dentomaxilofacial. En las figuras 3A y 3B se muestran ejemplos de tales datos de imágenes de TC 3D y datos de exploración óptica 3D. La figura 3A representa cortes de DICOM asociadas con diferentes planos de una exploración de TC 3D de una estructura dentomaxilofacial, por ejemplo, un plano axial 302, un plano frontal o coronal 304 y el plano sagital 306. La figura 3B representa datos de exploración óptica 3D de una estructura dentomaxilofacial. El ordenador puede formar mallas de superficie 3D 208 de la estructura dentomaxilofacial sobre la base de los datos de exploración óptica. Además, puede emplearse una función de alineación 210 que se configura para alinear las mallas de superficie 3D con los datos de imágenes de TC 3D. Después de la alineación, las representaciones de estructuras 3D que se proporcionan a la entrada del ordenador usan el mismo sistema de coordenadas espaciales. Basándose en los datos de imágenes de TC alineados y de mallas de superficie 3D 212 pueden determinarse características posicionales y datos de vóxel clasificados del modelo 3D 214 explorados ópticamente. Las características posicionales y los datos de vóxel clasificados pueden proporcionarse entonces a la entrada de la red neuronal profunda 216, junto con la pila de imágenes 204.

Por tanto, durante la fase de entrenamiento, la red neuronal de aprendizaje profundo 3D recibe datos de entrenamiento

de TC 3D y características posicionales extraídas de los datos de entrenamiento de TC 3D como datos de entrada y los vóxeles de entrenamiento clasificados asociados con los datos de entrenamiento de TC 3D se usan como datos objetivo. Puede usarse un método de optimización para aprender los valores óptimos de los parámetros de red de la red neuronal profunda minimizando una función de pérdida que representa la desviación de la salida de la red neuronal profunda con respecto a los datos objetivo (es decir, datos de vóxel clasificados), que representa la salida deseada para un entrada predeterminada. Cuando la minimización de la función de pérdida converge a un determinado valor, el proceso de entrenamiento puede considerarse adecuado para su aplicación.

El proceso de entrenamiento representado en la figura 2 que usa características posicionales 3D en combinación con los vóxeles de entrenamiento, que pueden derivarse (al menos parcialmente) de datos de exploración de manera óptica 3D, proporciona un conjunto de entrenamiento de alta calidad para la red neuronal de aprendizaje profundo 3D. Después del proceso de entrenamiento, la red entrenada es capaz de clasificar de manera precisa vóxeles a partir de una pila de datos de imágenes de TC 3D.

Las figura 4A y 4B representan esquemas de alto nivel de arquitecturas de redes neuronales profundas para su uso en los métodos y sistemas descritos en esta divulgación. Las redes neuronales profundas pueden implementarse usando una o más redes neuronales convolucionales 3D (CNN 3D). Las capas convolucionales pueden emplear una función de activación asociada con las neuronas en las capas tales como una función sigmoide, función tanh, función ReLu, función softmax, etc. Una red neuronal profunda puede incluir una pluralidad de capas convolucionales 3D en las que variaciones menores en el número de capas y sus parámetros definitorios, por ejemplo, diferentes funciones de activación, cantidades y tamaños de núcleo, y capas funcionales adicionales como capas de normalización de lotes y de pérdida de señal pueden usarse en la implementación sin perder la esencia del diseño de la red neuronal profunda.

Como se muestra en figura 4A, la red puede incluir una pluralidad de trayectorias convolucionales en las que cada trayectoria convolucional se asocia con un conjunto de capas convolucionales 3D. En una realización, la red puede incluir al menos dos trayectorias convolucionales, una primera trayectoria convolucional asociada a un primer conjunto de capas convolucionales 3D 406 y una segunda trayectoria convolucional asociada a un segundo conjunto de capas convolucionales 3D 408. Las trayectorias convolucionales primera y segunda pueden entrenarse para codificar características 3D derivadas de datos de imágenes 3D recibidos asociados a los vóxeles que se ofrecen a la entrada de las trayectorias de convolución primera y segunda respectivamente. Además, en algunas realizaciones, la red puede incluir al menos una (tercera) trayectoria convolucional adicional asociada a un tercer conjunto de capas convolucionales 3D 407. La tercera trayectoria convolucional puede entrenarse para codificar características 3D derivadas de datos de características posicionales 3D recibidos asociados con vóxeles que se ofrecen a la entrada de la tercera trayectoria.

Alternativamente, en otra realización, en lugar de una trayectoria de convolución adicional que se entrena basándose en datos de características posicionales 3D, los datos de características posicionales 3D pueden asociarse con los valores de intensidad de vóxeles que se ofrecen a la entrada de las trayectorias de convolución primera y segunda. Por tanto, en esta realización, las trayectorias convolucionales primera y segunda pueden entrenarse basándose en datos de entrenamiento que incluyen una pila de datos 3D de valores de vóxel que incluyen valores de intensidad e información de características posicionales.

La función de las diferentes trayectorias se ilustra con más detalle en la figura 4B. Como se muestra en esta figura, los vóxeles se alimentan a la entrada de la red neuronal. Estos vóxeles están asociados con un volumen predeterminado, que puede denominarse como el volumen de imagen 401<sub>3</sub>. El volumen total de vóxeles puede dividirse en primeros bloques de vóxeles y las capas de convolución 3D de la primera trayectoria 403<sub>1</sub> pueden realizar una operación de convolución 3D en cada uno de los primeros bloques de vóxeles 401<sub>1</sub> de los datos de imágenes 3D. Durante el procesamiento, la salida de cada capa de convolución 3D puede ser la entrada de una capa de convolución 3D subsecuente. De esta manera, cada capa convolucional 3D puede generar un mapa de características 3D que representa características de los datos de imágenes 3D que se alimentan a la entrada. Por lo tanto, una capa convolucional 3D que se configura para generar tales mapas de características puede denominarse capa de características de CNN 3D.

Como se muestra en la figura 4B, las capas convolucionales de la segunda trayectoria convolucional 403<sub>2</sub> pueden configurarse para procesar segundos bloques de vóxeles 401<sub>2</sub> de los datos de imágenes 3D. Cada segundo bloque de vóxeles se asocia con un primer bloque de vóxeles, en el que el bloque de vóxeles primero y segundo tiene el mismo origen centrado en el volumen de imagen. El volumen del segundo bloque es mayor que el volumen del primer bloque. Además, el segundo bloque de vóxeles representa una versión de muestra reducida de un primer bloque de vóxeles asociado. El muestreo reducido puede basarse usando un algoritmo de interpolación conocido. El factor de muestreo reducido puede ser cualquier valor adecuado. En una realización, el factor de muestreo reducido puede seleccionarse entre 20 y 2, preferiblemente entre 10 y 3.

Por tanto, la red neuronal profunda 3D puede comprender al menos dos trayectorias convolucionales. Una primera trayectoria convolucional 403<sub>1</sub> puede definir un primer conjunto de capas de características de CNN 3D (por ejemplo, 5-20 capas), que se configuran para procesar datos de entrada (por ejemplo, los primeros bloques de vóxeles en

posiciones predeterminadas en el volumen de imagen) en una primera resolución de vóxel, por ejemplo, la resolución de vóxel del objetivo (es decir, la resolución de los vóxeles de los datos de imágenes 3D que van a clasificarse). Del mismo modo, una segunda trayectoria convolucional puede definir un segundo conjunto de capas de características de CNN 3D (por ejemplo, 5-20 capas), que se configuran para procesar datos de entrada en una segunda resolución de vóxel (por ejemplo, segundos bloques de vóxeles en los que cada bloque de los segundos bloques de vóxeles  $401_2$  tiene el mismo punto central que su bloque asociado desde el primer bloque de vóxeles  $401_1$ ). En el presente documento, la segunda resolución es menor que la primera resolución. Por tanto, los segundos bloques de vóxeles representan un volumen mayor en dimensiones reales que los primeros bloques. De esta manera, las segundas capas de características de CNN 3D procesan vóxeles con el fin de generar mapas de características 3D que incluyan información sobre la vecindad (directa) de vóxeles asociados que se procesan por las primeras capas de características de CNN 3D.

La segunda trayectoria por tanto permite a la red neuronal determinar información contextual, es decir, información sobre el contexto (por ejemplo, su entorno) de vóxeles de los datos de imágenes 3D que se presentan a la entrada de la red neuronal. Al usar múltiples trayectorias convolucionales (paralelas), tanto los datos de imágenes 3D (los datos de entrada) como la información contextual sobre vóxeles de los datos de imágenes 3D pueden procesarse en paralelo. La información contextual es útil para clasificar estructuras dentomaxilofaciales, que normalmente incluyen estructuras dentales estrechamente empaquetadas que son difíciles de distinguir, especialmente en el caso de datos de imágenes CBCT.

En una realización, la red neuronal de 4B puede incluir además una tercera trayectoria convolucional  $403_3$  de un tercer conjunto de capas convolucionales 3D que están entrenadas para procesar representaciones específicas de características posicionales 3D  $404$  que pueden extraerse de los datos de imágenes 3D. La extracción de las características posicionales 3D a partir de los datos de imágenes 3D puede realizarse como una etapa de procesamiento previo. En una realización alternativa, en lugar de usar una tercera trayectoria convolucional para procesar características posicionales 3D, la información posicional 3D, que incluye características posicionales 3D, puede asociarse con los datos de imágenes 3D que se ofrecen a la entrada de la red neuronal profunda. En particular, puede formarse una pila de datos 3D en la que cada vóxel se asocia con un valor de intensidad e información posicional. Por tanto, la información posicional puede emparejarse según el vóxel recibido aplicable, por ejemplo, mediante la adición de la información de características posicionales 3D como canales adicionales a la información de imagen 3D recibida. Por tanto, en esta realización, un vóxel de una representación de vóxel de una estructura dentomaxilofacial 3D en la entrada de la red neuronal profunda puede no solo asociarse con un valor de vóxel que representa, por ejemplo, un valor de intensidad de radio, sino también con información posicional 3D. Por tanto, en esta realización, durante el entrenamiento de las capas convolucionales de la trayectoria convolucional primera y segunda, tanto la información derivada de las características de imágenes 3D como de características posicionales 3D puede codificarse en estas capas convolucionales.

La salida de los conjuntos de capas de características de CNN 3D se fusionan y se alimentan a la entrada de un conjunto de capas de CNN 3D completamente conectadas  $410$ , que están entrenadas para derivar la clasificación prevista de vóxeles  $412$  que se ofrecen en la entrada de la red neuronal y se procesan por las capas de características de CNN 3D.

Los conjuntos de capas de características de CNN 3D están entrenados (a través de sus parámetros que pueden aprenderse) para derivar y transmitir la información óptimamente útil que puede determinarse a partir de su entrada específica, las capas completamente conectadas codifican parámetros que determinarán la manera en la que la información de las trayectorias anteriores debe combinarse para proporcionar vóxeles clasificados de manera óptima  $412$ . Después de eso, pueden presentarse vóxeles clasificados en el espacio de imagen  $414$ . Por tanto, la salida de la red neuronal son vóxeles clasificados en un espacio de imagen que corresponde al espacio de imagen de los vóxeles en la entrada.

En el presente documento, la salida (la última capa) de las capas completamente conectadas puede proporcionar una pluralidad de activaciones para cada vóxel. Una activación de vóxel de este tipo puede representar una medida de probabilidad (una predicción) que define la probabilidad de que un vóxel pertenezca a una de una pluralidad de clases, por ejemplo, clases de estructura dental, por ejemplo, estructura de diente, mandíbula y/o nervio. Para cada vóxel, las activaciones de vóxel asociadas a diferentes estructuras dentales pueden limitarse con el fin de obtener un vóxel clasificado.

Las figuras 5-7 ilustran métodos para determinar características posicionales 3D en una pila de datos de imágenes 3D que representa una estructura dentomaxilofacial 3D y ejemplos de tales características posicionales. Específicamente, en el caso de características diseñadas manualmente, y tal como se describe con referencia a la figura 1, tanto la pila de datos de imágenes 3D como las características posicionales 3D asociadas se ofrecen como entrada a la red neuronal profunda 3D para que la red pueda clasificar de manera precisa los vóxeles sin riesgo de sobreajuste. Una conversión basándose en dimensiones reales garantiza una entrada comparable independientemente de la resolución de la imagen de entrada.

Una característica posicional 3D diseñada manualmente pueden proporcionar la información de red neuronal profunda

3D sobre las posiciones de vóxeles en el volumen de imagen en relación con un plano de referencia o un objeto de referencia en el volumen de imagen. Por ejemplo, en una realización, un plano de referencia puede ser un plano axial en el volumen de imagen que separa vóxeles asociados a la mandíbula superior y vóxeles con la mandíbula inferior. En otra realización, un objeto de referencia puede incluir una curva, por ejemplo, una curva 3D, aproximando al menos parte de un arco dental de dientes en los datos de imágenes 3D de la estructura dentomaxilofacial. De esta manera, las características posicionales proporcionan a la primera red neuronal profunda los medios para codificar abstracciones que indican una probabilidad por vóxel asociado a mandíbula, dientes y/o tejidos de nervio en diferentes posiciones en el volumen de imagen. Estas características posicionales pueden ayudar a la red neuronal profunda a clasificar de manera eficiente y precisa los vóxeles de una pila de datos de imágenes 3D y están diseñadas para reducir el riesgo de sobreajuste.

Con el fin de determinar planos de referencia y/u objetos de referencia en el volumen de imagen que sean útiles en el proceso de clasificación, la función de análisis de características puede determinar vóxeles de un valor de intensidad predeterminado o por encima o por debajo de un valor de intensidad predeterminado. Por ejemplo, los vóxeles asociados con valores de intensidad de brillo pueden relacionarse con tejido de dientes y/o mandíbula. De esta manera, información sobre la posición de los dientes y/o mandíbula y la orientación (por ejemplo, un ángulo de rotación) en el volumen de imagen puede determinarse por el ordenador. Si la función de análisis de características determina que el ángulo de rotación es mayor que una cantidad predeterminada (por ejemplo, mayor a 15 grados), la función puede corregir el ángulo de rotación a cero, ya que esto es más beneficioso para obtener resultados precisos.

La figura 5A ilustra un ejemplo de un diagrama de flujo 502 de un método para determinar características posicionales 3D diseñadas manualmente en datos de imágenes 3D 504, por ejemplo, una pila de datos de imágenes de TC 3D. Este proceso puede incluir determinar una o más características posicionales 3D de la estructura dentomaxilofacial, en el que una o más características posicionales 3D se configuran para la entrada a la red neuronal profunda 3D (como se comenta con referencia a la figura 4B anteriormente). Una característica posicional 3D diseñada manualmente define información de posición de vóxeles en el volumen de imagen con respecto a planos de referencia u objetos de referencia en el volumen de imagen, por ejemplo, una distancia, por ejemplo, una distancia perpendicular, entre vóxeles en el volumen de imagen y un plano de referencia en el volumen de imagen que separa la mandíbula superior de la mandíbula inferior. También puede definir la distancia entre vóxeles en el volumen de la imagen y un objeto de referencia dental, por ejemplo, un arco dental en el volumen de imagen. Puede además definir posiciones de valores de intensidad acumulados en un segundo plano de referencia del volumen de imagen, un valor de intensidad acumulado en un punto en el segundo plano de referencia que incluye valores de intensidad acumulados de vóxeles sobre o en la proximidad del paso normal a través del punto en el plano de referencia. A continuación, se describen ejemplos de características posicionales 3D.

Con el fin de determinar un objeto de referencia que proporcione información posicional del arco dental en los datos de imágenes 3D de la estructura dentomaxilofacial. Puede usarse un algoritmo de ajuste para determinar una curva, por ejemplo, una curva que sigue una fórmula polinómica, que ajusta puntos predeterminados en una nube de puntos de diferentes valores de intensidad (acumulados).

En una realización, puede determinarse una nube de puntos de valores de intensidad en un plano axial (un plano xy) del volumen de imagen. Un valor de intensidad acumulado de un punto en tal plano axial puede determinarse sumando los valores de vóxel de vóxeles situados en la normal que atraviesa un punto en el plano axial. Por tanto, los valores de intensidad obtenidos en el plano axial pueden utilizarse para encontrar una curva que se aproxime a un arco dental de los dientes.

La figura 5B representa un ejemplo de un método de aprendizaje automático como puede utilizarse para generar características posicionales 3D relevantes (no diseñadas manualmente) según una realización de la invención. En particular, la figura 5B representa una arquitectura de red neuronal profunda 3D a modo de ejemplo que puede entrenarse para generar características deseadas para procesarse por la red neuronal 3D de segmentación. Después del entrenamiento, tal modelo entrenado puede emplearse análogo al método 502 como un preprocesador que deriva características posicionales 3D relevantes basándose en todo el conjunto de datos 3D recibidos.

Al igual que con las características posicionales 3D diseñadas manualmente, el objetivo es incorporar a la información de características posicionales 3D considerando todo el conjunto de datos 3D recibidos (o al menos una parte sustancial del mismo) para su uso en la red de aprendizaje profundo 3D de segmentación que es potencialmente relevante para la tarea de clasificación y segmentación automatizadas, y puede no estar disponible de otro modo a partir del conjunto o conjuntos de submuestras ofrecidas a la red de aprendizaje profundo 3D de segmentación. De nuevo, al igual que con las características posicionales 3D diseñadas manualmente, tal información debe estar disponible por vóxel en el conjunto de datos 3D recibidos.

Una de las maneras posibles de implementar tal método de aprendizaje automático para generar automáticamente características posicionales 3D es una red neuronal profunda entrenada. Tal red puede entrenarse para derivar características posicionales 3D basándose en un conjunto de datos 3D de entrada (por ejemplo, una representación de vóxel de una estructura dentomaxilofacial) que se ofrece a la entrada de la red neuronal profunda de segmentación 3D. En una realización, la red neuronal profunda de procesamiento previo puede ser una red neuronal profunda del



tipo U-net 3D tal como se ilustra en la figura 5B. Debido a los límites de procesamiento disponibles (en su mayoría requisitos de memoria), una arquitectura de este tipo no se hará funcionar en las resoluciones de las representaciones de vóxel recibidas. Por tanto, un primer conjunto de datos 3D de entrada, una primera representación de vóxel de una primera resolución (por ejemplo, 0,2 x 0,2 x 0,2 mm por vóxel) puede muestrearse de manera reducida a una segunda representación de vóxel de una segunda resolución inferior, por ejemplo, una resolución de 1 x 1 x 1 mm por vóxel, usando un algoritmo de interpolación. Después de eso, una red neuronal profunda 3D que está entrenada basándose en representaciones de vóxel de la segunda resolución puede generar por vóxel de entrada información de característica posicional 3D. Un algoritmo de interpolación puede usarse para escalar esta información hasta la primera resolución original. De esta manera, las características posicionales 3D resultantes coinciden (espacialmente) con los vóxeles de la primera representación de vóxel dando información relevante para cada vóxel del primer conjunto de datos 3D de entrada, teniendo en cuenta la información que considera (una versión agregada de) todo el conjunto de datos 3D recibido.

Tal red neuronal profunda 3D de procesamiento previo puede entrenarse para aproximar los valores objetivo deseados (que son las características posicionales 3D deseadas). En este ejemplo específico, los objetivos pueden ser, por ejemplo, una indicación de clase por vóxel en la resolución a la que se hace funcionar la red neuronal profunda 3D de procesamiento previo. Tales indicaciones de clase pueden, por ejemplo, proceder del mismo grupo de vóxeles de entrenamiento clasificados 136, pero muestreados de forma reducida de la misma manera que se ha muestreado de forma reducida el conjunto de datos 3D recibido.

Obsérvese que tal implementación a modo de ejemplo de un método de aprendizaje automático de procesamiento previo puede considerarse de manera efectiva como una segmentación previa grosera, específicamente una que potencialmente tiene acceso a información de la representación entera (o una parte sustancial de la representación) de vóxel 3D recibida. Al emparejar la información de segmentación previa en curso con los vóxeles aplicables del espacio de imagen 3D recibido, por ejemplo, mediante la ampliación de escala, conduce a que estas características posicionales 3D se procesen en paralelo con los datos de imagen 3D recibidos, hacia un resultado en la resolución de imagen 3D recibida.

La red de procesamiento previo puede implementarse usando una variedad de capas de red neuronal 3D, tales como capas convolucionales (CNN 3D), capas de agrupamiento máximo 3D, capas deconvolucionales 3D (de-CNN 3D), y capas conectadas de manera densa. Las capas pueden usar una variedad de funciones de activación tales como lineal, tanh, ReLu, PreLu, sigmoide, etc. Las capas de CNN y de-CNN 3D pueden variar en su cantidad de filtros, tamaños de filtro y parámetros de submuestreo. Las capas de CNN y de-CNN 3D, así como las capas conectadas de manera densa, pueden variar en sus métodos de inicialización de parámetros. Las capas de normalización de lotes y/o pérdida de señal pueden emplearse en toda la arquitectura.

Siguiendo una arquitectura U-net 3D, durante el entrenamiento, los diversos filtros dentro de las capas de CNN 3D y de-CNN 3D aprenden a codificar características significativas lo que ayudaría al esfuerzo de precisión de predicción. Durante el entrenamiento, se usan conjuntos coincidentes de datos de imágenes 3D 522 y características posicionales 3D coincidentes codificadas 560 para optimizar hacia la predicción de estas últimas a partir de las primeras. Puede emplearse una función de pérdida como medida que debe minimizarse. Este esfuerzo de optimización puede ayudarse haciendo uso de optimizadores como SGD, Adam, etc.

Una arquitectura de este tipo puede emplear diversas escalas de resolución interna, reduciendo la escala de manera efectiva 526, 530, 534 como resultado de un conjunto anterior de capas de CNN 3D 524, 528, 532, a través de, por ejemplo, agrupamiento máximo o convoluciones 3D submuestreadas. El término "características significativas" se refiere en el presente documento a derivaciones (sucesivas) de información relevante para determinar los valores de salida objetivo, y también se codifican a través de las capas de de-CNN 3D, que de manera efectiva realizan una ampliación de escala empleando filtros. Al combinar datos 540, 546, 552 resultantes de tales capas de de-CNN 3D 538, 544, 554 con los datos de las "últimas" capas de CNN 3D que funcionan en la misma resolución (532 a 540, 528 a 546 y 524 a 552), pueden lograrse predicciones muy precisas. A lo largo de la trayectoria de ampliación de escala, pueden usarse capas de CNN 3D adicionales 542, 548, 554.

Cuando se utiliza para inferencia, que se ha entrenado para tener parámetros internos codificados de tal manera que la validación proporciona resultados suficientemente precisos, puede presentarse una muestra de entrada y la red de aprendizaje profundo 3D puede proporcionar características posicionales 3D predictivas 542.

En la figura 6 se proporciona un ejemplo de un objeto de referencia para su uso en determinación de características posicionales 3D diseñadas manualmente, en este caso una curva que se aproxima a un arco dental. En este ejemplo, una nube de puntos en el plano axial (xy) indica áreas de valores de alta intensidad (áreas blancas brillantes) puede indicar áreas de estructuras de dientes o mandíbula. Con el fin de determinar una curva de arco dental, el ordenador puede determinar áreas en un plano axial del volumen de imagen asociadas con vóxeles brillantes (por ejemplo, vóxeles que tienen un valor de intensidad por encima de un valor umbral predeterminado) que pueden identificarse como vóxeles de dientes o mandíbula. Estas áreas de alta intensidad pueden usarse para determinar una disposición creciente de áreas brillantes que se aproxima al arco dentomaxilofacial. De esta manera, puede determinarse una curva de arco dental, que se aproxima a un promedio de los arcos dentomaxilofaciales de la mandíbula superior y la

mandíbula inferior respectivamente. En otra realización, pueden determinarse las curvas de arco dental separadas asociadas a la mandíbula superior e inferior.

5 Las figura 7A-7E representan ejemplos de características posicionales 3D de datos de imágenes 3D según diversas realizaciones de la invención.

10 La figura 7A representa (izquierda) una imagen de un corte del plano sagital de una pila de datos de imágenes 3D y (derecha) una visualización asociada de una denominada característica de altura del mismo corte. Tal característica de altura puede codificar una posición z (una altura 704) de cada vóxel en el volumen de imagen de la pila de datos de imágenes de TC 3D en relación con un plano de referencia 702. El plano de referencia (por ejemplo, el plano axial o xy que se determina que es (la mejor aproximación de) el plano xy con una distancia aproximadamente igual tanto a la mandíbula superior como a la mandíbula inferior y sus dientes constituyentes.

15 Pueden definirse otras características posicionales 3D para codificar información espacial en un espacio xy de una pila de datos de imágenes 3D. En una realización, tal característica posicional puede basarse en una curva que se aproxima a (parte del) el arco dental. Una característica posicional de este tipo se ilustra en la figura 7B, que representa (izquierda) un corte de una pila de datos de imágenes 3D y (derecha) una visualización de la denominada característica de recorrido para el mismo corte. Esta característica de recorrido se basa en la curva que se aproxima al arco dental 706 y define la distancia relativa 708 medida a lo largo de la curva. En el presente documento, puede definirse distancia cero como el punto 710 de la curva donde la derivada del polinomio de segundo grado es (aproximadamente) cero. La distancia recorrida aumenta cuando se mueve en cualquier dirección en el eje x, desde este punto (por ejemplo, el punto donde la derivada es cero).

20 Una característica posicional 3D adicional que se basa en la curva de arco dental puede definir la distancia más corta (perpendicular) de cada vóxel en el volumen de imagen a la curva de arco dental 706. Por lo tanto, esta característica posicional puede denominarse "característica de distancia". Un ejemplo de tal característica se proporciona en la figura 7C, que representa (izquierda) un corte de la pila de datos de imágenes 3D y (derecha) una visualización de la característica de distancia para el mismo corte. Para esta característica, distancia cero significa que el vóxel está situado en la curva de arco dental 708.

25 Sin embargo, una característica posicional 3D adicional puede definir la información posicional de dientes individuales. Un ejemplo de tal característica (que también puede denominarse característica dental) se proporciona en la figura 7D, que representa (izquierda) un corte de la pila de datos de imágenes 3D y (derecha) una visualización de la característica dental para el mismo corte. La característica dental puede proporcionar información para determinar la probabilidad de encontrar vóxeles de determinados dientes en una determinada posición en el espacio de vóxel. Esta característica puede, siguiendo un plano de referencia determinado tal como el 702, codificar una suma separada de vóxeles sobre la normal a cualquier plano (por ejemplo, el plano xy o cualquier otro plano). Esta información por tanto dota a la red neuronal de una "vista" de toda la información del espacio original como se suma sobre el plano normal. Esta vista es más grande de lo que se procesaría al excluir esta característica y puede proporcionar un medio para diferenciar si una estructura dura está presente basándose en toda la información en la dirección elegida del espacio (como se ilustra en 712<sub>1,2</sub> para el plano xy).

30 La figura 7E muestra una visualización de características posicionales 3D que pueden generarse por un preprocesador de aprendizaje automático, en particular una red neuronal profunda 3D tal como se describe con respecto a la figura 5B. Estas características posicionales 3D se han renderizado por ordenador en 3D y los volúmenes 3D mostrados son el resultado de umbrales de valores predichos. A partir de la "rugosidad" relativa de la superficie que define los volúmenes puede verse que tal red y sus datos de entrada y objetivo operaban en una resolución 3D menor que la de la representación de vóxel definitiva que va a segmentarse (en el caso de este ejemplo, se empleó una resolución de 1 x 1 x 1 mm por vóxel). Como objetivos, pueden usarse los mismos datos de entrenamiento que pueden haberse empleado para la red de aprendizaje profundo 3D de segmentación, pero se muestrearon de forma reducida a una resolución aplicable que se adhiere a los requisitos de procesamiento para su uso por una red neuronal profunda 3D de procesamiento previo. Esto conduce, en efecto, a tales características posicionales 3D que contienen una segmentación previa "rugosa" de, en el caso de este ejemplo, estructuras de mandíbula 720, diente 722 y nervio 724. Para el propósito de esta ilustración, la mandíbula inferior de este paciente en particular no se ha renderizado para mostrar los vóxeles clasificados que sean más propensos a formar parte de la estructura de nervio.

35 Tal segmentación previa rugosa puede muestrearse adecuadamente, por ejemplo, por medio de interpolación, garantizando que por vóxel a la resolución de segmentación deseada (que es la resolución de vóxel recibida originalmente), la información de tal segmentación previa coincide espacialmente a la resolución deseada. Por ejemplo, la información de un vóxel en la visualización mostrada puede coincidir espacialmente con vóxeles 5 x 5 x 5 a la resolución deseada, y esta información debe emparejarse con todos los 125 vóxeles aplicables a la resolución deseada. Posteriormente, esta información puede presentarse como, o incluirse en, un conjunto de características posicionales 3D y, como se describe con referencia a la figura 4, alimentarse en la red neuronal profunda 3D de segmentación como entrada.

60 Por tanto, las figuras 5-7 muestran que una característica posicional 3D define información sobre vóxeles de una

representación de vóxel que se proporcionan a la entrada de una red neuronal profunda que está entrenada para clasificar vóxeles. La información puede agregarse a partir de toda (o una parte sustancial de) la información disponible en la representación de vóxel, en la que durante la agregación puede tenerse en cuenta la posición de un vóxel en relación con un objeto de referencia dental. Además, la información se agrega de manera que puede procesarse por posición de un vóxel en la primera representación de vóxel.

Las figuras 8A-8D representan ejemplos de la salida de una red neuronal de aprendizaje profundo entrenada según una realización de la invención. En particular, las figuras 8A-8D representan imágenes 3D de vóxeles que se clasifican usando una red neuronal de aprendizaje profundo que está entrenada usando un método de entrenamiento como se describe con referencia a la figura 2. Como se muestra en las figuras 8B y 8C, los vóxeles pueden clasificarse por la red neuronal en vóxeles que pertenecen a estructuras de dientes (figura 8B), estructuras de mandíbula (figura 8C) o estructuras de nervio (figura 8D). La figura 8A representa una imagen 3D que incluye los vóxeles que la red neuronal de aprendizaje profundo ha clasificado como dientes, mandíbula y tejido nervioso. Como muestran las figuras 8B-8D, el proceso de clasificación es preciso pero existen todavía bastante vóxeles que se pierden o que se clasifican erróneamente. Por ejemplo, como se muestra en las figuras 8B y 8C los vóxeles que pueden formar parte de la estructura de mandíbula se clasifican como vóxeles de dientes mientras que en las superficies que pertenecen a las raíces de los dientes se pierden vóxeles. Como se muestra en la figura 8D, este problema es aún más pronunciado con los vóxeles de nervio clasificados.

Con el fin de abordar el problema de los valores atípicos en los vóxeles clasificados (que forman la salida de la primera red neuronal de aprendizaje profundo), los vóxeles pueden procesarse posteriormente. La figura 9 representa un diagrama de flujo de procesamiento posterior de vóxeles clasificados de estructuras dentomaxilofaciales 3D según una realización de la invención. En particular, la figura 9 representa un diagrama de flujo de procesamiento posterior de datos de vóxel de estructuras dentomaxilofaciales que se clasifican usando una red neuronal de aprendizaje profundo como se describe con referencia a las figuras 1-8 de esta solicitud.

Como se muestra en la figura 9, el proceso puede incluir una etapa de dividir los datos de vóxel clasificados 902 de una estructura dentomaxilofacial en vóxeles que se clasifican como vóxeles de mandíbula 904, vóxeles de dientes 906 y vóxeles que se clasifican como datos de nervios 908. Como se describirá a continuación con más detalle, los vóxeles de mandíbula y dientes se procesarán posteriormente usando una segunda red neuronal de aprendizaje profundo red adicional 910. En contraste con la primera red neuronal de aprendizaje profundo inicial (que usa una pila de datos de imágenes de TC 3D de una estructura dentomaxilofacial y características posicionales asociadas como entrada), que genera la mejor clasificación posible de vóxel basándose en los datos de imágenes, la segunda red neuronal de aprendizaje profundo de "procesamiento posterior" traduce partes de la salida de la primera red neuronal de aprendizaje profundo a vóxeles de modo que la salida coincida más estrechamente con las estructuras 3D deseadas.

La red neuronal de aprendizaje profundo de procesamiento posterior codifica representaciones tanto de dientes como de mandíbula. Durante el entrenamiento de la red neuronal de aprendizaje profundo de procesamiento posterior, los parámetros de la red neuronal se ajustan de manera que la salida de la primera red neuronal de aprendizaje profundo se traduce a la representación 3D más factible de estas estructuras dentomaxilofaciales. De esta manera, pueden reconstruirse imperfecciones en los vóxeles clasificados 912. Además, la superficie de las estructuras 3D puede suavizarse 914 de modo que puedan generarse el mejor modelo factible de mandíbula y modelos de dientes 3D. Al omitir la pila de datos de imágenes de TC 3D como una fuente de información para la red neuronal de procesamiento posterior, esta etapa de procesamiento posterior es robusta frente a las variaciones no deseadas dentro de la pila de imágenes.

Debido a la naturaleza de las imágenes de (CB)CT, la salida de la primera red neuronal de aprendizaje profundo sufrirá artefactos potenciales (antes mencionada) como el promediado debido al movimiento del paciente, endurecimiento del haz, etc. Otra fuente de ruido es la varianza en los datos de imágenes capturados por diferentes elementos de obtención de imágenes de TC. Esta varianza da como resultado la introducción de diversos factores, como cantidades variables de ruido dentro de la pila de imágenes, valores variables de intensidad vóxel que representan la misma densidad (real), y potencialmente otros. Los efectos que los artefactos y fuentes de ruido mencionados anteriormente tienen en la salida de la primera red neuronal de aprendizaje profundo pueden retirarse o al menos reducirse sustancialmente por la red neuronal de aprendizaje profundo de procesamiento posterior, dando lugar a vóxeles de mandíbula segmentados 918 y vóxeles de dientes segmentados 920.

Los datos de nervios clasificados 908 pueden procesarse posteriormente por separado de los datos de mandíbula y dientes. La naturaleza de los datos de nervios, que representan estructuras de filamentos largos y delgados en la pila de datos de imágenes de TC, hace que estos datos sean menos adecuados para el procesamiento posterior por una red neuronal de aprendizaje profundo. En su lugar, los datos de nervios clasificados se procesan posteriormente usando un algoritmo de interpolación con el fin de someter a procedimiento a datos de nervios segmentados 916. Para ese fin, los vóxeles clasificados como vóxeles de nervios y que están asociados con una alta probabilidad (por ejemplo, una probabilidad del 95 % o más) se usan por el algoritmo de ajuste con el fin de construir un modelo 3D de las estructuras de nervio. Después de eso, los modelos 3D de mandíbula, dientes y nervio se combinan en un modelo 3D de la estructura dentomaxilofacial.

La figura 10 representa un ejemplo de una arquitectura de una red neuronal de aprendizaje profundo que se configura para el procesamiento posterior de vóxeles clasificados de una estructura dentomaxilofacial 3D según una realización de la invención. La red neuronal de aprendizaje profundo de procesamiento posterior puede tener una arquitectura que es similar a la primera red neuronal de aprendizaje profundo, que incluye una primera trayectoria formada por un primer conjunto de capas características de CNN 3D 1004, que se configura para procesar los datos de entrada (en este caso una parte de datos de vóxel clasificados) en la resolución del objetivo. La red neuronal de aprendizaje profundo además incluye un segundo conjunto de capas de características de CNN 3D 1006, que se configura para procesar el contexto de los datos de entrada que se procesan por las primeras capas de características de CNN 3D pero entonces a una resolución más baja que el objetivo. La salida de las capas de características de CNN 3D primera y segunda se alimentan entonces a la entrada de un conjunto de capas de CNN 3D completamente conectadas 1008 con el fin de reconstruir los datos de vóxel clasificados de manera que representan estrechamente un modelo 3D de la estructura dentomaxilofacial 3D. La salida de la capa de CNN 3D completamente conectada proporciona los datos de vóxel reconstruidos.

La red neuronal de procesamiento posterior puede entrenarse usando los mismos objetivos que la primera red neuronal de aprendizaje profundo, que representan la misma salida deseada. Durante el entrenamiento, la red se hace lo más ampliamente aplicable posible al proporcionar ruido a las entradas para representar casos excepcionales a regularizar. Inherente a la naturaleza de la red neuronal de aprendizaje profundo de procesamiento posterior, el procesamiento que realiza también da como resultado la retirada de aspectos no factibles de los datos de vóxel recibidos. Los factores en el presente documento incluyen el suavizado y llenado de estructuras dentomaxilofaciales deseadas, y la retirada directa de datos de vóxel no factibles.

Las figuras 11A y 11B representan una iteración de la red de procesamiento posterior que da como resultado la reconstrucción superficial de vóxeles clasificados según una realización de la invención. En particular, la figura 11A representa una imagen de vóxeles clasificados de estructuras de dientes, en la que los vóxeles son la salida de la primera red neuronal de aprendizaje profundo. Como se muestra en la figura, el ruido y otros artefactos en los datos de entrada dan como resultado irregularidades y artefactos en la clasificación de vóxel y, por lo tanto, estructuras de superficie 3D que incluyen huecos en conjuntos de vóxeles que representan una estructura de diente. Estas irregularidades y artefactos son especialmente visibles en la estructura de nervio alveolar inferior 1102<sub>1</sub>, y las estructuras de raíces dentales 1104<sub>1</sub> de los dientes, es decir, las áreas donde la red neuronal de aprendizaje profundo tiene que distinguir entre los vóxeles de dientes y vóxeles que forman parte del hueso de mandíbula.

La figura 11B representa el resultado del procesamiento posterior según el proceso tal como se describe con referencia a las figuras 9 y 10. Como se muestra en esta figura, la red neuronal de aprendizaje profundo de procesamiento posterior retira con éxito los artefactos que estaban presentes en los datos de entrada (los vóxeles clasificados). La etapa de procesamiento posterior reconstruye con éxito partes que estaban sustancialmente afectadas por las irregularidades y artefactos, como las estructuras de raíz 1104<sub>1</sub> de los dientes que ahora exhiben superficies uniformes que proporcionan un modelo 3D preciso de las estructuras de diente individuales 1104<sub>2</sub>. Los vóxeles de nervios de alta probabilidad 1102<sub>1</sub> (por ejemplo, una probabilidad del 95 % o más) se usan por un algoritmo de ajuste para construir un modelo 3D de las estructuras de nervio 1102<sub>2</sub>.

Aunque las figuras que representan las redes neuronales profundas 3D funcionan como redes neuronales separadas, en las que cada una de las redes neuronales tiene una determinada función, por ejemplo, procesamiento previo, clasificación y segmentación y procesamiento posterior, estas redes neuronales también pueden conectarse entre sí formando una o dos redes neuronales profundas que incluyen la funcionalidad deseada. En ese caso, diferentes redes neuronales pueden entrenarse por separado (como se describe con referencia a las figuras en esta divulgación). Después de eso, las redes entrenadas pueden conectarse entre sí formando una red neuronal profunda.

La figura 12 es un diagrama de bloques que ilustra los sistemas de procesamiento de datos a modo de ejemplo descritos en esta divulgación. El sistema de procesamiento de datos 1200 puede incluir al menos un procesador 1202 acoplado a elementos de memoria 1204 a través de un bus de sistema 1206. Como tal, el sistema de procesamiento de datos puede almacenar código de programa dentro de los elementos de memoria 1204. Además, el procesador 1202 puede ejecutar el código de programa al que se accede desde los elementos de memoria 1204 a través del bus de sistema 1206. En un aspecto, el sistema de procesamiento de datos puede implementarse como un ordenador que es adecuado para almacenar y/o ejecutar código de programa. Sin embargo, debe apreciarse que el sistema de procesamiento de datos 1200 puede implementarse en forma de cualquier sistema, incluyendo un procesador y una memoria que es capaz de realizar las funciones descritas en esta memoria descriptiva.

Los elementos de memoria 1204 pueden incluir uno o más dispositivos de memoria física como, por ejemplo, la memoria local 1208 y los uno o más dispositivos de almacenamiento masivo 1210. La memoria local puede referirse a la memoria de acceso aleatorio u otro(s) dispositivo(s) de memoria no persistente(s) generalmente usados durante la ejecución real del código de programa. Puede implementarse un dispositivo de almacenamiento masivo como un disco duro u otro dispositivo de almacenamiento de datos persistente. El sistema de procesamiento 1200 también puede incluir una o más memorias de caché (no mostradas) que proporcionan almacenamiento temporal de al menos algún código de programa con el fin de reducir el número de veces que el código de programa debe recuperarse del dispositivo de almacenamiento masivo 1210 durante la ejecución.

Los dispositivos de entrada/salida (E/S) representados como dispositivo de entrada 1212 y dispositivo de salida 1214 opcionalmente pueden acoplarse al sistema de procesamiento de datos. Los ejemplos de dispositivo de entrada pueden incluir, pero no limitarse a, por ejemplo, un teclado, un dispositivo señalador, como un ratón, o similares. Ejemplos de dispositivo de salida pueden incluir, pero no limitarse a, por ejemplo, un monitor o elemento de visualización, altavoces, o similares. El dispositivo de entrada y/o el dispositivo de salida puede acoplarse al sistema de procesamiento de datos o bien directamente o bien a través de controladores de E/S intervinientes. Un adaptador de red 1216 también puede acoplarse al sistema de procesamiento de datos para que pueda acoplarse a otros sistemas, sistemas informáticos, dispositivos de red remotos y/o dispositivos de almacenamiento remoto a través de redes privadas o públicas intervinientes. El adaptador de red puede comprender un receptor de datos para recibir datos que se transmitan por dichos sistemas, dispositivos y/o redes a dichos datos y un transmisor de datos para transmitir datos a dichos sistemas, dispositivos y/o redes. Los módems, los módems por cable y las tarjetas Ethernet son ejemplos de diferentes tipos de adaptador de red que pueden usarse con el sistema de procesamiento de datos 1200.

Como se representa en la figura 12, los elementos de memoria 1204 pueden almacenar una aplicación 1218. Debe apreciarse que el sistema de procesamiento de datos 1200 puede además ejecutar un sistema operativo (no mostrado) que puede facilitar la ejecución de la aplicación. La aplicación, que se implementa en forma de código de programa ejecutable, puede ejecutarse por el sistema de procesamiento de datos 1200, por ejemplo, por el procesador 1202. En respuesta a la ejecución de la aplicación, el sistema de procesamiento de datos puede configurarse para realizar una o más operaciones que se describirán en el presente documento con mayor detalle.

En un aspecto, por ejemplo, el sistema de procesamiento de datos 1200 puede representar un sistema de procesamiento de datos de cliente. En tal caso, la aplicación 1218 puede representar una aplicación cliente que, al ejecutarse, configura el sistema de procesamiento de datos 1200 para realizar las diversas funciones descritas en el presente documento con referencia a un "cliente". Los ejemplos de un cliente pueden incluir, pero no limitarse a, un ordenador personal, un ordenador portátil, un teléfono móvil o similares.

En otro aspecto, el sistema de procesamiento de datos puede representar un servidor. Por ejemplo, el sistema de procesamiento de datos puede representar un servidor (HTTP) en cuyo caso la aplicación 1218, cuando se ejecuta, puede configurar el sistema de procesamiento de datos para realizar operaciones de servidor (HTTP). En otro aspecto, el sistema de procesamiento de datos puede representar un módulo, unidad o función como se denomina en esta memoria descriptiva.

La terminología usada en el presente documento tiene como finalidad describir únicamente realizaciones particulares y no pretende limitar la invención. Como se usa en el presente documento, las formas singulares "un", "una", y "el/la" se pretende que también incluyan las formas plurales, a menos que el contexto indique claramente lo contrario. Se entenderá además que los términos "comprende" y/o "que comprende", cuando se usan en esta memoria descriptiva, especifican la presencia de características indicadas, números enteros, etapas, operaciones, elementos y/o componentes, pero no excluyen la presencia o adición de una o más otras características, números enteros, etapas, operaciones, elementos, componentes y/o grupos de los mismos.

**REIVINDICACIONES**

1. Un método implementado por ordenador para el procesamiento de datos 3D que representan una estructura dentomaxilofacial que comprende:

5 recibir datos 3D (402), incluyendo los datos 3D una primera representación de vóxel de la estructura dentomaxilofacial, estando asociado un vóxel a un valor de intensidad de radiación, los vóxeles de la primera representación de vóxel que definen un volumen de imagen;

10 usar los datos 3D (402) para determinar una o más características posicionales 3D (404) asociadas a la primera representación de vóxel de la estructura dentomaxilofacial, una característica posicional 3D (404) que define información sobre posiciones de vóxeles de la primera representación de vóxel relativa a una posición de un plano de referencia dental o una posición de uno o más objetos de referencia dentales en el volumen de imagen;

15 proporcionar los datos 3D (402) y las una o más características posicionales 3D (404) a la entrada de una primera red neuronal profunda 3D, DNN, la primera DNN 3D que comprende primera y segunda red neuronal convolucional 3D, CNN, capas (406, 408), estando conectadas las primeras capas de CNN 3D y las segundas capas de CNN 3D a al menos una capa de CNN 3D completamente conectada (410) y estando configuradas para procesar primeros y segundos bloques de vóxeles asociados (401<sub>1,2</sub>) de los datos 3D respectivamente, un primer bloque de vóxeles (401<sub>1</sub>) que tiene el mismo o sustancialmente el mismo punto central en el volumen de imagen que un segundo bloque de vóxeles asociado (401<sub>2</sub>), el segundo bloque de vóxeles que representa un volumen en dimensiones reales que es mayor que un volumen en dimensiones reales representado por el primer bloque de vóxeles, estando configuradas las segundas capas de CNN 3D (408) para determinar información contextual asociada a vóxeles de los primeros bloques de vóxeles y estando configurada la al menos una capa de CNN 3D completamente conectada (410) para generar activaciones de vóxel, una activación de vóxel que representa una probabilidad de que un vóxel de los datos 3D pertenezca a una de una pluralidad de clases de estructura dental, entrenándose la primera DNN 3D basándose en un conjunto de entrenamiento, incluyendo el conjunto de entrenamiento datos 3D que incluyen representaciones de vóxel de estructuras dentomaxilofaciales y una o más características posicionales 3D derivadas de los datos de imagen 3D del conjunto de entrenamiento; y,

20 recibir activaciones de vóxel para vóxeles de los datos 3D de la salida de la primera DNN 3D y clasificar los vóxeles basándose en las activaciones de vóxel y, opcionalmente, determinar una representación de vóxel de al menos una de la pluralidad de clases de estructura dental basándose en los vóxeles clasificados.

2. Método según la reivindicación 1, en el que usar los datos 3D (402) para determinar una o más características posicionales 3D (404) incluye además:

35 determinar una distancia entre un vóxel de la primera representación de vóxel y el plano de referencia dental y/o los uno o más objetos de referencia dentales en el volumen de imagen;

40 determinar valores de intensidad acumulados de vóxeles en uno o más puntos del plano de referencia del volumen de imagen, un valor de intensidad acumulado en un punto en el plano de referencia que incluye valores de intensidad acumulados de vóxeles sobre o en la proximidad del paso normal a través del punto en el plano de referencia.

45 3. Método según la reivindicación 2, en el que el plano de referencia dental incluye un plano axial en el volumen de imagen posicionado a una distancia predeterminada de la mandíbula superior y/o inferior de la estructura dentomaxilofacial; preferiblemente a una distancia aproximadamente igual a una mandíbula superior y una mandíbula inferior de la estructura dentomaxilofacial; o, en el que el uno o más objetos de referencia dentales incluye una curva de arco dental que se aproxima al menos a parte de un arco dental de la estructura dentomaxilofacial, estando determinada preferiblemente una curva de la curva de arco dental en un plano axial del volumen de imagen; y/o, incluyendo los uno o más objetos de referencia dentales uno o más dientes.

50 4. Método según reivindicación 1, en el que usar los datos 3D (402) para determinar una o más características posicionales 3D (404) incluye:

55 proporcionar una segunda representación de vóxel a la entrada de una segunda DNN 3D, preferiblemente una DNN 3D que comprende una arquitectura U-net 3D, estando entrenada la segunda DNN 3D para determinar para cada vóxel de la segunda representación de vóxel una característica posicional 3D, incluyendo preferiblemente la característica posicional 3D una medida que indica una probabilidad de que un vóxel represente mandíbula, dientes y/o tejido nervioso, en el que la primera representación de vóxel tiene una primera resolución y la segunda representación de vóxel tiene una segunda resolución, siendo la segunda resolución menor que la primera resolución, siendo preferiblemente la segunda resolución al menos tres veces menor que la primera resolución, estando entrenada más preferiblemente la segunda DNN 3D

basándose en el conjunto de entrenamiento.

5. Método según la reivindicación 1, en el que proporcionar la primera representación de vóxel y las una o más características posicionales 3D asociadas a la primera representación de vóxel a la entrada de una primera DNN 3D comprende además:

asociar cada vóxel de la primera representación de vóxel a al menos información definida por una característica posicional 3D;

dividir la primera representación de vóxel en primeros bloques de vóxeles;

proporcionar un primer bloque de los primeros bloques de vóxeles a la entrada de la primera DNN 3D en el que cada vóxel del primer bloque se asocia a un valor de intensidad de radiación y la información definida por una característica posicional 3D.

6. Método según las reivindicaciones 5 o 5, en el que la primera DNN 3D comprende además terceras capas de CNN 3D, estando conectadas las terceras capas de CNN 3D a la al menos una capa completamente conectada, en el que las terceras capas de CNN 3D están configuradas para procesar las una o más características posicionales 3D.

7. Método según cualquiera de las reivindicaciones 1-6, que comprende además:

una tercera DNN 3D configurada para procesar posteriormente los vóxeles clasificados por la primera DNN 3D para recibir los vóxeles clasificados en su entrada, estando entrenada la tercera DNN 3D para corregir vóxeles que se clasifican de manera incorrecta por la primera DNN 3D, estando entrenada preferiblemente la tercera DNN 3D basándose en vóxeles que se clasifican durante el entrenamiento de la primera DNN 3D como entrada y basándose en uno o más modelos 3D de partes de las estructuras dentomaxilofaciales del conjunto de entrenamiento como objetivo.

8. Método según cualquiera de las reivindicaciones 1-5, en el que los datos de cada vóxel proporcionados a la entrada de la primera DNN 3D comprende un valor de intensidad de radiación y al menos una característica posicional 3D.

9. Un método implementado por ordenador para entrenar una red neuronal profunda, DNN, sistema para procesar datos 3D de una estructura dentomaxilofacial que comprende:

recibir datos de entrenamiento, incluyendo los datos de entrenamiento: datos 3D, preferiblemente datos de TC de haz cónico 3D (CBCT), definiendo los datos 3D una o más representaciones de vóxel de una o más estructuras dentomaxilofaciales respectivamente, estando asociado un vóxel a un valor de intensidad de radiación, definiendo los vóxeles de una representación de vóxel un volumen de imagen; opcionalmente, incluyendo además los datos de formación: modelos 3D de partes de las estructuras dentomaxilofaciales representadas por los datos 3D de los datos de entrenamiento;

usar los datos 3D para determinar una o más características posicionales 3D para vóxeles en las una o más representaciones de vóxel, definiendo una característica posicional 3D información sobre una posición de al menos un vóxel de una representación de vóxel de una estructura dentomaxilofacial relativa a la posición de un plano de referencia dental o la posición de uno o más objetos de referencia dentales en el volumen de imagen; y,

usar los datos de entrenamiento y las una o más características posicionales 3D para entrenar una primera DNN 3D para determinar activaciones de vóxel, representando una activación de vóxel una probabilidad de que un vóxel de los datos 3D pertenezca a una de una pluralidad de clases de estructura dental, comprendiendo la primera DNN 3D primera y segunda red neuronal convolucional 3D, CNN, capas (406, 408), estando conectadas las primeras capas de CNN 3D y las segundas capas de CNN 3D a al menos una capa de CNN 3D completamente conectada (410) y estando configuradas para procesar primeros y segundos bloques de vóxeles asociados (401<sub>1,2</sub>) de los datos 3D respectivamente, un primer bloque de vóxeles (401<sub>1</sub>) que tiene el mismo o sustancialmente el mismo punto central en el volumen de imagen que un segundo bloque de vóxeles asociado (401<sub>2</sub>), representando el segundo bloque de vóxeles un volumen en dimensiones reales que es mayor que un volumen en dimensiones reales representado por el primer bloque de vóxeles, estando configuradas las segundas capas de CNN 3D (408) para determinar información contextual asociada a vóxeles de los primeros bloques de vóxeles y estando configurada la al menos una capa de CNN 3D completamente conectada (410) para generar las activaciones de vóxel.

10. Método según la reivindicación 9, que comprende además:

clasificar los vóxeles basándose en las activaciones de vóxel; y,

usar los clasificados y los uno o más modelos 3D de partes de las estructuras dentomaxilofaciales del conjunto de entrenamiento para entrenar a una tercera DNN 3D para procesar posteriormente los vóxeles clasificados, en el que el procesamiento posterior por la tercera DNN 3D incluye corregir vóxeles que están clasificados de manera incorrecta por la primera DNN 3D.

5 11. Un sistema informático adaptado para procesar datos 3D de una estructura dentomaxilofacial que comprende:

10 un medio de almacenamiento legible por ordenador que tiene código de programa legible por ordenador incorporado con el mismo, comprendiendo el ordenador el código de programa legible por ordenador; y al menos un procesador, preferiblemente un microprocesador, acoplado al medio de almacenamiento legible por ordenador, en el que en respuesta a la ejecución del código de programa legible por ordenador, el al menos un procesador se configura para realizar operaciones ejecutables que comprenden:

15 recibir datos 3D (402), incluyendo los datos 3D una primera representación de vóxel de la estructura dentomaxilofacial, estando asociado un vóxel a un valor de intensidad de radiación, definiendo los vóxeles de la primera representación de vóxel un volumen de imagen;

20 usar los datos 3D (402) para determinar una o más características posicionales 3D (404) asociadas a la primera representación de vóxel de la estructura dentomaxilofacial, definiendo una característica posicional 3D (404) información sobre posiciones de vóxeles de la primera representación de vóxel relativa a una posición de un plano de referencia dental o una posición de uno o más objetos de referencia dentales en el volumen de imagen;

25 proporcionar los datos 3D (402) y las una o más características posicionales 3D (404) a la entrada de una primera red neuronal profunda 3D, DNN, la primera DNN 3D que comprende primera y segunda red neuronal convolucional 3D, CNN, capas (406, 408), estando conectadas las primeras capas de CNN 3D y las segundas capas de CNN 3D a al menos una capa de CNN 3D completamente conectada (410) y estando configuradas para procesar primeros y segundos bloques asociados de vóxeles (401<sub>1,2</sub>) de los datos 3D respectivamente, un primer bloque de vóxeles (401<sub>1</sub>) que tiene el mismo o sustancialmente el mismo punto central en el volumen de imagen que un segundo bloque de vóxeles asociado (401<sub>2</sub>), representando el segundo bloque de vóxeles un volumen en dimensiones reales que es mayor que un volumen en dimensiones reales representado por el primer bloque de vóxeles, estando configuradas las segundas capas de CNN 3D (408) para determinar información contextual asociada a vóxeles de los primeros bloques de vóxeles y estando configurada la al menos una capa de CNN 3D completamente conectada (410) para generar activaciones de vóxel, representando una activación de vóxel una probabilidad de que un vóxel de los datos 3D pertenezca a una de una pluralidad de clases de estructura dental, estando entrenada la primera DNN 3D basándose en un conjunto de entrenamiento, incluyendo el conjunto de entrenamiento datos 3D que incluyen representaciones de vóxel de estructuras dentomaxilofaciales y una o más características posicionales 3D derivadas de los datos de imagen 3D del conjunto de entrenamiento; y,

40 recibir activaciones de vóxel para vóxeles de los datos 3D de la salida de la primera DNN 3D y clasificar los vóxeles basándose en las activaciones de vóxel y, opcionalmente, determinar una representación de vóxel de al menos una de la pluralidad de clases de estructura dental basándose en los vóxeles clasificados.

45 12. Un sistema informático según la reivindicación 11, en el que las operaciones ejecutables comprenden además:

50 proporcionar una segunda representación de vóxel a la entrada de una segunda DNN 3D, preferiblemente una DNN 3D que comprende una arquitectura U-net 3D, estando entrenada la segunda DNN 3D para determinar para cada vóxel de la segunda representación de vóxel una característica posicional 3D, incluyendo preferiblemente la característica posicional 3D una medida que indica una probabilidad de que un vóxel represente mandíbula, dientes y/o tejido nervioso, en el que la primera representación de vóxel tiene una primera resolución y la segunda representación de vóxel tiene una segunda resolución, siendo la segunda resolución menor que la primera resolución, siendo preferiblemente la segunda resolución al menos tres veces menor que la primera resolución, estando entrenada más preferiblemente la segunda DNN 3D basándose en el conjunto de entrenamiento.

60 13. Un sistema informático según la reivindicación 12, en el que las operaciones ejecutables comprenden además:

65 proporcionar los vóxeles clasificados a la entrada de una tercera DNN 3D, estando entrenada la tercera DNN 3D para corregir vóxeles que se clasifican de manera incorrecta por la primera DNN 3D, estando entrenada preferiblemente la tercera DNN 3D basándose en vóxeles que se clasifican durante el entrenamiento de la primera DNN 3D como entrada y basándose en uno o más modelos 3D de partes de las estructuras dentomaxilofaciales del conjunto de entrenamiento como objetivo.



- 5
14. Un sistema informático según cualquiera de las reivindicaciones 11-13, en el que los datos de cada vóxel proporcionados a la entrada de la primera DNN 3D comprenden un valor de intensidad de radiación y al menos una característica posicional 3D.
  15. Producto de programa informático que comprende partes de código de software configuradas para, cuando se ejecuta en la memoria de un ordenador, ejecutar las etapas de método según cualquiera de las reivindicaciones 1-10.

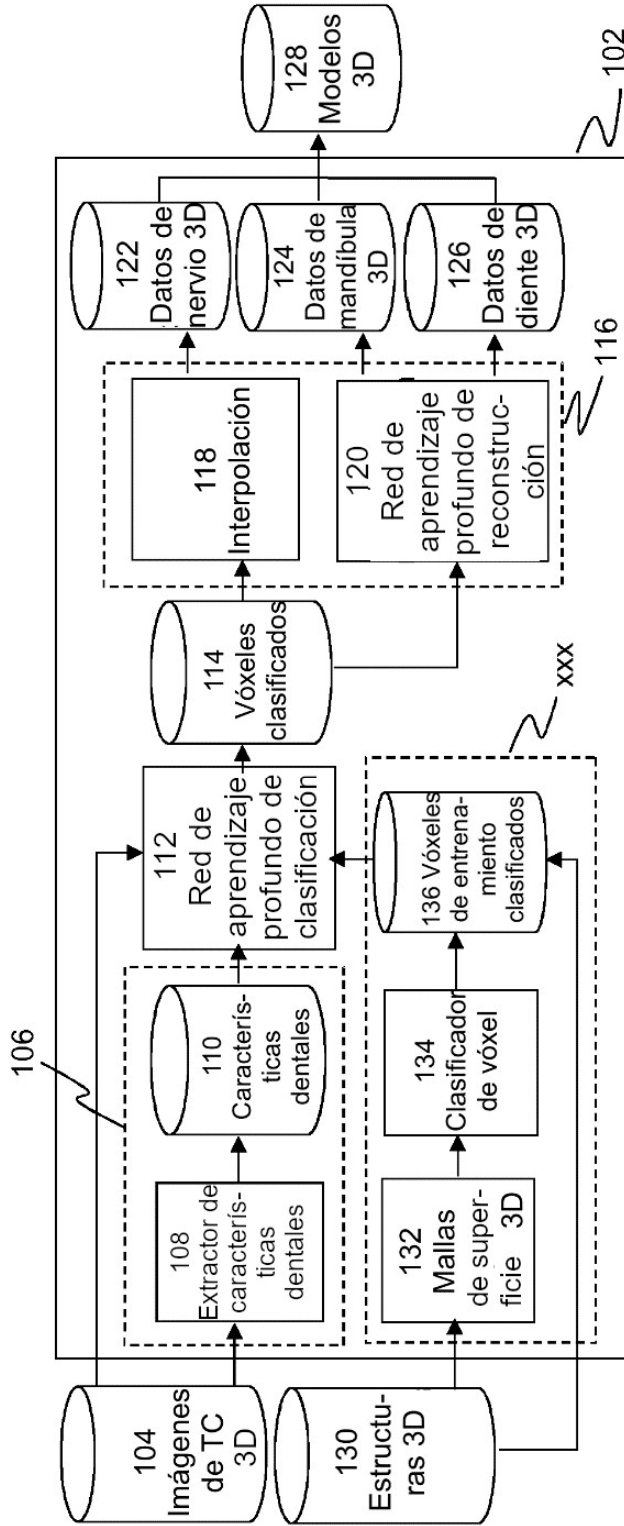


FIG. 1

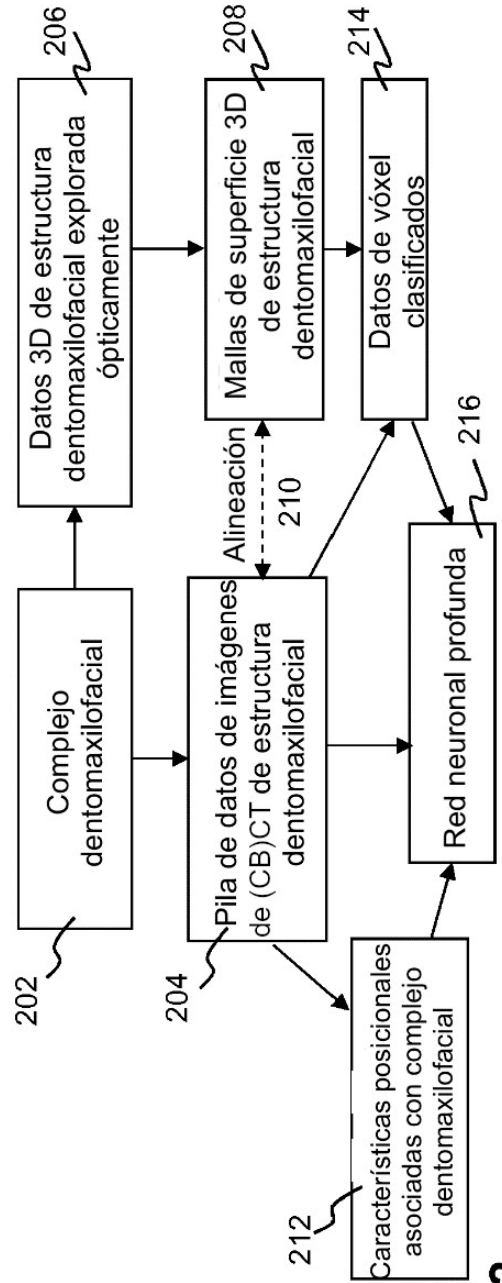
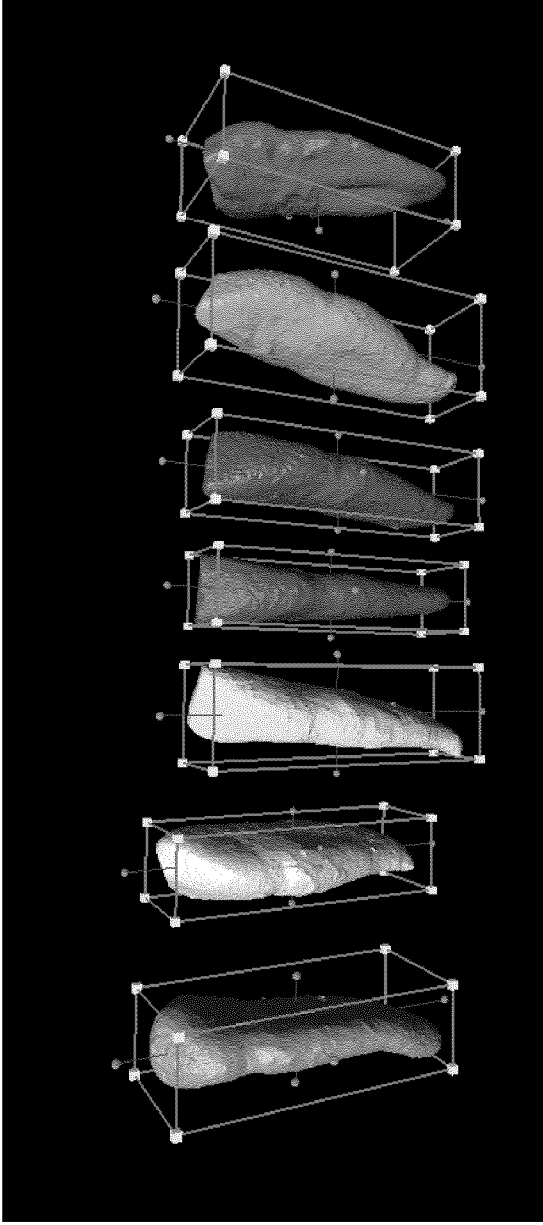
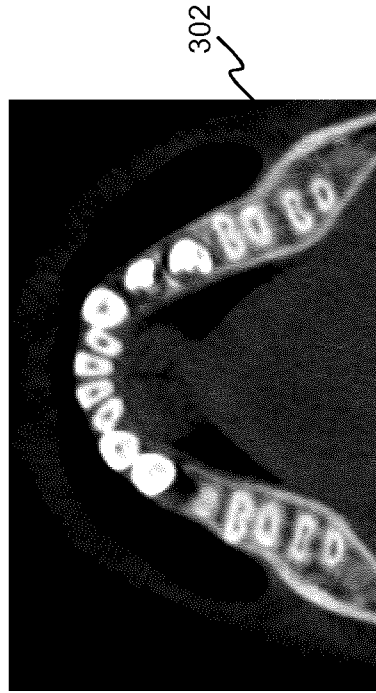
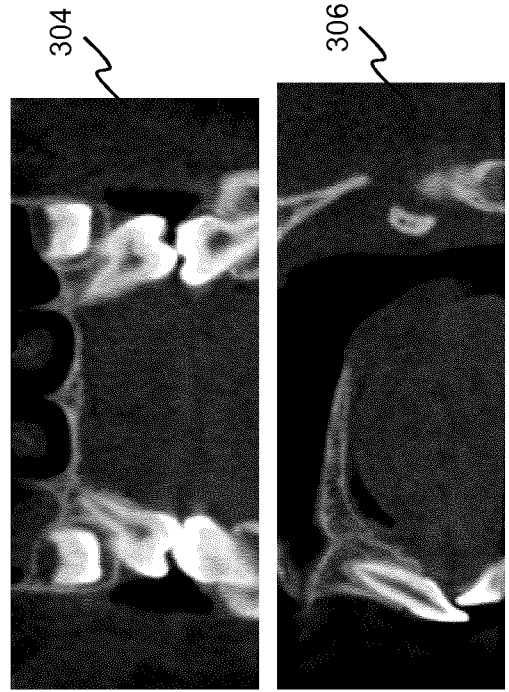


FIG. 2



**FIG. 3B**



**FIG. 3A**

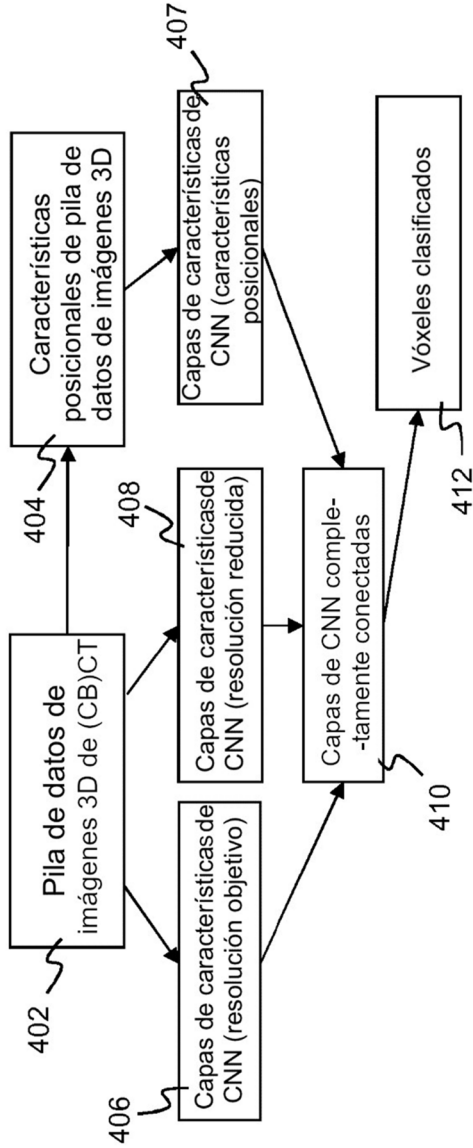


FIG. 4A

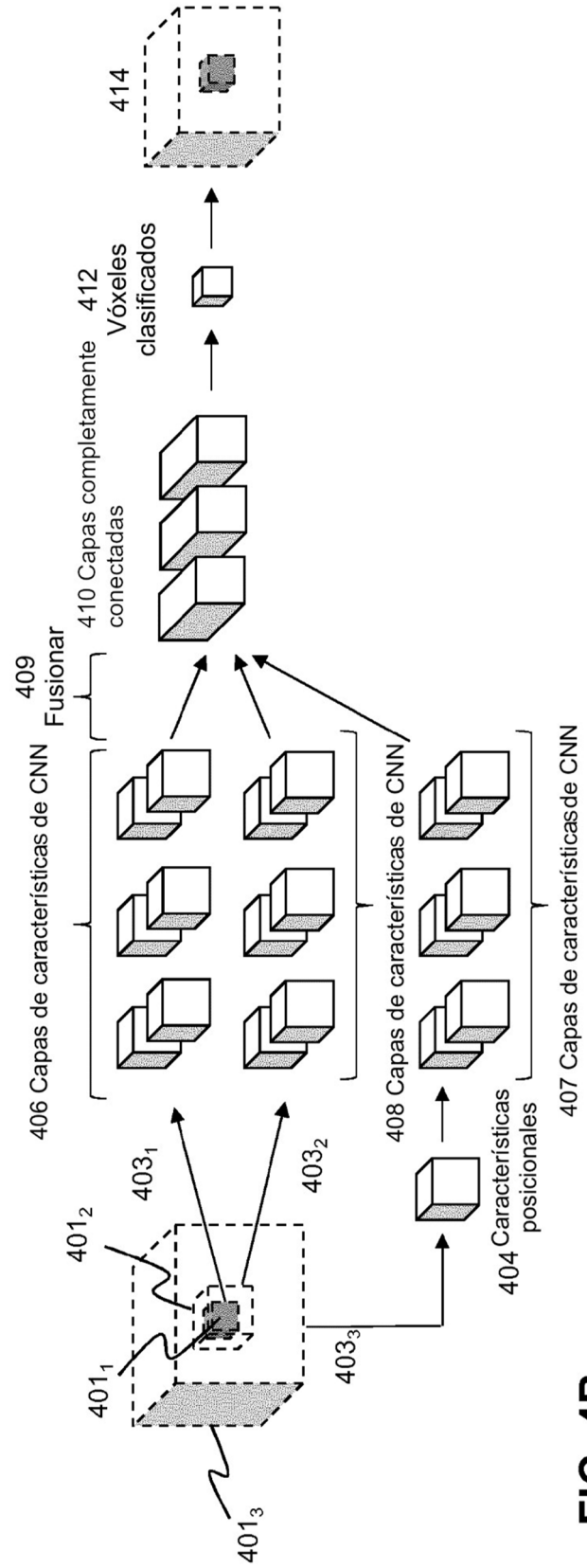


FIG. 4B

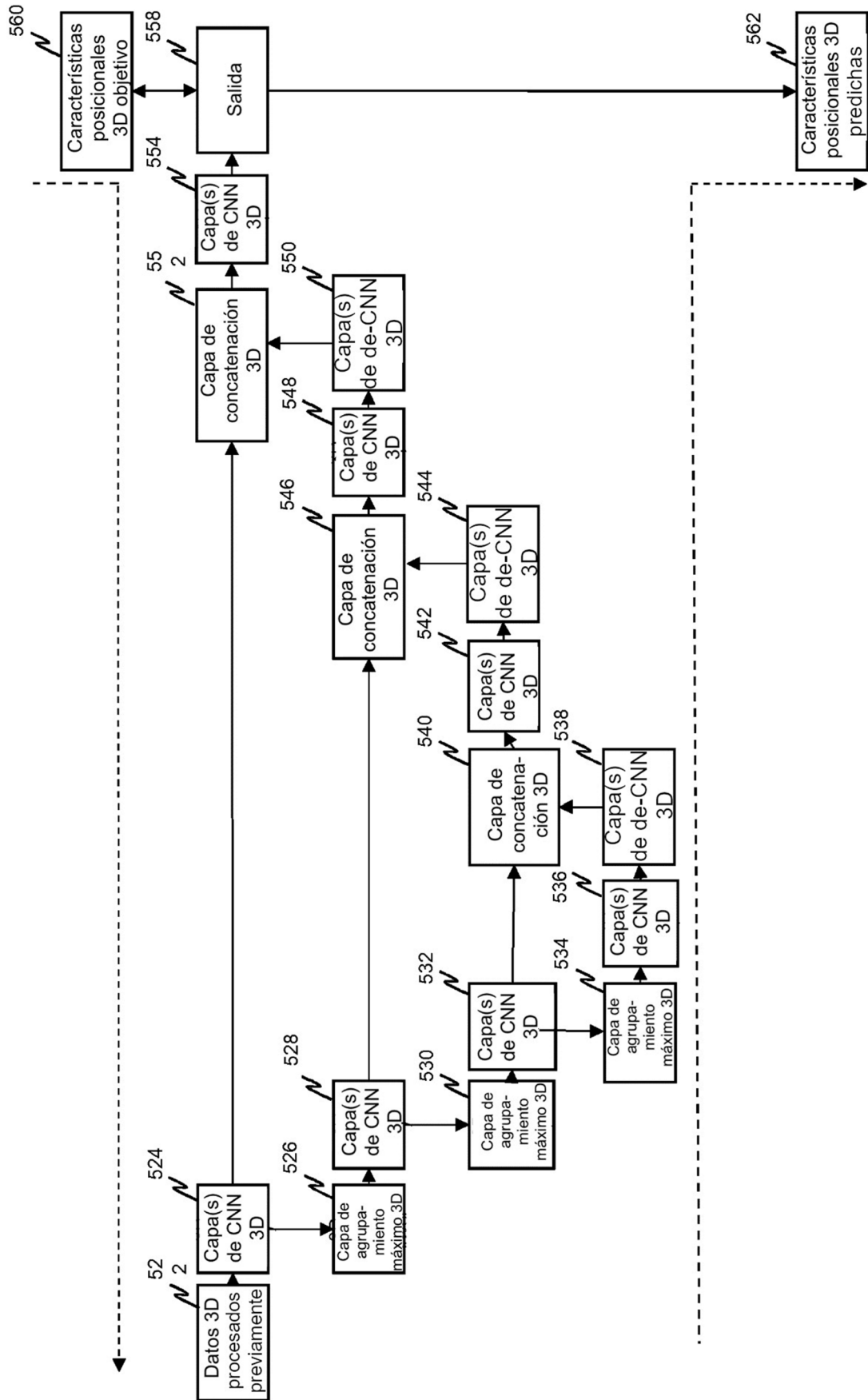


FIG. 5B

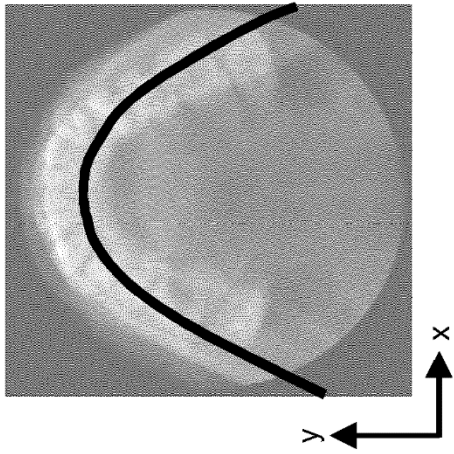


FIG. 6

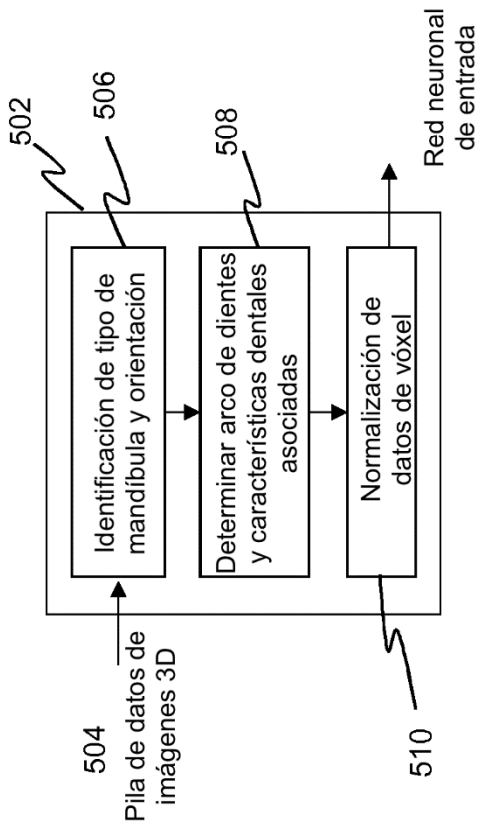


FIG. 5A

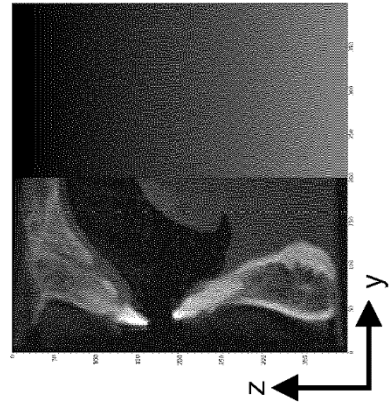
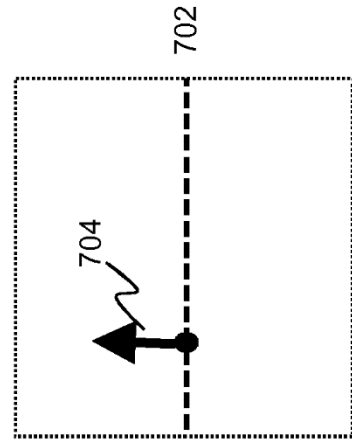
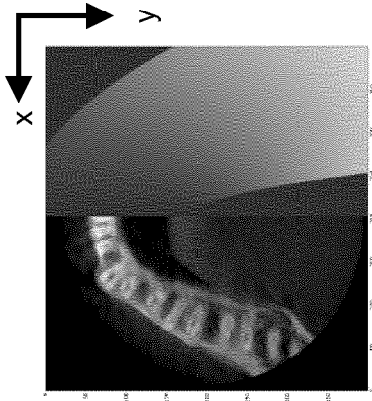
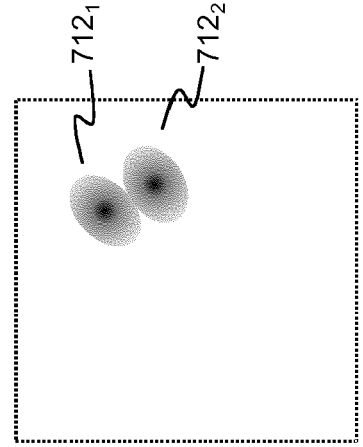
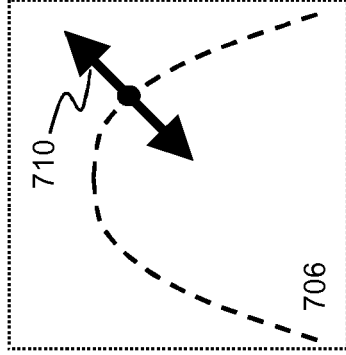
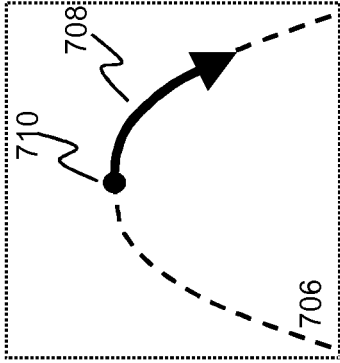
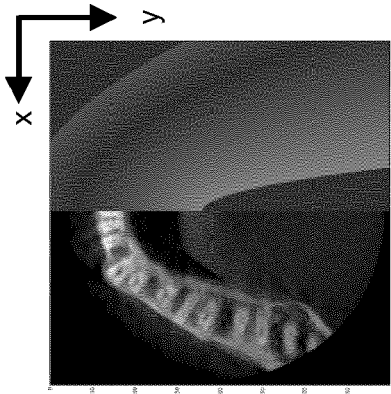


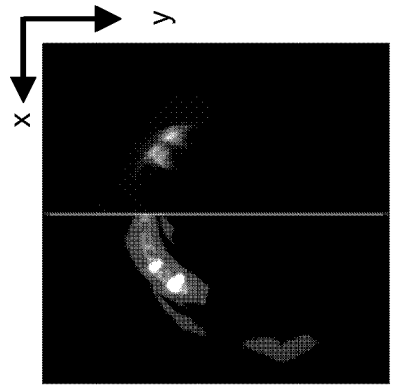
FIG. 7A



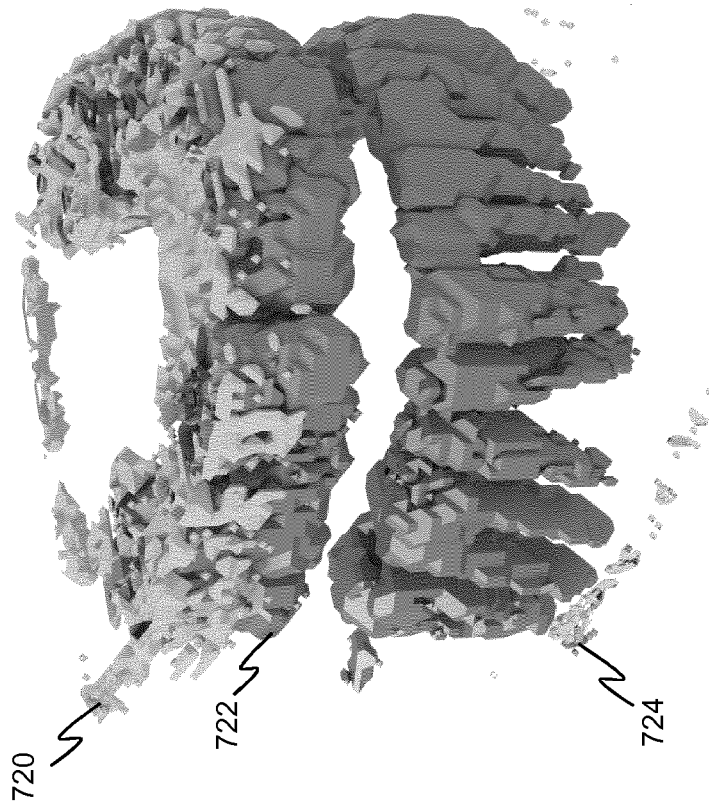
**FIG. 7B**



**FIG. 7C**

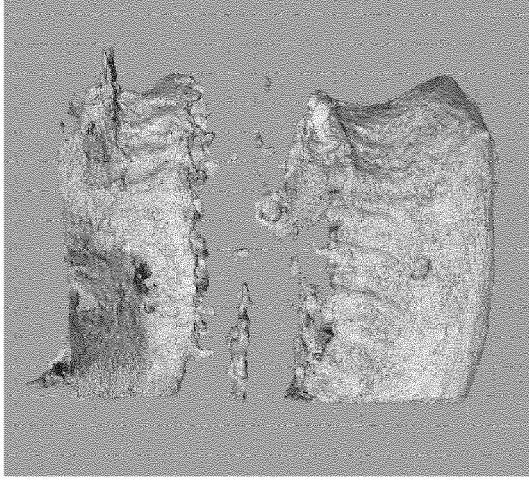


**FIG. 7D**

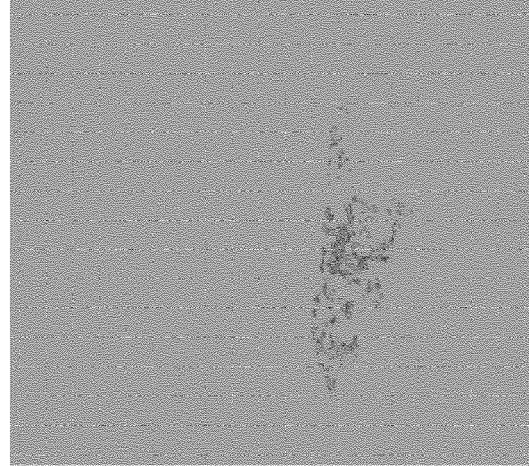


**FIG. 7E**

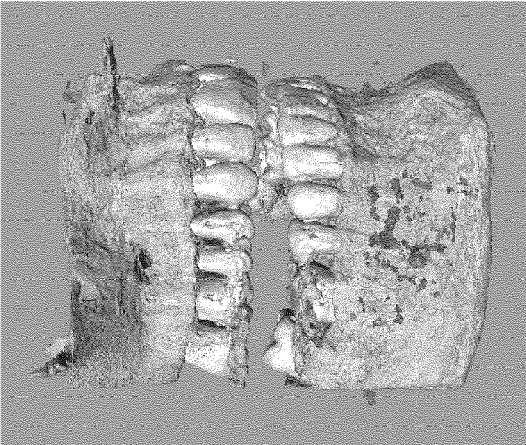




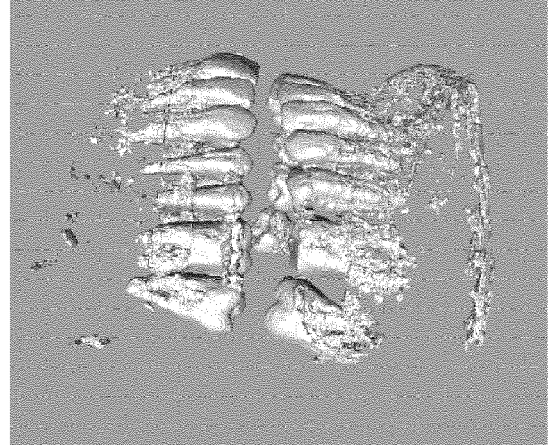
**FIG. 8C**



**FIG. 8D**



**FIG. 8A**



**FIG. 8B**

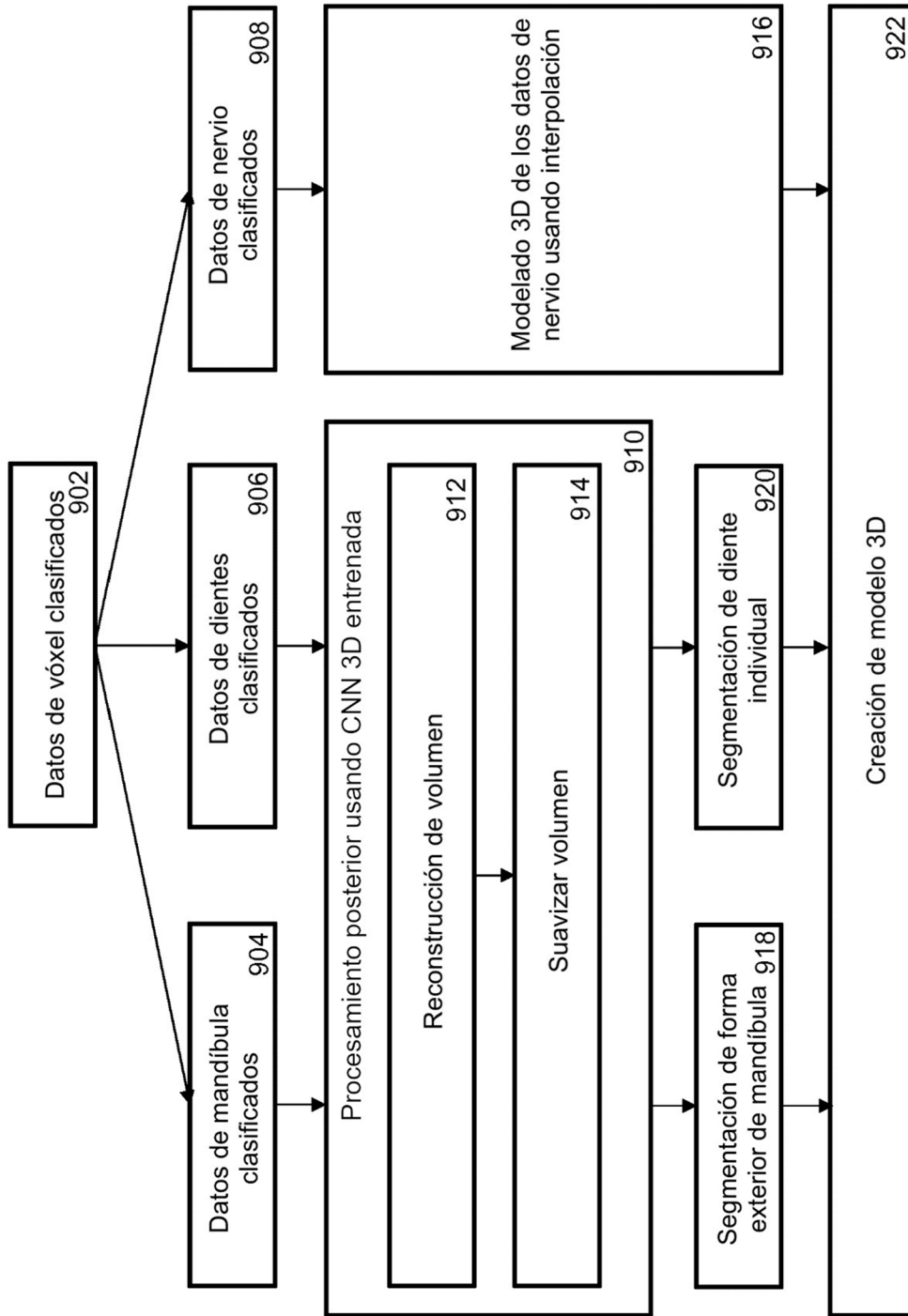
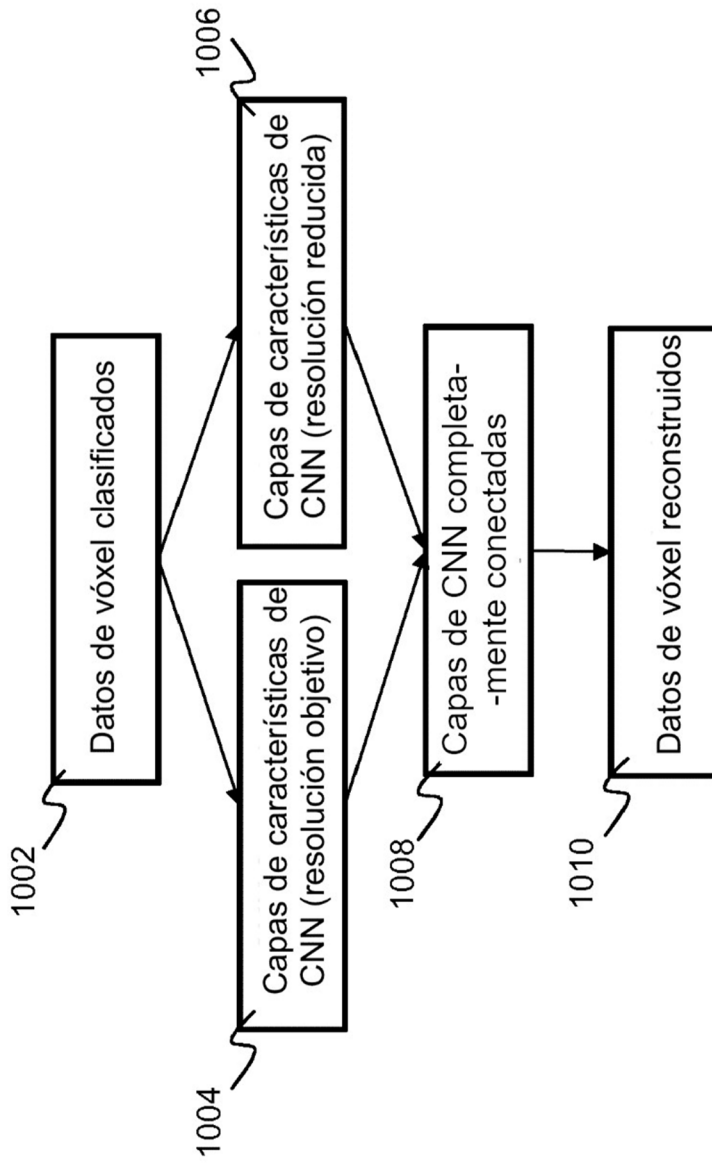
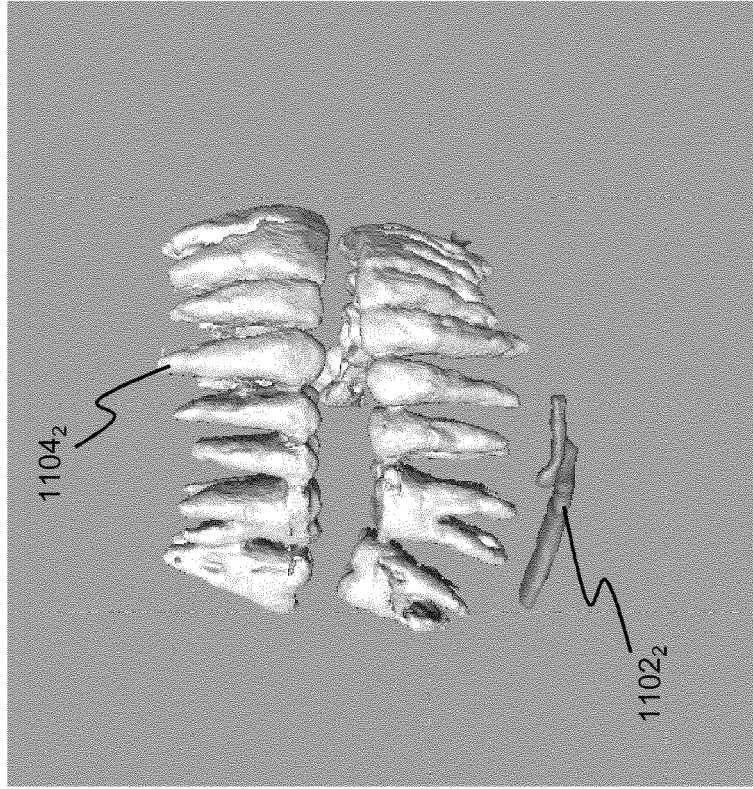


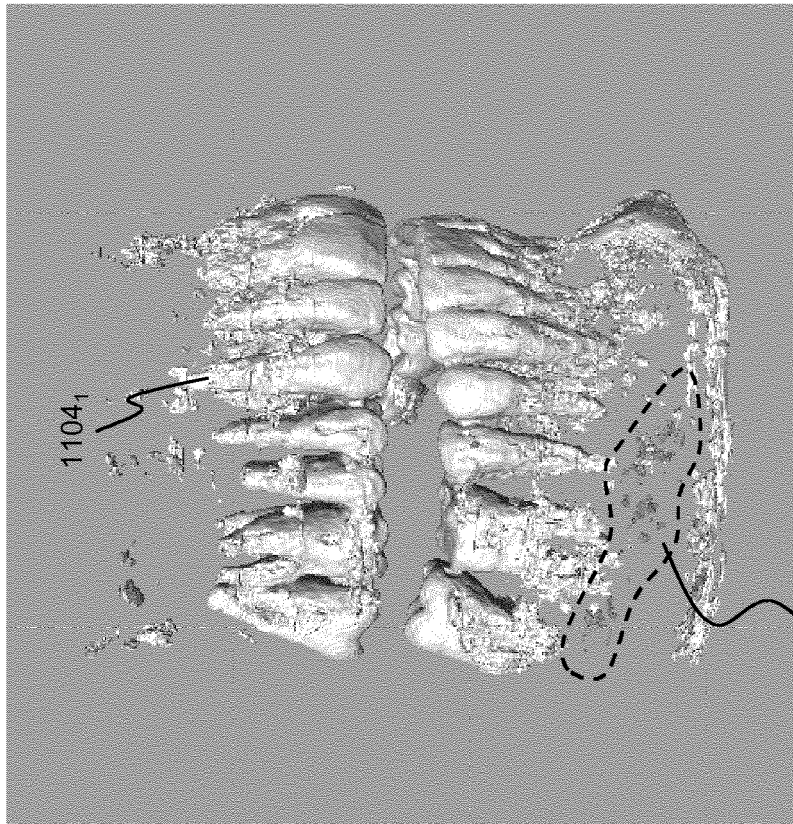
FIG. 9



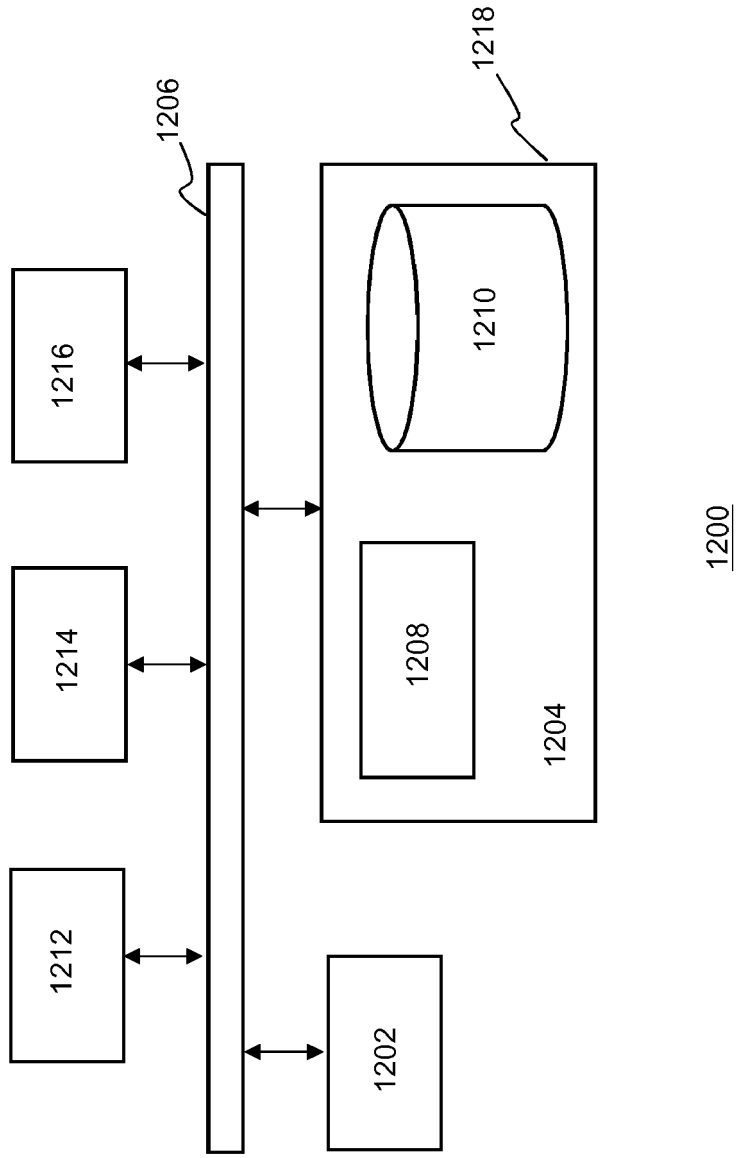
**FIG. 10**



**FIG. 11B**



**FIG. 11A**



**FIG. 12**