

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 014 296**

51 Int. Cl.:

G06V 10/764 (2012.01)

G06V 10/82 (2012.01)

G06T 7/11 (2007.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **29.03.2019** **PCT/US2019/025020**

87 Fecha y número de publicación internacional: **03.10.2019** **WO19191697**

96 Fecha de presentación y número de la solicitud europea: **29.03.2019** **E 19776693 (4)**

97 Fecha y número de publicación de la concesión europea: **01.01.2025** **EP 3776342**

54 Título: **Método y sistema para la tinción digital de imágenes de fluorescencia sin marcadores usando aprendizaje profundo**

30 Prioridad:

30.03.2018 US 201862651005 P

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

21.04.2025

73 Titular/es:

**THE REGENTS OF THE UNIVERSITY OF
CALIFORNIA (100.00%)
1111 Franklin Street 12th Floor
Oakland CA 94607-5200, US**

72 Inventor/es:

**OZCAN, AYDOGAN;
RIVENSON, YAIR;
WANG, HONGDA y
WEI, ZHENSONG**

74 Agente/Representante:

TOMAS GIL, Tesifonte Enrique

ES 3 014 296 T3

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

DESCRIPCIÓN

Método y sistema para la tinción digital de imágenes de fluorescencia sin marcadores usando aprendizaje profundo

Campo técnico

[0001] El campo técnico se refiere generalmente a métodos y sistemas usados para obtener imágenes de tejido sin teñir (es decir, sin marcadores). En particular, el campo técnico se refiere a métodos y sistemas de microscopía que utilizan aprendizaje de red neural profundo para teñir digital o virtualmente imágenes de tejido sin teñir o sin marcar. El aprendizaje profundo en redes neuronales, una clase de algoritmos de aprendizaje automático, se usa para teñir digitalmente imágenes de secciones de tejido sin marcar en imágenes que son equivalentes a imágenes de microscopía de las mismas muestras que se tiñen o marcan.

Antecedentes de la invención

[0002] La obtención microscópica de imágenes de muestras de tejido es una herramienta fundamental usada para el diagnóstico de diversas enfermedades y forma el caballo de batalla de la patología y las ciencias biológicas. La imagen estándar de referencia clínicamente establecida de una sección de tejido es el resultado de un proceso laborioso, que incluye que la muestra de tejido esté incrustada en parafina fijada en formalina (FFPE), seccionada en cortes delgados (normalmente -2-10 μm), marcada/teñida y montada en un portaobjetos de vidrio, a lo que sigue su obtención de imágenes microscópicas usando, por ejemplo, un microscopio de campo claro. Todos estos pasos usan múltiples reactivos e introducen efectos irreversibles en el tejido. Se han realizado esfuerzos recientes para cambiar este flujo de trabajo usando diferentes modalidades de obtención de imágenes. Se han realizado intentos para obtener imágenes de muestras de tejido frescas no embebidas en parafina usando métodos de microscopía no lineal basados en, por ejemplo, fluorescencia de dos fotones, generación de segundo armónico, generación de tercer armónico, así como dispersión Raman. Otros intentos han usado una fuente supercontinua controlable para adquirir imágenes multimodales para el análisis químico de muestras de tejido frescas. Estos métodos requieren el uso de láseres ultrarrápidos o fuentes supercontinuas, que podrían no estar fácilmente disponibles en la mayoría de los ajustes y requerir tiempos de exploración relativamente largos debido a señales ópticas más débiles. Además de estos, también han surgido otros métodos de microscopía para obtener imágenes de muestras de tejido no seccionadas usando excitación UV en muestras teñidas, o aprovechando la emisión de fluorescencia de tejido biológico a longitudes de onda cortas.

[0003] De hecho, la señal de fluorescencia crea algunas oportunidades únicas para obtener imágenes de muestras de tejido haciendo uso de la luz fluorescente emitida por fluoróforos endógenos. Se ha demostrado que tales firmas de fluorescencia endógena llevan información útil que puede asociarse a propiedades funcionales y estructurales de la muestra biológica y, por lo tanto, se han usado ampliamente para fines de diagnóstico e investigación. Una de las principales áreas de interés de estos esfuerzos ha sido la investigación espectroscópica de la relación entre diferentes moléculas biológicas y sus propiedades estructurales en diferentes condiciones. Algunos de estos constituyentes biológicos bien caracterizados incluyen vitaminas (por ejemplo, vitamina A, riboflavina, tiamina), colágeno, coenzimas, ácidos grasos, entre otros.

[0004] Aunque algunas de las técnicas analizadas anteriormente tienen capacidades únicas para discriminar, por ejemplo, tipos celulares y componentes subcelulares en muestras de tejido usando diversos mecanismos de contraste, los patólogos, así como el *software* de clasificación tumoral, en general, están entrenados para examinar muestras de tejido teñidas con "estándar de referencia" para tomar decisiones de diagnóstico. Motivadas parcialmente por esto, algunas de las técnicas mencionadas anteriormente se han aumentado para crear imágenes de pseudo-hematoxilina y eosina (H&E), que se basan en una aproximación lineal que relaciona la intensidad de fluorescencia de una imagen con la concentración de colorante por volumen de tejido, usando constantes determinadas empíricamente que representan la respuesta espectral media de diversos colorantes embebidos en el tejido. Estos métodos también usaron tinción exógena para potenciar el contraste de la señal de fluorescencia con el fin de crear imágenes de H&E virtuales de muestras de tejido. El artículo de las actas de conferencia de Neslihan Bayramoglu; Kaakinen Mika; Eklund Lauri; Heikkilä Janne, "Towards Virtual H&E Staining of Hiperespectral Lung Histology Images Using Conditional Generative Adversarial Network" de la conferencia internacional de talleres de visión por ordenador del IEEE (ICCVW) 2017 describe un método para teñir virtualmente especímenes sin teñir. El método utiliza reducción de dimensiones y redes generativas adversarias condicionales (cGAN) que construyen mapeos altamente no lineales entre imágenes de entrada y salida.

Resumen

[0005] En una forma de realización, se proporcionan un sistema y un método que utilizan una red neuronal profunda entrenada que se usa para la tinción digital o virtual de secciones de tejido delgadas sin marcador u otras muestras usando sus imágenes de fluorescencia obtenidas de tejido no teñido químicamente (u otras muestras). Por tejido químicamente no teñido se entiende la falta de tinciones o marcadores estándar usados en

tinción histoquímica de tejido. La fluorescencia de tejido químicamente no teñido puede incluir autofluorescencia de tejido de fluoróforos naturales o endógenos u otros emisores endógenos de luz a frecuencias diferentes de la frecuencia de iluminación (es decir, luz con desplazamiento de frecuencia). La fluorescencia de tejido químicamente no teñido puede incluir además fluorescencia de tejido de marcadores fluorescentes añadidos exógenamente u otros emisores de luz exógenos. Se obtienen imágenes de las muestras con un microscopio de fluorescencia tal como un microscopio de fluorescencia de campo amplio (o un microscopio de fluorescencia estándar). El microscopio puede utilizar un conjunto de filtros de excitación/emisión de UV cercano estándar u otros conjuntos de fuente/filtros de luz de excitación/emisión que son conocidos por los expertos en la técnica. La tinción digital o virtual se realiza, en algunas formas de realización, en una única imagen de fluorescencia obtenida de la muestra usando, en una forma de realización preferida, una red neuronal profunda entrenada.

[0006] En una forma de realización, la red neuronal profunda entrenada es una red neuronal convolucional (CNN) que se entrena usando un modelo de red generativa adversarial (GAN) para coincidir con las imágenes microscópicas de campo claro correspondientes de muestras de tejido después de que se marquen con una determinada tinción histológica. En esta forma de realización, se introduce una imagen de fluorescencia de la muestra sin teñir (por ejemplo, tejido) en la red neuronal profunda entrenada para generar la imagen teñida digitalmente. Por lo tanto, en esta forma de realización, los pasos de tinción histoquímica y formación de imágenes de campo claro se reemplazan completamente por el uso de la red neuronal profunda entrenada que genera la imagen teñida digitalmente. Como se explica en el presente documento, la inferencia de red realizada por la red neuronal entrenada es rápida tardando, en algunas formas de realización, menos de un segundo usando un ordenador de escritorio estándar para un campo de visión de formación de imágenes de $\sim 0,33 \text{ mm} \times 0,33 \text{ mm}$ usando, por ejemplo, una lente objetivo de $40\times$. Utilizando un objetivo de $20\times$ para escanear tejido, se logró un tiempo de inferencia de la red de 1,9 segundos/mm².

[0007] El método de tinción histológica digital/virtual basado en aprendizaje profundo que usa autofluorescencia se ha demostrado obteniendo imágenes de muestras de tejido humano sin marcadores que incluyen glándula salival, tiroides, riñón, hígado, pulmón y piel, donde la salida de red neuronal profunda entrenada creó imágenes equivalentes, que coinciden sustancialmente con las imágenes de las mismas muestras que se marcaron con tres tinciones diferentes, es decir, H&E (glándula salival y tiroides), tinción de Jones (riñón) y tricrómico de Masson (hígado y pulmón). Debido a que la imagen de entrada de la red neuronal profunda entrenada es capturada por un microscopio de fluorescencia convencional con un conjunto de filtros estándar, este enfoque tiene potencial transformativo para usar muestras de tejido sin teñir para aplicaciones de patología e histología, evitando completamente el proceso de tinción histoquímica, ahorrando tiempo y los costes concomitantes. Esto incluye el coste de la mano de obra, los reactivos, el tiempo adicional implicado en los procesos de tinción y similares. Por ejemplo, para las tinciones histológicas que se aproximaron usando el proceso de tinción digital o virtual descrito en el presente documento, cada procedimiento de tinción de una sección de tejido tarda de media $\sim 45 \text{ min}$ (H&E) y 2-3 horas (tinción de Jones y tricrómico de Masson), con un coste estimado, incluyendo la mano de obra, de \$2-5 para H&E y $> \$16-35$ para el tricrómico de Masson y la tinción de Jones. Además, algunos de estos procesos de tinción histoquímica requieren pasos sensibles al tiempo, lo que exige al experto monitorizar el proceso bajo un microscopio, lo que hace que todo el proceso no solo sea largo y relativamente costoso, sino que también sea laborioso. El sistema y método descritos en el presente documento evitan todos estos pasos de tinción, y también permiten la conservación de secciones de tejido no marcadas para análisis posterior, tal como micromarqueo de subregiones de interés en la muestra de tejido sin teñir que pueda usarse para análisis inmunológico y molecular más avanzados para facilitar, por ejemplo, terapias personalizadas. Además, la eficacia de tinción de este enfoque para imágenes de portaobjetos completos (WSI) correspondientes a algunas de estas muestras fue evaluada a ciegas por un grupo de patólogos, que fueron capaces de reconocer características histopatológicas con la técnica de tinción digital/virtual, logrando un alto grado de concordancia con las imágenes histológicamente teñidas de las mismas muestras.

[0008] Además, este marco de tinción histológica digital/virtual basado en aprendizaje profundo puede aplicarse ampliamente a otras longitudes de onda de excitación o conjuntos de filtros de fluorescencia, así como a otras modalidades de microscopía (como microscopía no lineal) que utilizan mecanismos de contraste endógenos adicionales. En los experimentos, se usaron muestras de tejido seccionado y fijado para poder proporcionar comparaciones significativas con los resultados del proceso de tinción histoquímica estándar. Sin embargo, el enfoque presentado también funcionaría con muestras de tejido no fijadas y no seccionadas, haciéndolo potencialmente aplicable a su uso en salas quirúrgicas o en el sitio de una biopsia para aplicaciones de diagnóstico rápido o telepatología. Más allá de sus aplicaciones clínicas, este método podría beneficiar ampliamente al campo histológico y a sus aplicaciones en la investigación y educación de las ciencias de la vida.

[0009] En una forma de realización, un método para generar una imagen microscópica teñida digitalmente de una muestra sin marcadores incluye proporcionar una red neuronal profunda entrenada que se ejecuta usando *software* de procesamiento de imágenes ejecutado usando uno o más procesadores de un dispositivo informático, donde la red neuronal profunda entrenada se entrena con una pluralidad de imágenes teñidas químicamente coincidentes o parches de imagen y sus imágenes de fluorescencia correspondientes o parches de imagen de la misma muestra. La muestra sin marcador puede incluir tejidos, células, patógenos, frotis de fluidos biológicos u otros microobjetos de interés. En algunas formas de realización, la red neuronal profunda

puede entrenarse usando una o más combinaciones de tipo de tejido/tipo de tinción química. Por ejemplo, esto puede incluir tejido de tipo A con tinción #1, tinción #2, tinción #3, etc. En algunas formas de realización, la red neuronal profunda puede ser entrenada usando tejido que ha sido teñido con múltiples tinciones.

[0010] Se introduce una imagen de fluorescencia de la muestra en la red neuronal profunda entrenada. La red neuronal profunda entrenada genera entonces una imagen microscópica teñida digitalmente de la muestra basándose en la imagen de fluorescencia de entrada de la muestra. En una forma de realización, la red neuronal profunda entrenada es una red neuronal convolucional (CNN). Esto puede incluir un CNN que usa un modelo de red generativa adversarial (GAN). La imagen de entrada de fluorescencia de la muestra se obtiene usando un microscopio de fluorescencia y una fuente de luz de excitación (por ejemplo, fuente de luz emisora de UV o UV cercano). En algunas formas de realización alternativas, se introducen múltiples imágenes de fluorescencia en la red neuronal profunda entrenada. Por ejemplo, se puede obtener una imagen de fluorescencia en una primera longitud de onda filtrada o intervalo de longitud de onda mientras que se puede obtener otra imagen de fluorescencia en una segunda longitud de onda filtrada o intervalo de longitud de onda. Estas dos imágenes de fluorescencia se introducen después en la red neuronal profunda entrenada para producir una única imagen teñida digital/virtualmente. En otra forma de realización, la imagen de fluorescencia obtenida puede someterse a una o más operaciones de preprocesamiento lineales o no lineales seleccionadas entre mejora de contraste, inversión de contraste, filtrado de imágenes que puedan introducirse solas o en combinación con la imagen de fluorescencia obtenida en la red neuronal profunda entrenada.

[0011] Por ejemplo, un método para generar una imagen microscópica teñida digitalmente de una muestra sin marcador incluye proporcionar una red neuronal profunda entrenada que se ejecuta mediante un *software* de procesamiento de imágenes usando uno o más procesadores de un dispositivo informático, donde la red neuronal profunda entrenada se entrena con una pluralidad de imágenes teñidas químicamente o parches de imagen coincidentes y sus imágenes de fluorescencia o parches de imagen correspondientes de la misma muestra. Se obtiene una primera imagen de fluorescencia de la muestra usando un microscopio de fluorescencia y donde se emite luz fluorescente a una primera longitud de onda de emisión o intervalo de longitud de onda desde fluoróforos endógenos u otros emisores endógenos de luz desplazada en frecuencia dentro de la muestra. Se obtiene una segunda imagen de fluorescencia de la muestra usando un microscopio de fluorescencia y donde se emite luz fluorescente a una segunda longitud de onda de emisión o intervalo de longitud de onda desde fluoróforos endógenos u otros emisores endógenos de luz desplazada en frecuencia dentro de la muestra. La primera y segunda imágenes de fluorescencia se pueden obtener usando diferentes combinaciones de longitud de onda de excitación/emisión. La primera y segunda imágenes de fluorescencia de la muestra se introducen a continuación en la red neuronal profunda entrenada, emitiendo la red neuronal profunda entrenada la imagen microscópica teñida digitalmente de la muestra que es sustancialmente equivalente a una imagen de campo claro correspondiente de la misma muestra que se ha teñido químicamente.

[0012] En otra forma de realización, un sistema para generar imágenes microscópicas teñidas digitalmente de una muestra no teñida químicamente incluye un dispositivo informático que tiene un *software* de procesamiento de imágenes ejecutado en el mismo o, por lo tanto, el *software* de procesamiento de imágenes que comprende una red neuronal profunda entrenada que se ejecuta usando uno o más procesadores del dispositivo informático. La red neuronal profunda entrenada se entrena con una pluralidad de imágenes teñidas químicamente coincidentes o parches de imagen y sus correspondientes imágenes de fluorescencia o parches de imagen de la misma muestra. El *software* de procesamiento de imágenes está configurado para recibir una o más imágenes de fluorescencia de la muestra y generar la imagen microscópica teñida digitalmente de la muestra que sea sustancialmente equivalente a una imagen de campo claro correspondiente de la misma muestra que se ha teñido químicamente.

Breve descripción de los dibujos

[0013]

La Figura 1 ilustra esquemáticamente un sistema que se usa para generar una imagen de salida teñida digital/virtualmente de una muestra a partir de una imagen de microscopio sin teñir de la muestra según una forma de realización.

La Figura 2 ilustra una representación esquemática de las operaciones de tinción histológica digital/virtual basada en aprendizaje profundo usando una imagen de fluorescencia de tejido no teñido.

Las Figuras 3A-3H ilustran resultados de tinción digital/virtual que coinciden con las muestras de H&E teñidas químicamente. Las dos primeras (2) columnas (Figuras 3A y 3E) muestran las imágenes de autofluorescencia de secciones de tejido de glándula salival sin teñir (usadas como entrada a la red neuronal profunda), y la tercera columna (Figuras 3C y 3G) muestra los resultados de tinción digital/virtual. La última columna (Figuras 3D y 3H) muestra las imágenes de campo claro de las mismas secciones de tejido, después del proceso de tinción histoquímica. La evaluación tanto de la Figura 3C como de la Figura 3D demuestra una pequeña isla de células tumorales infiltrantes dentro del tejido fibroadiposo subcutáneo. Obsérvese que el detalle nuclear, incluyendo la distinción de nucléolos (flechas en 3C y 3D) y la textura de la cromatina, se aprecia claramente en ambos paneles. De manera similar, en las figuras 3G y 3H las tinciones con H&E demuestran carcinoma

de células escamosas infiltrante. La reacción desmoplástica con cambio mixoide edematoso (asterisco en las figuras 3G y 3H) en el estroma adyacente es claramente identificable en ambas tinciones/paneles.

Las Figuras 4A-4H ilustran resultados de tinción digital/virtual para que coincidan con las muestras de Jones teñidas químicamente. Las dos primeras (2) columnas (Figuras 4A, 4E) muestran las imágenes de autofluorescencia de secciones de tejido renal sin teñir (usadas como entrada a la red neuronal profunda), y la tercera columna (Figuras 4C y 4G), muestra los resultados de tinción digital/virtual. La última columna (Figuras 4D, 4H) muestra las imágenes de campo claro de las mismas secciones de tejido, después del proceso de tinción histoquímica.

Las Figuras 5A-5P ilustran resultados de tinción digital/virtual para coincidir con la tinción tricrómica de Masson para secciones de tejido hepático y pulmonar. Las dos primeras (2) columnas muestran las imágenes de autofluorescencia de una sección de tejido hepático sin teñir (filas 1 y 2 - figuras 5A, 5B, 5E, 5F) y una sección de tejido pulmonar sin teñir (filas 3 y 4 - figuras 5I, 5J, 5M, 5N), usadas como entrada a la red neuronal profunda. La tercera columna (figuras 5C, 5G, 5K, 5O) muestra los resultados de tinción digital/virtual para estas muestras de tejido. La última columna (figuras 5D, 5H, 5L, 5P) muestra las imágenes de campo claro de las mismas secciones de tejido, después del proceso de tinción histoquímica.

La Figura 6A ilustra un gráfico de función de pérdida combinada frente al número de iteraciones para inicialización aleatoria e inicialización de aprendizaje de transferencia. La figura 6A ilustra cómo se logra convergencia superior usando aprendizaje de transferencia. Se inicializa una nueva red neuronal profunda usando los pesos y sesgos aprendidos de las secciones de tejido de glándula salival para lograr tinción virtual de tejido tiroideo con H&E. En comparación con la inicialización aleatoria, el aprendizaje de transferencia permite una convergencia mucho más rápida, logrando también un mínimo local más bajo.

La Figura 6B ilustra imágenes de salida de red en diferentes etapas del proceso de aprendizaje tanto para inicialización aleatoria como para aprendizaje de transferencia para ilustrar mejor el impacto del aprendizaje de transferencia para traducir el enfoque presentado a nuevas combinaciones de tejido/tinción.

La Figura 6C ilustra la imagen de campo claro teñida químicamente con H&E correspondiente.

La Figura 7A ilustra la tinción virtual (tinción H&E) de tejido cutáneo usando solamente el canal DAPI.

La Figura 7B ilustra la tinción virtual (tinción H&E) de tejido cutáneo usando los canales DAPI y Cy5. Por Cy5 se entiende un colorante de cianina marcado fluorescente de rojo lejano usado para marcar biomoléculas.

La Figura 7C ilustra el tejido correspondiente teñido histológicamente (es decir, teñido químicamente con H&E).

La Figura 8 ilustra el proceso de coincidencia y registro del campo de visión de las imágenes de autofluorescencia de muestras de tejido sin teñir con respecto a las imágenes de campo claro de las mismas muestras, después del proceso de tinción química.

La Figura 9 ilustra esquemáticamente el proceso de entrenamiento de la red de tinción virtual mediante una GAN.

La Figura 10 ilustra la arquitectura de red generativa adversarial (GAN) para el generador y discriminador según una forma de realización.

Descripción detallada de las formas de realización ilustradas

[0014] La figura 1 ilustra esquemáticamente una forma de realización de un sistema 2 para emitir imágenes teñidas digitalmente 40 desde una imagen de microscopio de entrada 20 de una muestra 22. Como se explica en el presente documento, la imagen de entrada 20 es una imagen de fluorescencia 20 de una muestra 22 (como tejido en una forma de realización) que no está teñida o marcada con un tinte o marcador fluorescente. Es decir, la imagen de entrada 20 es una imagen de autofluorescencia 20 de la muestra 22 en la que la luz fluorescente que se emite por la muestra 22 es el resultado de uno o más fluoróforos endógenos u otros emisores endógenos de luz desplazada en frecuencia contenidos en la misma. La luz desplazada en frecuencia es luz que se emite a una frecuencia (o longitud de onda) diferente que difiere de la frecuencia (o longitud de onda) incidente. Los fluoróforos endógenos o emisores endógenos de luz desplazada en frecuencia pueden incluir moléculas, compuestos, complejos, especies moleculares, biomoléculas, pigmentos, tejidos y similares. En algunas formas de realización, la imagen de entrada 20 (por ejemplo, la imagen fluorescente en bruto) se somete a una o más operaciones de preprocesamiento lineales o no lineales seleccionadas de mejora de contraste, inversión de contraste, filtrado de imagen. El sistema incluye un dispositivo informático 100 que contiene uno o más procesadores 102 en el mismo y *software* de procesamiento de imágenes 104 que incorpora la red neuronal profunda entrenada 10 (por ejemplo, una red neuronal convolucional como se explica en el presente documento en una o más formas de realización). El dispositivo informático 100 puede incluir, como se explica en el presente documento, un ordenador personal, un ordenador portátil, un dispositivo informático móvil, un servidor remoto o similares, aunque pueden usarse otros dispositivos informáticos (por ejemplo, dispositivos que incorporan una o más unidades de procesamiento gráfico (GPU)) u otros circuitos integrados de aplicación específica (ASIC). Las GPU o ASIC se pueden usar para acelerar el entrenamiento, así como la salida de imagen final. El dispositivo informático 100 puede estar asociado con o conectado a un monitor o pantalla 106 que se usa para mostrar las imágenes teñidas digitalmente 40. La pantalla 106 puede usarse para mostrar una interfaz gráfica de usuario (GUI) que usa el usuario para mostrar y ver las imágenes teñidas digitalmente 40. En una forma de realización, el usuario puede ser capaz de activar o conmutar manualmente entre múltiples tinciones digitales/virtuales diferentes para una muestra 22 particular usando, por ejemplo, la GUI. Alternativamente, la activación o conmutación entre diferentes tinciones puede realizarse automáticamente por el dispositivo informático 100. En

una forma de realización preferida, la red neuronal profunda entrenada 10 es una red neuronal convolucional (CNN).

[0015] Por ejemplo, en una forma de realización preferida como se describe en el presente documento, la red neuronal profunda entrenada 10 se entrena usando un modelo GAN. En una red neuronal profunda entrenada con GAN 10, se usan dos modelos para entrenar. Se usa un modelo generativo que captura la distribución de datos mientras que un segundo modelo estima la probabilidad de que una muestra provenga de los datos de entrenamiento en lugar del modelo generativo. Los detalles con respecto a GAN pueden encontrarse en Goodfellow *et al.*, Generative Adversarial Nets., Advances in Neural Information Processing Systems, 27, págs. 2672-2680 (2014), que se incorpora por referencia en el presente documento. El entrenamiento de red de la red neuronal profunda 10 (por ejemplo, GAN) puede realizarse en el mismo o diferente dispositivo informático 100. Por ejemplo, en una forma de realización, se puede usar un ordenador personal para entrenar la GAN, aunque tal entrenamiento puede llevar una cantidad considerable de tiempo. Para acelerar este proceso de entrenamiento, se pueden usar una o más GPU dedicadas para el entrenamiento. Como se explica en el presente documento, tal entrenamiento y prueba se realizó en GPU obtenidas de una tarjeta gráfica disponible comercialmente. Una vez que la red neuronal profunda 10 se ha entrenado, la red neuronal profunda 10 puede usarse o ejecutarse en un dispositivo informático diferente 110 que puede incluir uno con menos recursos computacionales usados para el proceso de entrenamiento (aunque las GPU también pueden integrarse en la ejecución de la red neuronal profunda entrenada 10).

[0016] El *software* de procesamiento de imágenes 104 puede implementarse usando Python y TensorFlow, aunque pueden usarse otros paquetes y plataformas de *software*. La red neuronal profunda entrenada 10 no se limita a una plataforma de *software* o lenguaje de programación particular y la red neuronal profunda entrenada 10 puede ejecutarse usando cualquier número de lenguajes o plataformas de *software* disponibles comercialmente. El *software* de procesamiento de imágenes 104 que incorpora o se ejecuta en coordinación con la red neuronal profunda entrenada 10 puede ejecutarse en un entorno local o un entorno de tipo nube. En algunas formas de realización, alguna funcionalidad del *software* de procesamiento de imágenes 104 puede ejecutarse en un lenguaje o plataforma particular (por ejemplo, normalización de imágenes) mientras que la red neuronal profunda entrenada 10 puede ejecutarse en otro lenguaje o plataforma particular. No obstante, ambas operaciones son llevadas a cabo por el *software* de procesamiento de imágenes 104.

[0017] Como se ve en la Figura 1, en una forma de realización, la red neuronal profunda entrenada 10 recibe una única imagen de fluorescencia 20 de una muestra no marcada 22. En otras formas de realización, por ejemplo, cuando se usan múltiples canales de excitación (véase la discusión de melanina en el presente documento), puede haber múltiples imágenes de fluorescencia 20 de la muestra no marcada 22 que se introducen en la red neuronal profunda entrenada 10 (por ejemplo, una imagen por canal). Las imágenes de fluorescencia 20 pueden incluir una imagen de fluorescencia de campo amplio 20 de una muestra de tejido no marcado 22. Por campo amplio se entiende que se obtiene un campo de visión amplio (FOV) escaneando un FOV más pequeño, estando el FOV amplio en el intervalo de tamaño de 10-2,000 mm². Por ejemplo, los FOV más pequeños pueden obtenerse mediante un microscopio de fluorescencia de barrido 110 que usa *software* de procesamiento de imágenes 104 para coser digitalmente los FOV más pequeños entre sí para crear un FOV más amplio. Se pueden usar FOV amplios, por ejemplo, para obtener imágenes de portaobjetos completos (WSI) de la muestra 22. La imagen de fluorescencia se obtiene usando un dispositivo de formación de imágenes 110. Para las formas de realización de fluorescencia descritas en el presente documento, esto puede incluir un microscopio de fluorescencia 110. El microscopio de fluorescencia 110 incluye una fuente de luz de excitación que ilumina la muestra 22 así como uno o más sensores de imagen (por ejemplo, sensores de imagen CMOS) para capturar luz fluorescente que es emitida por fluoróforos u otros emisores endógenos de luz desplazada en frecuencia contenida en la muestra 22. El microscopio de fluorescencia 110 puede, en algunas formas de realización, incluir la capacidad de iluminar la muestra 22 con luz de excitación a múltiples longitudes de onda diferentes o intervalos/bandas de longitud de onda. Esto puede conseguirse usando múltiples fuentes de luz diferentes y/o diferentes conjuntos de filtros (por ejemplo, conjuntos de filtros de excitación/emisión de UV o UV cercano estándar). Además, el microscopio de fluorescencia 110 puede incluir, en algunas formas de realización, múltiples conjuntos de filtros que pueden filtrar diferentes bandas de emisión. Por ejemplo, en algunas formas de realización, se pueden capturar múltiples imágenes de fluorescencia 20, cada una capturada en una banda de emisión diferente usando un conjunto de filtros diferente.

[0018] La muestra 22 puede incluir, en algunas formas de realización, una porción de tejido que está dispuesta sobre o en un sustrato 23. El sustrato 23 puede incluir un sustrato ópticamente transparente en algunas formas de realización (por ejemplo, un portaobjeto de vidrio o plástico o similar). La muestra 22 puede incluir secciones de tejido que se cortan en secciones finas usando un dispositivo de micrótopo o similar. Las secciones delgadas de tejido 22 pueden considerarse un objeto de fase de dispersión débil, que tiene modulación de contraste de amplitud limitada bajo iluminación de campo claro. La muestra 22 puede visualizarse con o sin una cubierta de cristal/cubreobjetos. La muestra puede implicar secciones congeladas o secciones de parafina (cera). La muestra de tejido 22 puede fijarse (por ejemplo, usando formalina) o no fijarse. La muestra de tejido 22 puede incluir tejido de mamífero (por ejemplo, humano o animal) o tejido vegetal. La muestra 22 también puede incluir otras muestras biológicas, muestras ambientales y similares. Los ejemplos incluyen partículas, células, orgánulos

celulares, patógenos u otros objetos de interés a microescala (aquellos con dimensiones de tamaño micrométrico o más pequeños). La muestra 22 puede incluir frotis de fluidos o tejidos biológicos. Entre ellos se cuentan, por ejemplo, frotis de sangre, frotis de Papanicolaou o de Pap. Como se explica en el presente documento, para las formas de realización basadas en fluorescencia, la muestra 22 incluye uno o más fluoróforos de origen natural o endógenos que tienen fluorescencia y son capturados por el dispositivo de microscopio de fluorescencia 110. La mayoría de los tejidos vegetales y animales muestran algo de autofluorescencia cuando se excitan con luz ultravioleta o ultravioleta cercana. Los fluoróforos endógenos pueden incluir a modo de ilustración proteínas tales como colágeno, elastina, ácidos grasos, vitaminas, flavinas, porfirinas, lipofuscinas, coenzimas (por ejemplo, NAD(P)H). En algunas formas de realización opcionales, también se pueden añadir marcadores fluorescentes añadidos exógenamente u otros emisores de luz exógenos. Como se explica en el presente documento, la muestra 22 también puede contener otros emisores endógenos de luz desplazada en frecuencia.

[0019] La red neuronal profunda entrenada 10 en respuesta a la imagen de entrada 20 produce o genera una imagen de salida marcada o teñida digitalmente 40. La imagen de salida teñida digitalmente 40 tiene "tinción" que se ha integrado digitalmente en la imagen de salida teñida 40 usando la red neuronal profunda entrenada 10. En algunas formas de realización, como las que implican secciones de tejido, la red neuronal profunda entrenada 10 parece a un observador experto (por ejemplo, un histopatólogo entrenado) sustancialmente equivalente a una imagen de campo claro correspondiente de la misma muestra de sección de tejido 22 que se ha teñido químicamente. De hecho, como se explica en el presente documento, los resultados experimentales obtenidos usando la red neuronal profunda entrenada 10 muestran que los patólogos entrenados fueron capaces de reconocer características histopatológicas tanto con técnicas de tinción (teñidas químicamente frente a teñidas digital/virtualmente) como con un alto grado de acuerdo entre las técnicas, sin una técnica de tinción clara preferible (virtual frente a histológica). Esta tinción digital o virtual de la muestra de la sección de tejido 22 aparece como si la muestra de la sección de tejido 22 hubiera sido sometida a tinción histoquímica aunque no se hubiera realizado tal operación de tinción.

[0020] La figura 2 ilustra esquemáticamente las operaciones implicadas en una forma de realización típica basada en fluorescencia. Como se observa en la figura 2, se obtiene una muestra 22 tal como una sección de tejido sin teñir. Esto puede obtenerse a partir de tejido vivo, tal como a través de una biopsia B o similar. La muestra de la sección de tejido 22 sin teñir se somete entonces a formación de imágenes fluorescentes usando un microscopio de fluorescencia 110 y genera una imagen de fluorescencia 20. Esta imagen de fluorescencia 20 se introduce a continuación en una red neuronal profunda entrenada 10 que emite rápidamente una imagen teñida digitalmente 40 de la muestra de sección de tejido 22. Esta imagen teñida digitalmente 40 se asemeja mucho a la apariencia de una imagen de campo claro de la misma muestra de sección de tejido 22 que tenía la muestra de sección de tejido real 22 sujeta a tinción histoquímica. La figura 2 ilustra (usando flechas discontinuas) el proceso convencional por el cual la muestra de la sección de tejido 22 se somete a tinción histoquímica 44 seguido de formación de imágenes microscópicas de campo claro 46 convencionales para generar una imagen de campo claro 48 convencional de la muestra de la sección de tejido teñida 22. Como se ve en la figura 2, la imagen teñida digitalmente 40 se asemeja mucho a la imagen teñida químicamente 48 real. Se obtienen perfiles de resolución y color similares usando la plataforma de tinción digital descrita en el presente documento. Esta imagen teñida digitalmente 40 puede, como se ilustra en la figura 1, mostrarse o visualizarse en un monitor de ordenador 106, pero debe apreciarse que la imagen teñida digitalmente 40 puede mostrarse en cualquier pantalla adecuada (por ejemplo, monitor de ordenador, tableta, dispositivo informático móvil, teléfono móvil, etc.). Se puede visualizar una GUI en el monitor de ordenador 106 de modo que el usuario pueda ver y opcionalmente interactuar con la imagen teñida digitalmente 40 (por ejemplo, zoom, corte, resalte, marca, ajuste de la exposición y similares).

Experimental - Tinción digital de tejido sin marcador usando autofluorescencia

Tinción virtual de muestras de tejido

[0021] El sistema 2 y los métodos descritos en el presente documento se probaron y demostraron usando diferentes combinaciones de muestras de secciones de tejido 22 y tinciones. Después del entrenamiento de una red neural profunda basada en CNN 10, su inferencia se probó ciegamente alimentándola con las imágenes de autofluorescencia 20 de secciones de tejido sin marcador 22 que no se solaparon con las imágenes que se usaron en los conjuntos de entrenamiento o validación. Las figuras 4A-4H ilustran los resultados para una sección de tejido de glándula salival, que se teñió digital/virtualmente para coincidir con imágenes de campo brillante teñidas 48 con H&E (es decir, imágenes reales) de la misma muestra 22. Estos resultados demuestran la capacidad del sistema 2 para transformar una imagen de fluorescencia 20 de una sección de tejido sin marcador 22 en una imagen equivalente de campo claro 40, que muestra el esquema de color correcto que se espera de un tejido teñido con H&E, que contiene diversos constituyentes tales como células epitelioides, núcleos celulares, nucléolos, estroma y colágeno. La evaluación tanto de la figura 3C como de la 3D muestra que las tinciones con H&E demuestran una pequeña isla de células tumorales infiltrantes dentro del tejido fibroadiposo subcutáneo. Obsérvese que el detalle nuclear, incluyendo la distinción de nucléolos (flecha) y la textura de la cromatina, se aprecia claramente en ambos paneles. De manera similar, en las figuras 3G y 3H, las tinciones con H&E demuestran la infiltración de un carcinoma de células escamosas. La reacción desmoplástica

con cambio mixoide edematoso (asterisco) en el estroma adyacente es claramente identificable en ambas tinciones.

[0022] A continuación, la red profunda 10 se entrenó para teñir digital/virtualmente otros tipos de tejidos con dos tinciones diferentes, es decir, la tinción de plata con metenamina de Jones (riñón) y la tinción tricrómica de Masson (hígado y pulmón). Las figuras 4A-4H y 5A-5P resumen los resultados para la tinción digital/virtual basada en aprendizaje profundo de estas secciones de tejido 22, que coinciden muy bien con las imágenes de campo claro 48 de las mismas muestras 22, capturadas después del proceso de tinción histoquímica. Estos resultados ilustran que la red neuronal profunda entrenada 10 es capaz de inferir los patrones de tinción de diferentes tipos de tinciones histológicas usadas para diferentes tipos de tejidos, a partir de una única imagen de fluorescencia 20 de una muestra sin marcador (es decir, sin ninguna tinción histoquímica). Con la misma conclusión global que en las figuras 3A-3H, también fue confirmado por un patólogo que las imágenes de salida de red neuronal de las figuras 4C y 5G revelan correctamente las características histológicas correspondientes a hepatocitos, espacios sinusoidales, colágeno y gotitas de grasa (figura 5G), consecuente con la manera en que aparecen en las imágenes de campo claro 48 de las mismas muestras de tejido 22, capturadas después de la tinción química (figuras 5D y 5H). De manera similar, también fue confirmado por el mismo experto que las imágenes de salida de red neuronal profunda 40 proporcionadas en las figuras 5K y 5O (pulmón) revelan características histológicas teñidas de manera consistente correspondientes a vasos, colágeno y espacios alveolares tal como aparecen en las imágenes de campo claro 48 de la misma muestra de tejido 22 de la que se forman imágenes después de la tinción química (figuras 6L y 6P).

[0023] Las imágenes de salida teñidas digital/virtualmente 40 de la red neuronal profunda entrenada 10 se compararon con las imágenes de tinción histoquímica estándar 48 para diagnosticar múltiples tipos de afecciones en múltiples tipos de tejidos, que eran secciones embebidas en parafina y fijadas en formalina (FFPE) o congeladas. Los resultados se resumen en la Tabla 1 a continuación. El análisis de quince (15) secciones de tejido por cuatro patólogos certificados (que no conocían la técnica de tinción virtual) demostró una discordancia no mayor del 100 %, definida como ninguna diferencia clínicamente significativa en el diagnóstico entre observadores profesionales. El "tiempo para el diagnóstico" varió considerablemente entre los observadores, desde un promedio de 10 segundos por imagen para el observador 2 a 276 segundos por imagen para el observador 3. Sin embargo, la variabilidad intra-observador fue muy pequeña y tendía hacia un tiempo más corto para diagnosticar con las imágenes de portaobjetos teñidas virtualmente 40 para todos los observadores excepto el observador 2 que fue igual, es decir, ~10 segundos por imagen tanto para la imagen de portaobjetos virtual 40 como para la imagen de portaobjetos teñida histológicamente 48. Estos indican una utilidad diagnóstica muy similar entre las dos modalidades de imagen.

Tabla 1

Número de serie	Tejido, fijación, tipo de tinción	Patólogo #	Teñido Histoquímicamente/ Virtualmente	Diagnóstico	Tiempo para diagnosticar
1	Ovario, Sección congelada, H&E	1	VS	Adenocarcinoma	30 seg
		2	VS	Tumor seroso límite	15 seg
		3	HS	Adenocarcinoma mucinoso	10 min
		4	HS	Adenocarcinoma, endometroide	2 min

ES 3 014 296 T3

(continuada)

Número de serie	Tejido, fijación, tipo de tinción	Patólogo #	Teñido Histoquímicamente/ Virtualmente	Diagnóstico	Tiempo para diagnosticar
2	Ovario, Sección congelada, H&E	1	VS	Ovario benigno	10 seg
		2	VS	Ovario benigno	10 seg
		3	HS	Ovario normal con quiste del cuerpo lúteo	15 min
		4	HS	Normal	1 min
3	Glándula salival, FFPE, H&E	1	VS	Glándulas salivales benignas con inflamación crónica leve	10 seg
		2	VS	Tejido parótido benigno	5 seg
		3	HS	Glándula salival normal	1 min
		4	HS	No hay anomalía histopatológica	1 min
4	Glándula salival, sección congelada, H&E	1	HS	Adenoma pleomórfico	5 seg
		2	HS	Adenoma pleomórfico	5 seg
		3	VS	Adenoma pleomórfico	3 min
		4	VS	Adenoma pleomórfico	2 seg
5	Glándula salival, FFPE, H&E	1	HS	Carcinoma mucocépidermoide, de bajo grado	5 seg
		2	HS	Carcinoma de conducto salival	5 seg
		3	VS	Carcinoma mucocépidermoide	10 min
		4	VS	Carcinoma mucocépidermoide	10 seg
6	Mama, FFPE, H&E	1	VS	Carcinoma ductal invasivo y DCIS	15 seg
		2	VS	Carcinoma ductal	10 seg
		3	HS	Carcinoma ductal invasivo con DCIS	2min
		4	HS	Carcinoma invasivo	1 minuto
7	Piel, FFPE, H&E	1	HS	Melanoma maligno	30 seg
		2	HS	melanoma	30 seg
		3	VS	Melanoma	5 min
		4	VS	Melanoma	1 min
8	Próstata, FFPE, H&E	1	HS	Adenocarcinoma prostático 3+4	1 min
		2	HS	Adenocarcinoma prostático 4+3	5 seg
		3	VS	Adenocarcinoma prostático, patrón de Gleason 3+4	5 min
		4	VS	HG-PIN con cribiforme frente a carcinoma	5 min
9	Higado, FFPE, tinción de Masson	1	VS	Higado benigno con esteatosis leve	10 seg
		2	VS	Higado benigno con esteatosis	5 seg
		3	HS	Hepatosteatosis, predominantemente macrovesicular	3 min
		4	HS	Esteatosis mínima, sin fibrosis	5 min

(continuada)

Número de serie	Tejido, fijación, tipo de tinción	Patólogo #	Teñido Histoquímicamente/ Virtualmente	Diagnóstico	Tiempo para diagnosticar
10	Hígado, FFPE, tricrómico de Masson	1	HS	Hígado benigno con fibrosis en puente	10 seg
		2	HS	Hígado benigno, fibrosis en puente	5 seg
		3	VS	Cirrosis moderada	1 min
		4	VS	Inflamación portal leve, fibrosis en puente focal (Etapa 2-3)	5 minutos
11	Glándula salival, FFPE, H&E	1	VS	Carcinoma	5 seg
		2	VS	Ca intraductal	20 seg
		3	HS	Carcinoma poco diferenciado	1 min
		4	HS	Neoplasia de glándula salival de bajo grado	1 minuto
12	Glándula salival, FFPE, H&E	1	HS	Adenocarcinoma	5 seg
		2	HS	Carcinoma de conducto salival	5 seg
		3	VS	Carcinoma de conducto salival	2min
		4	VS	Neoplasia de glándula salival de bajo grado	1 minute
13	Tiroides, FFPE, H&E	1	VS	Carcinoma papilar de tiroides, tipo de célula alta	10 seg
		2	VS	Ca papilar de tiroides, célula alta	20 seg
		3	HS	Carcinoma papilar de tiroides, variante de células altas	5 min
		4	HS	PTC	10 seg
14	Tiroides, FFPE, H&E	1	HS	Carcinoma papilar de tiroides	5 seg
		2	HS	Ca medular	5 seg
		3	VS	Carcinoma papilar de tiroides, variante oncocítica	7 min
		4	VS	PTC	10 seg
15	Tiroides, FFPE, H&E	1	VS	Carcinoma papilar de tiroides	5 seg
		2	VS	Ca papilar de tiroides	5 seg
		3	HS	Carcinoma papilar de tiroides	1 min
		4	HS	PTC	10 seg

Evaluación ciega de la eficacia de tinción para imágenes de portaobjetos completos (WSI)

- 5 [0024] Después de evaluar las diferencias en la sección de tejido y las tinciones, se probó la capacidad del sistema de tinción virtual 2 en el flujo de trabajo de histología de tinción especializada. En particular, se obtuvieron imágenes de la distribución de autofluorescencia de 15 muestras sin marcador de secciones de tejido hepático y 13 secciones de tejido sin marcador de riñón con una lente objetivo de 20x/0,75NA. Todas las secciones de tejido hepático y renal se obtuvieron de diferentes pacientes e incluían biopsias pequeñas y resecciones más grandes. Todas las secciones de tejido se obtuvieron de FFPE pero no se cubrieron con cubreobjetos. Después del barrido de autofluorescencia, las secciones de tejido se tiñeron histológicamente con tricrómico de Masson (secciones de tejido hepático de 4 µm) y tinción de Jones (secciones de tejido renal de 2 µm). Los WSI se dividieron entonces en conjuntos de entrenamiento y prueba. Para la cohorte de portaobjetos de hígado, se usaron 7 WSI para entrenar el algoritmo de tinción virtual y se usaron 8 WSI para pruebas ciegas; para la cohorte de portaobjetos de riñón, se usaron 6 WSI para entrenar el algoritmo y se usaron 7 WSI para pruebas. Los patólogos del estudio no conocían las técnicas de tinción para cada WSI y se les pidió aplicar un grado numérico 1-4 para la calidad de las diferentes tinciones: 4 = perfecto, 3 = muy bueno, 2 = aceptable, 1 = inaceptable. En segundo lugar, los patólogos del estudio aplicaron la misma escala de puntuación (1-4) para

características específicas: detalle nuclear (ND), detalle citoplasmático (CD) y fibrosis extracelular (EF), para el hígado solamente. Estos resultados se resumen en la Tabla 2 (Hígado) y la Tabla 3 (Riñón) a continuación (el ganador está en negrita). Los datos indican que los patólogos fueron capaces de reconocer características histopatológicas tanto con técnicas de tinción como con un alto grado de acuerdo entre las técnicas, sin una clara técnica de tinción preferible (virtual frente a histológica).

5

Tabla 2

Tej. #	Patólogo 1				Patólogo 2				Patólogo 3				Promedio			
	ND	CD	EF	SQ	ND	CD	EF	SQ	ND	CD	EF	SQ	ND	CD	EF	SQ
1-HS	3	2	1	1	4	4	3	4	1	1	1	3	2.67	2.33	1.67	2.67
1-VS	3	3	3	3	3	3	2	3	2	2	3	3	2.67	2.67	2.67	3.00
2-HS	3	2	4	4	4	4	3	4	1	2	2	2	2.67	2.67	3.00	3.33
2-VS	3	3	4	4	4	3	3	3	2	2	3	3	3.00	2.67	3.33	3.33
3-HS	3	3	2	2	3	3	4	3	1	1	1	1	2.33	2.33	2.33	2.00
3-VS	3	2	1	1	3	3	1	4	1	1	1	1	2.33	2.00	1.00	2.00
4-HS	3	2	4	4	3	4	4	4	1	2	1	2	2.33	2.67	3.00	3.33
4-VS	3	3	4	4	4	3	4	4	2	2	3	3	3.00	2.67	3.67	3.67
5-HS	3	3	4	4	3	3	2	1	1	3	2	2	2.33	3.00	2.67	2.33
5-VS	3	2	3	3	3	3	4	2	2	1	3	3	2.67	2.00	3.33	2.67
6-HS	3	2	3	3	4	4	4	3	2	2	2	2	3.00	2.67	3.00	2.67
6-VS	3	3	4	3	4	3	4	3	1	1	1	1	2.67	2.33	3.00	2.33
7-HS	3	3	4	4	3	4	4	3	2	1	2	2	2.67	2.67	3.33	3.00
7-VS	3	2	3	3	4	4	4	3	2	2	3	3	3.00	2.67	3.33	3.00
8-HS	3	3	4	4	4	4	4	3	1	1	1	1	2.67	2.67	3.00	2.67
8-VS	3	2	4	4	4	3	4	4	2	2	3	2	3.00	2.33	3.67	3.33

Tabla 3

Tejido #	Patólogo 1			Patólogo 2			Patólogo 3			Promedio		
	ND	CD	SQ	ND	CD	SQ	ND	CD	SQ	ND	CD	SQ
1 - HS	3	3	3	2	2	4	2	2	2	2.33	2.33	3.00
1 - VS	2	3	3	3	3	4	3	3	3	2.67	3.00	3.33
2 - HS	2	4	4	3	3	2	1	1	2	2.00	2.67	2.67
2 - VS	2	3	4	3	3	3	1	2	3	2.00	2.67	3.33
3 - HS	2	3	3	3	3	2	2	3	4	2.33	3.00	3.00
3 - VS	2	3	3	3	3	3	1	2	3	2.00	2.67	3.00
4 - HS	3	3	3	2	2	2	1	2	3	2.00	2.33	2.67
4 - VS	3	3	3	2	2	3	1	2	2	2.00	2.33	2.67
5 - HS	3	3	2	3	3	1	3	3	3	3.00	3.00	2.00
5 - vs	3	3	2	4	3	4	3	3	4	3.33	3.00	3.33
6 - HS	2	3	3	3	3	1	2	2	2	2.33	2.67	2.00
6 - VS	2	2	3	2	2	2	2	2	2	2.00	2.00	2.33
7 - HS	3	3	2	3	2	2	3	3	3	3.00	2.67	2.33
7 - VS	3	3	2	4	3	1	3	2	3	3.33	2.67	2.00

10

Cuantificación de la calidad de imagen de salida de red

[0025] A continuación, más allá de la comparación visual proporcionada en las FIGS. 3A-3H, 4A-4H, 5A-5P, los resultados de la red neuronal profunda entrenada 10 se cuantificaron calculando en primer lugar las diferencias de nivel de píxel entre las imágenes de campo claro 48 de las muestras teñidas químicamente 22 y las imágenes teñidas digital/virtualmente 40 que se sintetizan usando la red neuronal profunda 10 sin el uso de ningún marcador/tinción. La Tabla 4 a continuación resume esta comparación para diferentes combinaciones de tipos de tejido y tinciones, usando el espacio de color YCbCr, donde los componentes de croma Cb y Cr definen completamente el color, e Y define el componente de brillo de la imagen. Los resultados de esta comparación revelan que la diferencia media entre estos dos conjuntos de imágenes es $< -5\%$ y $< -16\%$, para los canales de croma (Cb, Cr) y de brillo (Y), respectivamente. A continuación, se usó una segunda métrica para cuantificar aún más la comparación, es decir, el índice de similitud estructural (SSIM), que en general se usa para predecir la puntuación que un observador humano dará para una imagen, en comparación con una imagen de referencia (Ecuación 8 en el presente documento). El SSIM oscila entre 0 y 1, donde 1 define la puntuación para imágenes idénticas. Los resultados de esta cuantificación SSIM también se resumen en la Tabla 4, que ilustra muy bien la fuerte similitud estructural entre las imágenes de salida de red 40 y las imágenes de campo claro 48 de las muestras teñidas químicamente.

Tabla 4

Tinción histológica virtual usando una red profunda	Número de imágenes de prueba	SSIM		Diferencia de Y (%)		Diferencia de Cb (%)		Diferencia de Cr (%)	
		media	std	media	std	media	std	media	std
Glándula salival (H&E)	10	0.826	0.059	11.5	9.0	2.5	2.4	2.5	2.5
Tiroides (H&E)	30	0.789	0.044	10.1	7.9	3.4	2.7	2.8	2.7
Tiroides (H&E, aprendizaje de transferencia)	30	0.839	0.029	14.0	8.4	2.4	2.2	2.6	2.6
Hígado (tricrómico de Masson)	30	0.847	0.023	11.0	8.9	3.1	2.7	4.0	3.5
Pulmón (tricrómico de Masson)	48	0.776	0.039	15.9	11.7	4.0	3.6	5.3	4.9
Riñón (tinción de Jones)	30	0.841	0.021	16.1	10.4	2.5	2.2	3.6	3.4

[0026] Se debe observar que las imágenes de campo claro 48 de las muestras de tejido teñidas químicamente 22 no proporcionan el verdadero estándar de referencia para este análisis específico de SSIM y YCbCr de las imágenes de salida de la red 40, porque hay variaciones incontroladas y cambios estructurales que experimenta el tejido durante el proceso de tinción histoquímica y pasos de deshidratación y aclaramiento relacionados. Otra variación que se observó para algunas de las imágenes fue que el *software* de escaneo automático de microscopio seleccionó diferentes planos de autoenfoco para las dos modalidades de obtención de imágenes. Todas estas variaciones crean algunos desafíos para la comparación cuantitativa absoluta de los dos conjuntos de imágenes (es decir, la salida de red 40 para un tejido sin marcador frente a la imagen de campo claro 48 del mismo tejido después del proceso de tinción histológica).

Estandarización de la tinción

[0027] Un subproducto interesante del sistema de tinción digital/virtual 2 puede ser la estandarización de la tinción. En otras palabras, la red neural profunda entrenada 10 converge a un esquema de coloración de "tinción común" por lo que la variación en las imágenes de tejido histológicamente teñidas 48 es mayor que la de las imágenes de tejido virtualmente teñidas 40. La coloración de la tinción virtual es únicamente el resultado de su entrenamiento (es decir, la tinción histológica estándar de referencia usada durante la fase de entrenamiento) y puede ajustarse adicionalmente basándose en las preferencias de los patólogos, reentrenando la red con una nueva coloración de tinción. Tal entrenamiento "mejorado" puede crearse desde cero o acelerarse a través del aprendizaje por transferencia. Esta estandarización potencial de la tinción usando aprendizaje profundo puede remediar los efectos negativos de las variaciones de humano a humano en diferentes etapas de la preparación de muestras, crear un terreno común entre diferentes laboratorios clínicos, potenciar el flujo de trabajo de diagnóstico para médicos clínicos así como ayudar al desarrollo de nuevos algoritmos tales como detección automática de metástasis tisular o clasificación de diferentes tipos de cáncer, entre otros.

Transferencia del aprendizaje a otras combinaciones de tinción de tejidos

[0028] Usando el concepto de aprendizaje por transferencia, el procedimiento de entrenamiento para nuevos tipos de tejido y/o tinciones puede converger mucho más rápido, al tiempo que se alcanza un rendimiento mejorado, es decir, un mejor mínimo local en la función de coste/pérdida de entrenamiento. Esto significa que una red neuronal profunda 10 del modelo CNN preaprendida, de una combinación diferente de tinción de tejido, puede usarse para inicializar la red neuronal profunda 10 para aprender estadísticamente la tinción virtual de una

nueva combinación. Las figuras 6A-6C muestran los atributos favorables de dicho enfoque: se entrenó una nueva red neural profunda 10 para teñir virtualmente las imágenes de autofluorescencia 20 secciones de tejido tiroideo sin teñir, y se inicializó usando los pesos y sesgos de otra red neuronal profunda 10 que se adiestró previamente para la tinción virtual con H&E de la glándula salival. La evolución de la métrica de pérdida en función del número de iteraciones utilizadas en la fase de entrenamiento demuestra claramente que la nueva red profunda tiroidea 10 converge rápidamente a un mínimo inferior en comparación con la misma arquitectura de red que se entrenó desde cero, utilizando inicialización aleatoria como se ve en la figura 6A. La figura 6B compara las imágenes de salida 40 de esta red tiroidea 10 en diferentes etapas de su proceso de aprendizaje, lo que ilustra además el impacto del aprendizaje por transferencia para adaptar rápidamente el enfoque presentado a nuevas combinaciones de tejido/tinción. Las imágenes de salida de red 40, después de la fase de entrenamiento con, por ejemplo, $\geq 6,000$ iteraciones, revelan que los núcleos celulares muestran contornos irregulares, surcos nucleares y palidez de la cromatina, sugestivos de carcinoma papilar de tiroides; las células también muestran cantidades leves a moderadas de citoplasma granular eosinofílico y el núcleo fibrovascular en la imagen de salida de red muestra un aumento de células inflamatorias, incluyendo linfocitos y células plasmáticas. La figura 6C ilustra la correspondiente imagen 48 de campo claro teñida químicamente con H&E.

Uso de múltiples canales fluorescentes a diferentes resoluciones

[0029] El método de uso de la red neuronal profunda entrenada 10 puede combinarse con otras longitudes de onda de excitación y/o modalidades de obtención de imágenes para mejorar su rendimiento de inferencia para diferentes constituyentes de tejido. Por ejemplo, se intentó la detección de melanina en una muestra de sección de tejido cutáneo usando tinción de H&E virtual. Sin embargo, la melanina no se identificó claramente en la salida de la red, ya que presenta una señal autofluorescente débil a longitudes de onda de excitación/emisión DAPI medidas en el sistema experimental descrito en el presente documento. Un método potencial para aumentar la autofluorescencia de melanina es obtener imágenes de las muestras mientras están en una solución oxidante. Sin embargo, se usó una alternativa más práctica en la que se empleó un canal de autofluorescencia adicional, que se originó a partir de, por ejemplo, un filtro Cy5 (excitación 628 nm/emisión 692 nm) de manera que la señal de melanina puede potenciarse e inferirse con precisión en la red neuronal profunda entrenada 10. Entrenando la red 10 usando tanto los canales de autofluorescencia DAPI como Cy5, la red neuronal profunda entrenada 10 fue capaz de determinar con éxito dónde se produce melanina en la muestra, como se ilustra en las FIGS. 7A-7C. Por el contrario, cuando se utiliza solamente el canal DAPI (figura 7A), la red 10 no puede determinar las zonas que contienen melanina (las zonas aparecen blancas). Dicho de otro modo, la red 10 usó la información de autofluorescencia adicional del canal Cy5 para distinguir la melanina del tejido de fondo. Para los resultados que se muestran en las FIGS. 7A-7C, las imágenes 20 se adquirieron usando una lente objetivo de menor resolución ($10\times/0,45NA$) para el canal Cy5, para complementar la exploración DAPI de alta resolución ($20\times/0,75NA$), ya que se planteó la hipótesis de que la información más necesaria se encuentra en la exploración DAPI de alta resolución y la información adicional (por ejemplo, la presencia de melanina) puede codificarse con la exploración de menor resolución. De esta manera, se usaron dos canales diferentes, usándose uno de los canales a una resolución más baja para identificar la melanina. Esto puede requerir múltiples pasadas de exploración de la muestra 22 con el microscopio fluorescente 110. En otra forma de realización multicanal más, se pueden alimentar múltiples imágenes 20 a la red neuronal profunda entrenada 10. Esto puede incluir, por ejemplo, imágenes fluorescentes sin procesar en combinación con una o más imágenes que han experimentado un preprocesamiento de imágenes lineal o no lineal, tal como mejora de contraste, inversión de contraste y filtrado de imágenes.

[0030] El sistema 2 y los métodos descritos en el presente documento muestran la capacidad de teñir digital/virtualmente secciones de tejido libres de marcadores 22, usando una técnica de aprendizaje profundo supervisada que usa una única imagen de fluorescencia 20 de la muestra como entrada, capturada por un microscopio de fluorescencia estándar 110 y un conjunto de filtros (en otras formas de realización, se introducen múltiples imágenes de fluorescencia 20 cuando se usan múltiples canales de fluorescencia). Este método estadístico basado en aprendizaje tiene el potencial de reestructurar el flujo de trabajo clínico en histopatología y puede beneficiarse de diversas modalidades de obtención de imágenes tales como microscopía de fluorescencia, microscopía no lineal, microscopía holográfica, microscopía de dispersión Raman estimulada y tomografía de coherencia óptica, entre otras, para proporcionar potencialmente una alternativa digital a la práctica estándar de tinción histoquímica de muestras de tejido 22. Aquí, el método se demostró usando muestras de tejido no teñidas fijas 22 para proporcionar una comparación significativa con muestras de tejido teñidas químicamente, que es esencial para entrenar la red neuronal profunda 10 así como para probar ciegamente el rendimiento de la salida de red frente al método clínicamente aprobado. Sin embargo, el enfoque basado en aprendizaje profundo presentado es ampliamente aplicable a diferentes tipos y estados de una muestra 22 incluyendo muestras de tejido frescas no seccionadas (por ejemplo, después de un procedimiento de biopsia) sin el uso de ningún marcador o tinción. Después de su entrenamiento, la red neuronal profunda 10 puede usarse para teñir digital/virtualmente las imágenes de muestras de tejido fresco libres de marcadores 22, adquiridas usando, por ejemplo, excitación UV o UV profunda o incluso modalidades de microscopía no lineal. Por ejemplo, la microscopía Raman puede proporcionar firmas bioquímicas libres de marcadores muy ricas que pueden mejorar aún más la eficacia de la tinción virtual que aprende la red neuronal.

[0031] Una parte importante del proceso de entrenamiento implica hacer coincidir las imágenes de fluorescencia 20 de las muestras de tejido sin marcador 22 y sus correspondientes imágenes de campo claro 48 después del proceso de tinción histoquímica (es decir, imágenes teñidas químicamente). Debe observarse que durante el proceso de tinción y los pasos relacionados, algunos elementos constitutivos del tejido pueden perderse o deformarse de una manera que induzcan a error a la función de pérdida/coste en la fase de entrenamiento. Sin embargo, esto es solo un desafío relacionado con el entrenamiento y la validación y no plantea ninguna limitación en la práctica de una red neuronal profunda bien entrenada 10 para la tinción virtual de muestras de tejido libres de marcadores 22. Para garantizar la calidad de la fase de entrenamiento y validación y minimizar el impacto de este desafío en el rendimiento de la red, se estableció un umbral para un valor de correlación aceptable entre los dos conjuntos de imágenes (es decir, antes y después del proceso de tinción histoquímica) y se eliminaron los pares de imágenes no coincidentes del conjunto de entrenamiento/validación para asegurarse de que la red neuronal profunda 10 aprende la señal real, no las perturbaciones a la morfología del tejido debido al proceso de tinción química. De hecho, este proceso de limpieza de los datos de imagen de entrenamiento/validación puede realizarse iterativamente: puede comenzarse con una eliminación aproximada de las muestras obviamente alteradas y, por consiguiente, converger en una red neuronal 10 que está entrenada. Después de esta fase de entrenamiento inicial, las imágenes de salida 40 de cada muestra en el conjunto de imágenes disponible pueden compararse con sus imágenes de campo claro 48 correspondientes para establecer un umbral más refinado para rechazar algunas imágenes adicionales y limpiar adicionalmente el conjunto de imágenes de entrenamiento/validación. Con unas pocas iteraciones de este proceso, se puede, no solo refinar aún más el conjunto de imágenes, sino también mejorar el rendimiento de la red neuronal profunda 10 entrenada final.

[0032] La metodología descrita anteriormente mitigará algunos de los desafíos de entrenamiento debido a la pérdida aleatoria de algunas características tisulares después del proceso de tinción histológica. De hecho, esto destaca otra motivación para omitir los procedimientos laboriosos y costosos que están implicados en la tinción histoquímica ya que será más fácil conservar la histología tisular local en un método sin marcadores, sin la necesidad de que un experto maneje algunos de los procedimientos delicados del proceso de tinción, que a veces también requiere observar el tejido bajo un microscopio.

[0033] Usando un PC de sobremesa, la fase de entrenamiento de la red neuronal profunda 10 lleva una cantidad de tiempo considerable (por ejemplo, ~13 horas para la red de glándulas salivales). Sin embargo, todo este proceso puede acelerarse significativamente usando *hardware* informático dedicado, basado en GPU. Además, como ya se ha destacado en las figuras 6A-6C, el aprendizaje por transferencia proporciona un "inicio en caliente" a la fase de entrenamiento de una nueva combinación de tejido/tinción, haciendo que todo el proceso sea significativamente más rápido. Una vez que la red neuronal profunda 10 ha sido entrenada, la tinción digital/virtual de una imagen de muestra 40 se realiza de una manera única, no iterativa, que no requiere un enfoque de prueba y error o cualquier sintonización de parámetro para lograr el resultado óptimo. Basándose en su arquitectura de alimentación directa y no iterativa, la red neuronal profunda 10 emite rápidamente una imagen teñida virtualmente en menos de un segundo (por ejemplo, 0,59 segundos, correspondiente a un campo de visión de muestra de ~ 0,33 mm × 0,33 mm). Con una aceleración basada en GPU adicional, tiene el potencial de lograr un rendimiento en tiempo real o casi en tiempo real en la salida de imágenes teñidas digital/virtualmente 40 que podrían ser especialmente útiles en el quirófano o para aplicaciones de obtención de imágenes *in vivo*.

[0034] El procedimiento de tinción digital/virtual que se implementa se basa en entrenar una red neuronal profunda 10 de CNN separada para cada combinación de tejido/tinción. Si se alimenta una red neuronal profunda 10 basada en CNN con las imágenes de autofluorescencia 20 que tienen diferentes combinaciones de tejido/tinción, no se comportará como se desea. Esto, sin embargo, no es una limitación porque para aplicaciones histológicas, el tipo de tejido y el tipo de tinción están predeterminados para cada muestra 22 de interés y, por lo tanto, una selección de CNN específica para crear la imagen teñida digital/virtualmente 40 a partir de una imagen de autofluorescencia 20 de la muestra 22 no marcada no requiere una información o recurso adicional. Por supuesto, se puede aprender un modelo de CNN más general para múltiples combinaciones de tejido/tinción, por ejemplo, aumentando el número de parámetros entrenados en el modelo, a costa de un posible aumento en los tiempos de entrenamiento e inferencia. Otra vía es el potencial del sistema 2 y el método para realizar múltiples tinciones virtuales en el mismo tipo de tejido no marcado.

[0035] Una ventaja significativa del sistema 2 es que es bastante flexible. Puede adaptarse a la retroalimentación para modificar estadísticamente su rendimiento si se detecta un fallo de diagnóstico a través de una comparación clínica, penalizando en consecuencia tales fallos a medida que se detectan. Este entrenamiento iterativo y el ciclo de aprendizaje de transferencia, basado en evaluaciones clínicas del rendimiento de la salida de la red, ayudará a optimizar la robustez y el impacto clínico del enfoque presentado. Finalmente, este método y sistema 2 se pueden usar para microguiar el análisis molecular a nivel de tejido sin teñir, identificando localmente regiones de interés basándose en tinción virtual, y usando esta información para guiar el análisis posterior del tejido para, por ejemplo, microinmunohistoquímica o secuenciación. Este tipo de microguiado virtual en una muestra de tejido no marcado puede facilitar la identificación de alto rendimiento de subtipos de enfermedades, ayudando también al desarrollo de terapias personalizadas para los pacientes.

Preparación de la muestra

[0036] Se desparafinaron secciones de tejido de 2 μm de espesor embebidas en parafina y fijadas con formalina usando xileno y se montaron en un portaobjetos de vidrio convencional usando Cytoseal™ (Thermo-Fisher Scientific, Waltham, MA EE.UU.), seguido de la colocación de un cubreobjetos (Fisherfinest™, 24x50-1, Fisher Scientific, Pittsburgh, PA EE.UU.). Después del proceso de formación de imágenes de autofluorescencia inicial (usando un conjunto de filtros de excitación y emisión DAPI) de la muestra de tejido no marcado, el portaobjetos se puso entonces en xileno durante aproximadamente 48 horas o hasta que el cubreobjetos pudo retirarse sin dañar el tejido. Una vez retirado el cubreobjetos, se sumergió el portaobjetos (aproximadamente 30 inmersiones) en alcohol absoluto, alcohol al 95 % y luego se lavó en agua desionizada durante ~1 minuto. Este paso fue seguido por los procedimientos de tinción correspondientes, usados para las tinciones de H&E, tricrómico de Masson o Jones. Esta ruta de procesamiento de tejido solo se usa para el entrenamiento y validación del enfoque y no se necesita después de que la red se haya entrenado. Para ensayar el sistema y el método, se usaron diferentes combinaciones de tejido y tinción: las secciones de tejido de glándula salival y tiroides se tiñeron con H&E, las secciones de tejido renal se tiñeron con la tinción de Jones, mientras que las secciones de tejido hepático y pulmonar se tiñeron con el tricrómico de Masson.

[0037] En el estudio WSI, las secciones de tejido de 2-4 μm de espesor de FFPE no se cubrieron con portaobjetos durante la etapa de formación de imágenes de autofluorescencia. Después de la formación de imágenes de autofluorescencia, las muestras de tejido se tiñeron histológicamente como se ha descrito anteriormente (tricrómico de Masson para el hígado y Jones para las secciones de tejido de riñón). Las muestras congeladas sin teñir se prepararon embebiendo la sección de tejido en O.C.T. (Tissue Tek, SAKURA FINETEK USA INC) y se sumergieron en 2-metilbutano con hielo seco. La sección congelada se cortó a continuación en secciones de 4 μm y se puso en un congelador hasta que se obtuvieron imágenes. Después del proceso de obtención de imágenes, la sección de tejido se lavó con alcohol al 70 %, se tiñó con H&E y se cubrió con cubreobjetos. Las muestras se obtuvieron del Translational Pathology Core Laboratory (TPCL) y fueron preparadas por el Laboratorio de Histología de la UCLA. Las secciones de tejido renal de pacientes diabéticos y no diabéticos se obtuvieron bajo IRB 18-001029 (UCLA). Todas las muestras se obtuvieron después de la desidentificación de la información relacionada con el paciente, y se prepararon a partir de especímenes existentes. Por lo tanto, este trabajo no interfirió con las prácticas estándar de cuidados o los procedimientos de recogida de muestras.

Adquisición de datos

[0038] Las imágenes de autofluorescencia de tejido sin marcador 20 se capturaron usando un microscopio de fluorescencia convencional 110 (IX83, Olympus Corporation, Tokio, Japón) equipado con una platina motorizada, donde el proceso de adquisición de imágenes se controló mediante MetaMorph® software de automatización de microscopio (Molecular Devices, LLC). Las muestras de tejido sin teñir se excitaron con luz UV cercana y se tomaron imágenes usando un cubo de filtro DAPI (OSF13-DAPI-5060C, longitud de onda de excitación 377 nm/50 nm de ancho de banda, longitud de onda de emisión 447 nm/60 nm de ancho de banda) con una lente objetivo 40x/0,95NA (Olympus UPLSAPO 40x2/0,95NA, WD0,18) o una lente objetivo 20x/0,75NA (Olympus UPLSAPO 20x/0,75NA, WD0,65). Para la inferencia de la melanina, las imágenes de autofluorescencia de las muestras se adquirieron adicionalmente usando un cubo de filtro Cy5 (CY5-4040C-OFX, longitud de onda de excitación 628 nm/40 nm de ancho de banda, longitud de onda de emisión 692 nm/40 nm de ancho de banda) con una lente objetivo 10x/0,4NA (Olympus UPLSAPO10x2). Cada imagen de autofluorescencia se capturó con un sensor científico CMOS (ORCA-flash 4.0 v2, Hamamatsu Photonics K.K., Prefectura de Shizuoka, Japón) con un tiempo de exposición de ~500 ms. Las imágenes de campo claro 48 (usadas para el entrenamiento y validación) se adquirieron usando un microscopio de escáner de portaobjetos (Aperio AT, Leica Biosystems) usando un objetivo de 20x/0,75NA (Plan Apo), equipado con un adaptador de aumento de 2x.

Preprocesamiento y alineación de imágenes

[0039] Dado que la red neuronal profunda 10 tiene como objetivo aprender una transformación estadística entre una imagen de autofluorescencia 20 de una muestra de tejido sin teñir químicamente 22 y una imagen de campo claro 48 de la misma muestra de tejido 22 después de la tinción histoquímica, es importante hacer coincidir con precisión el FOV de las imágenes de entrada y de destino (es decir, imagen de autofluorescencia sin teñir 20 y la imagen de campo claro teñida 48). Un esquema global que describe el proceso de registro de imagen global y local se describe en la figura 8 que se implementó en MATLAB (The MathWorks Inc., Natick, MA, EE.UU.). El primer paso en este proceso es encontrar características candidatas para hacer coincidir imágenes de autofluorescencia sin teñir e imágenes de campo claro teñidas químicamente. Para esto, cada imagen de autofluorescencia 20 (2048x2048 píxeles) se reduce para coincidir con el tamaño de píxel efectivo de las imágenes de microscopio de campo claro. Esto da como resultado una imagen de tejido autofluorescente sin teñir de 1351x1351 píxeles, que mejora en contraste saturando el 1 % inferior y el 1 % superior de todos los valores de píxel, y se invierte el contraste (imagen 20a en la figura 8) para representar mejor el mapa de colores de la imagen de portaobjetos completo convertida en escala de grises. A continuación, se realiza un proceso de

parches de correlación 60 en el que se calcula una matriz de puntuación de correlación normalizada correlacionando cada uno de los parches de 1351×1351 píxeles con el parche correspondiente del mismo tamaño, extraído de la imagen de escala de grises de portaobjetos completo 48a. La entrada en esta matriz con la puntuación más alta representa el FOV más probablemente coincidente entre las dos modalidades de obtención de imágenes. Usando esta información (que define un par de coordenadas), el FOV coincidente de la imagen 48 de campo claro de portaobjetos completo original se recorta 48c para crear imágenes de destino 48d. Siguiendo este procedimiento de coincidencia de FOV 60, la autofluorescencia 20 y las imágenes de microscopio de campo claro 48 coinciden en gran medida. Sin embargo, todavía no están registrados con precisión a nivel de píxeles individuales, debido al ligero desajuste en la colocación de la muestra en los dos experimentos de formación de imagen microscópica diferentes (autofluorescencia, seguido de campo claro), que provoca aleatoriamente un ligero ángulo de rotación (por ejemplo, -1-2 grados) entre las imágenes de entrada y de destino de la misma muestra.

[0040] La segunda parte del proceso de coincidencia de entrada-destino implica un paso de registro global 64, que corrige este ángulo de rotación ligero entre las imágenes de autofluorescencia y de campo claro. Esto se realiza extrayendo vectores de características (descriptores) y sus ubicaciones correspondientes de los pares de imágenes, y haciendo coincidir las características usando los descriptores extraídos. A continuación, se encuentra una matriz de transformación correspondiente a los pares coincidentes usando el algoritmo de Consenso de Muestra del estimador de M (MSAC), que es una variante del algoritmo de Consenso de Muestra Aleatorio (RANSAC). Finalmente, la imagen corregida en ángulo 48e se obtiene aplicando esta matriz de transformación al parche original de imagen de microscopio de campo claro 48d. Después de la aplicación de esta rotación, las imágenes 20b, 48e se recortan aún más en 100 píxeles (50 píxeles en cada lado) para adaptarse a valores de píxeles indefinidos en los bordes de la imagen, debido a la corrección del ángulo de rotación.

[0041] Finalmente, para la operación de registro de características locales 68, un registro de imágenes elástico, que coincide con las características locales de ambos conjuntos de imágenes (autofluorescencia 20b frente a campo claro 48e), haciendo coincidir jerárquicamente los bloques correspondientes, de grande a pequeño. Se utiliza una red neuronal 71 para aprender la transformación entre las imágenes aproximadamente coincidentes. Esta red 71 utiliza la misma estructura que la red 10 de la figura 10. Se usa un bajo número de iteraciones de modo que la red 71 solo aprenda el mapeo de color preciso, y no cualquier transformación espacial entre las imágenes de entrada y de marcador. El mapa de transformación calculado de este paso se aplica finalmente a cada parche de imagen de campo claro 48e. Al final de estos pasos de registro 60, 64, 68, los parches de imagen de autofluorescencia 20b y sus parches de imagen de tejido de campo claro correspondientes 48f se emparejan con precisión entre sí y pueden usarse como pares de entrada y marcador para el entrenamiento de la red neuronal profunda 10, permitiendo que la red se enfoque únicamente en y aprenda el problema de la tinción histológica virtual.

[0042] Para las imágenes de lente de objetivo de $20\times$ (que se usaron para generar los datos de la Tabla 2 y la Tabla 3) se usó un proceso similar. En lugar de reducir las imágenes de autofluorescencia 20, las imágenes de microscopio de campo claro 48 se redujeron hasta el 75,85 % de su tamaño original de modo que coincidieran con las imágenes de menor aumento. Además, para crear imágenes de portaobjetos completos usando estas imágenes de $20\times$, se aplicaron técnicas de corrección y normalización de sombreado adicionales. Antes de ser introducidos en la red 71, cada campo de visión se normalizó restando el valor medio a través de todo el portaobjetos y dividiéndolo por la desviación estándar entre valores de píxeles. Esto normaliza la entrada de red tanto dentro de cada portaobjetos como entre portaobjetos. Finalmente, se aplicó corrección de sombreado a cada imagen para tener en cuenta la menor intensidad relativa medida en los bordes de cada campo de visión.

Arquitectura y entrenamiento de redes neuronales profundas

[0043] En este trabajo, se usó una arquitectura GAN para aprender la transformación de una imagen de entrada de autofluorescencia sin tinción sin marcador 20 a la imagen de campo claro correspondiente 48 de la muestra teñida químicamente. Un entrenamiento estándar basado en red neuronal convolucional aprende a minimizar una función de pérdida/coste entre la salida de la red y el marcador de destino. Por lo tanto, la elección de esta función de pérdida 69 (figuras 9 y 10) es un componente crítico del diseño de red profunda. Por ejemplo, simplemente eligiendo una penalización de norma- ℓ_2 como una función de coste tenderá a generar resultados borrosos, ya que la red promedia una probabilidad ponderada de todos los resultados plausibles; por lo tanto, generalmente se necesitan términos de regularización adicionales para guiar a la red para preservar las características de muestra pronunciada deseadas a la salida de la red. Las GAN evitan este problema al aprender un criterio que tiene como objetivo clasificar con precisión si la imagen de salida de la red profunda es real o falsa (es decir, correcta en su tinción virtual o incorrecta). Esto hace que las imágenes de salida que son inconsistentes con los marcadores deseados no sean toleradas, lo que hace que la función de pérdida sea adaptativa a los datos y a la tarea deseada en cuestión. Para lograr este objetivo, el procedimiento de entrenamiento de GAN implica el entrenamiento de dos redes diferentes, tal como se muestra en las figuras 9 y 10: (i) una red del generador 70, que en este caso tiene como objetivo aprender la transformación estadística entre las imágenes de entrada de autofluorescencia sin tinte 20 y las imágenes de campo claro 48

correspondientes de las mismas muestras 12, después del proceso de tinción histológica; y (ii) una red del discriminador 74 que aprende a cómo discriminar entre una imagen de campo claro real de una sección de tejido teñido y la imagen de salida de la red del generador. Finalmente, el resultado deseado de este proceso de entrenamiento es una red neuronal profunda entrenada 10, que transforma una imagen de entrada de autofluorescencia sin teñir 20 en una imagen teñida digitalmente 40 que será indistinguible de la imagen de campo claro teñida 48 de la misma muestra 22. Para esta tarea, las funciones de pérdida 69 del generador 70 y del discriminador 74 se definieron como tales:

$$\begin{aligned} \ell_{\text{generador}} &= \text{MSE}\{z_{\text{marcador}}, z_{\text{salida}}\} + \lambda \times \text{TV}\{z_{\text{salida}}\} + \alpha \times (1 - D(z_{\text{salida}}))^2 \\ \ell_{\text{discriminador}} &= D(z_{\text{salida}})^2 + (1 - D(z_{\text{marcador}}))^2 \end{aligned} \quad (1)$$

donde D se refiere a la salida de la red del discriminador, z_{marcador} denota la imagen de campo claro del tejido teñido químicamente, z_{salida} denota la salida de la red del generador. La función de pérdida del generador equilibra el error cuadrático medio (MSE) a nivel de píxeles de la imagen de salida de la red del generador con respecto a su marcador, el operador de variación total (TV) de la imagen de salida, y la predicción de la red del discriminador de la imagen de salida, usando los parámetros de regularización (λ, α) que se establecen empíricamente a diferentes valores, que se adaptan para ~2 % y -20 % de la pérdida de MSE por píxeles y la pérdida del generador combinada ($\ell_{\text{generador}}$), respectivamente. El operador de TV de una imagen z se define como:

$$\text{TV}(z) = \sum_p \sum_q \sqrt{(z_{p-1,q} - z_{p,q})^2 + (z_{p,q-1} - z_{p,q})^2} \quad (2)$$

donde p, q son índices de píxeles. Basado en la Ec. (1), el discriminador intenta minimizar la pérdida de salida, mientras maximiza la probabilidad de clasificar correctamente el marcador real (es decir, la imagen de campo claro del tejido teñido químicamente). Idealmente, la red del discriminador tendría como objetivo conseguir $D(z_{\text{marcador}}) = 1$ y $D(z_{\text{salida}}) = 0$, pero si el generador es entrenado con éxito por la GAN, $D(z_{\text{salida}})$ idealmente convergerá a 0,5.

[0044] La arquitectura de red neuronal profunda de generador 70 se detalla en la figura 10. La red 70 procesa una imagen de entrada 20 de una manera multiescala, usando trayectorias de muestreo descendente y ascendente, lo que ayuda a la red a aprender la tarea de tinción virtual a diversas escalas diferentes. La trayectoria de muestreo descendente consiste en cuatro pasos individuales (cuatro bloques #1, #2, #3, #4), conteniendo cada paso un bloque residual, cada uno de los cuales mapea un mapa de características x_k en el mapa de características x_{k+1} :

$$x_{k+1} = x_k + \text{LReLU}\left[\text{CONV}_{k3}\left\{\text{LReLU}\left[\text{CONV}_{k2}\left\{\text{LReLU}\left[\text{CONV}_{k1}\{x_k\}\right]\right\}\right]\right\}\right] \quad (3)$$

donde CONV{.} es el operador de convolución (que incluye los términos de sesgo), $k1$, $k2$, y $k3$ denotan el número de serie de las capas convolucionales, y LReLU[.] es la función de activación no lineal (es decir, una unidad lineal rectificada con fugas) que se usó en toda la red, definida como:

$$\text{LReLU}(x) = \begin{cases} x & \text{para } x > 0 \\ 0.1x & \text{de otro modo} \end{cases} \quad (4)$$

[0045] El número de los canales de entrada para cada nivel en la trayectoria de muestreo descendente se estableció en: 1, 64, 128, 256, mientras que el número de los canales de salida en la trayectoria de muestreo descendente se estableció en: 64, 128, 256, 512. Para evitar la falta de coincidencia de dimensiones para cada bloque, el mapa de características x_k se rellenó con cero para que coincidiera con el número de canales en x_{k+1} . La conexión entre cada nivel de muestreo descendente es una capa de agrupamiento promedio de 2x2 con un intervalo de 2 píxeles que reducen los mapas de características en un factor de 4 (2 veces para cada dirección). Después de la salida del cuarto bloque de muestreo descendente, otra capa convolucional (CL) mantiene el número de mapas de características como 512, antes de conectarlo a la trayectoria de muestreo ascendente. La

trayectoria de muestreo ascendente consiste en cuatro pasos de muestreo ascendente simétricos (#1, #2, #3, #4), conteniendo cada paso un bloque convolucional. La operación de bloque convolucional, que mapea el mapa de características y_k en el mapa de características y_{k+1} , se da por:

$$y_{k+1} = \text{LReLU} \left[\text{CONV}_{k6} \left\{ \text{LReLU} \left[\text{CONV}_{k5} \left\{ \text{LReLU} \left[\text{CONV}_{k4} \left\{ \text{CONCAT}(x_{k+15}, \text{US}\{y_k\}) \right\} \right] \right\} \right] \right\} \right] \quad (5)$$

donde $\text{CONCAT}(\cdot)$ es la concatenación entre dos mapas de características que combina el número de canales, $\text{US}\{\cdot\}$ es el operador de muestreo ascendente, $k4$, $k5$ y $k6$ denotan el número de serie de las capas de convolución. El número de los canales de entrada para cada nivel en la trayectoria de muestreo ascendente se estableció en 1024, 512, 256, 128 y el número de los canales de salida para cada nivel en la trayectoria de muestreo ascendente se estableció en 256, 128, 64, 32, respectivamente. La última capa es una capa convolucional (CL) que mapea 32 canales en 3 canales, representados por el mapa de color YCbCr. Tanto las redes del generador como del discriminador fueron entrenadas con un tamaño de parche de 256×256 píxeles.

[0046] La red del discriminador, resumida en la figura 10, recibe tres (3) canales de entrada, correspondientes al espacio de color YCbCr de una imagen de entrada 40YCbCr, 48YCbCr. Esta entrada se transforma entonces en una representación de 64 canales utilizando una capa convolucional, que va seguida de 5 bloques del siguiente operador:

$$z_{k+1} = \text{LReLU} \left[\text{CONV}_{k2} \left\{ \text{LReLU} \left[\text{CONV}_{k1} \{z_k\} \right] \right\} \right] \quad (6)$$

donde $k1$, $k2$ denotan el número de serie de la capa convolucional. El número de canales para cada capa fue 3, 64, 64, 128, 128, 256, 256, 512, 512, 1024, 1024, 2048. La siguiente capa fue una capa de agrupamiento promedio con un tamaño de filtro que es igual al tamaño de parche (256×256), lo que da como resultado un vector con 2048 entradas. La salida de esta capa de agrupamiento promedio se introduce entonces a dos capas completamente conectadas (FC) con la siguiente estructura:

$$z_{k+1} = \text{FC} \left[\text{LReLU} \left[\text{FC} \{z_k\} \right] \right] \quad (7)$$

donde FC representa la capa completamente conectada, con pesos y sesgos aprendibles. La primera capa completamente conectada genera un vector con 2048 entradas, mientras que la segunda genera un valor escalar. Este valor escalar se utiliza como una entrada a una función de activación sigmoidea $D(z) = 1 / (1 + \exp(-z))$ que calcula la probabilidad (entre 0 y 1) de que la entrada de red del discriminador sea real/genuina o falsa, es decir, idealmente $D(z_{\text{marcador}}) = 1$ como se ilustra por la salida 67 en la figura 10.

[0047] Los núcleos de convolución a lo largo de la GAN se establecieron en 3×3. Estos núcleos se inicializaron aleatoriamente usando una distribución normal truncada con una desviación estándar de 0,05 y una media de 0; todos los sesgos de red se inicializaron como 0. Los parámetros aprendibles se actualizan a través de la etapa de entrenamiento de la red neuronal profunda 10 mediante retropropagación (ilustrada en flechas discontinuas de la figura 10) usando un optimizador de estimación de momento adaptativo (Adam) con tasa de aprendizaje 1×10^{-4} para la red del generador 70 y 1×10^{-5} para la red del discriminador 74. Además, para cada iteración del discriminador 74, hubo 4 iteraciones de la red del generador 70, para evitar el estancamiento de entrenamiento después de un posible sobreajuste de la red del discriminador a los marcadores. Se usó un tamaño de lote de 10 en el entrenamiento.

[0048] Una vez que todos los campos de visión han pasado a través de la red 10, las imágenes de portaobjetos completos se unen usando el *plugin* de unión Fiji Grid/Collection (véase, por ejemplo, Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676-682 (2012), que se incorpora en el presente documento por referencia). Este *plugin* calcula la superposición exacta entre cada baldosa y las mezcla linealmente en una única imagen grande. En general, la inferencia y la costura tardaron ~5 minutos y 30 segundos, respectivamente, por cm^2 y puede mejorarse sustancialmente usando avances de *hardware* y *software*. Antes de mostrarse a los patólogos, se recortan secciones que están desenfocadas o tienen aberraciones importantes (debidas, por ejemplo, a partículas de polvo) en las imágenes de autofluorescencia o de campo claro. Finalmente, las imágenes se exportaron al formato Zoomify (diseñado para permitir la visualización de grandes imágenes usando un navegador web estándar; <http://zoomify.com/>) y se cargaron en el sitio web GIGAmacro (<https://viewer.gigamacro.com/>) para un fácil acceso y visualización por parte de los patólogos.

Detalles de implementación

- 5 [0049] Los demás detalles de implementación, incluyendo el número de parches entrenados, el número de épocas y los tiempos de entrenamiento se muestran en la Tabla 5 a continuación. La red neuronal profunda de tinción digital/virtual 10 se implementó usando Python versión 3.5.0. La GAN se implementó usando la versión 1.4.0 de la estructura TensorFlow. Otras bibliotecas de Python usadas fueron os, time, tqdm, la biblioteca de imágenes de Python (PIL), SciPy, glob, ops, sys y numpy. El *software* se implementó en un ordenador de escritorio con una CPU Core i7-7700K @ 4.2 GHz (Intel) y 64 GB de RAM, ejecutando un sistema operativo Windows 10 (Microsoft). El entrenamiento y pruebas de la red se realizaron usando doble GPU GeForce® GTX 1080Ti (NVIDIA).

Tabla 5

Red de tinción virtual	# de parches de entrenamiento	# de épocas	Tiempo de entrenamiento (horas)
Glándula salival (H&E)	2768	28	13.046
Tiroides (H&E)	8336	8	12.445
Tiroides (H&E, aprendizaje de transferencia)	8336	4	7.107
Hígado (tricrómico de Masson)	3840	26	18.384
Pulmón (tricrómico de Masson)	9162	10	16.602
Riñón (tinción de Jones)	4905	8	7.16
Hígado (tricrómico de Masson, WSI)	211475	3	39.64
Riñón (tinción de Jones, WSI)	59344	14	57.05
Ovario 1	4736	84	37.21
Ovario 2	11123	14	37.41
Glándula salival - 1	4417	65	24.61
Glándula salival - 2	2652	90	23.9
Glándula salival - 3	13262	24	30.58
Mama	67188	4	24.85
Piel	2568	124	27.02
Piel (DAPI+CY5)	2568	124	29.62
Próstata	677	472	30.27

- 15 [0050] Aunque se han mostrado y descrito formas de realización de la presente invención, se pueden realizar diversas modificaciones sin apartarse del alcance de la presente invención tal como se define en las reivindicaciones.

REIVINDICACIONES

1. Método para generar una imagen microscópica teñida digitalmente (40) de una muestra sin marcador (22) que comprende:

entrenar una red neuronal profunda (10) que se ejecuta mediante un *software* de procesamiento de imágenes (104) usando uno o más procesadores (102) de un dispositivo informático (100), donde la red neuronal profunda (10) se entrena con una pluralidad de pares coincidentes de imágenes teñidas químicamente o parches de imágenes y sus imágenes de autofluorescencia o parches de imágenes correspondientes de la misma muestra y donde los pares coincidentes se someten a un proceso de registro global (64) basado en vectores de características extraídos y sus ubicaciones correspondientes de los pares coincidentes y un proceso de registro local posterior (71, 68) que aprende, usando una red neuronal, el mapeo de color entre los pares coincidentes registrados globalmente seguido por un registro de características local (68) que coincide con las características locales de los pares coincidentes;

obtener una o más imágenes de autofluorescencia (20) de la muestra (22) usando un microscopio de fluorescencia (110) y una o más fuentes de luz de excitación y uno o más filtros de emisión, donde la luz fluorescente se emite desde fluoróforos endógenos u otros emisores endógenos de luz desplazada en frecuencia dentro de la muestra (22);

introducir una o más imágenes de autofluorescencia (20) de la muestra (22) en la red neuronal profunda entrenada (10); y

la red neuronal profunda entrenada (10) que produce la imagen microscópica teñida digitalmente (40) de la muestra (22) que es sustancialmente equivalente a una imagen de campo claro correspondiente (46) de la misma muestra (22) que se ha teñido químicamente (44).

2. Método según la reivindicación 1, donde la red neuronal profunda entrenada (10) comprende una red neuronal convolucional.

3. Método según la reivindicación 1, donde la red neuronal profunda se entrena usando un modelo de red generativa adversarial (GAN).

4. Método según la reivindicación 1, donde la red neuronal profunda se entrena usando una red del generador configurada para aprender la transformación estadística entre las imágenes o los parches de imágenes teñidas químicamente y de autofluorescencia coincidentes de la misma muestra y una red del discriminador configurada para discriminar entre una imagen real teñida químicamente de la muestra y la imagen de salida microscópica teñida digitalmente de la muestra.

5. Método según la reivindicación 1, donde la muestra comprende tejido de mamífero, tejido vegetal, células, patógenos, frotis de fluidos biológicos u otros objetos de interés.

6. Método según la reivindicación 1, donde la red neuronal profunda se entrena con muestras del mismo tipo que el tipo de muestra de la una o más imágenes de autofluorescencia obtenidas.

7. Método según la reivindicación 1, donde la una o más imágenes de autofluorescencia se someten a una o más operaciones de preprocesamiento de imágenes lineales o no lineales seleccionadas de entre mejora del contraste, inversión del contraste, filtrado de imágenes antes de introducirlas en la red neuronal profunda entrenada.

8. Método según la reivindicación 1, donde la red neuronal profunda entrenada se entrena y/o ejecuta usando una o más GPU o ASIC.

9. Método según la reivindicación 1, donde la muestra comprende una muestra *in vivo* y donde la imagen microscópica teñida digitalmente de la muestra se emite en tiempo real o casi en tiempo real después de obtener la una o más imágenes de autofluorescencia de la muestra.

10. Método según la reivindicación 1, donde la red neuronal profunda entrenada se entrena para una segunda combinación de tejido/tinción usando pesos de red neuronal iniciales y sesgos de una primera combinación de tejido/tinción que se optimizan adicionalmente para la segunda combinación de tejido/tinción usando aprendizaje de transferencia.

11. Método según la reivindicación 1, donde la red neuronal profunda entrenada se entrena para múltiples combinaciones de tejido/tinción.

12. Método según la reivindicación 1, donde una pluralidad de imágenes de autofluorescencia de la muestra se introducen en la red neuronal profunda entrenada, donde la pluralidad de imágenes de autofluorescencia se obtienen en diferentes canales de autofluorescencia.

13. Sistema (2) para generar imágenes microscópicas teñidas digitalmente de una muestra sin teñir químicamente (22) que comprende:
- 5 un dispositivo informático (100) con un *software* de procesamiento de imágenes (104) ejecutado en el mismo o de ese modo, comprendiendo el *software* de procesamiento de imágenes (104) una red neuronal profunda entrenada (10) que se ejecuta usando uno o más procesadores (102) del dispositivo informático (100), donde la red neuronal profunda entrenada (10) está entrenada con una pluralidad de imágenes o parches de imagen teñidos químicamente coincidentes y sus imágenes o parches de imagen de autofluorescencia correspondientes de la misma muestra y donde los pares coincidentes se someten a un proceso de registro global basándose en
- 10 vectores de características extraídos y sus ubicaciones correspondientes de los pares coincidentes y un proceso de registro local posterior que aprende, usando una red neuronal, el mapeo de color entre los pares coincidentes registrados globalmente seguido de registro de características local que coincide con características locales de los pares coincidentes, la red neuronal profunda entrenada (10) configurada para recibir una o más imágenes de autofluorescencia (20) obtenidas en una o más bandas de longitud de onda de emisión usando uno o más filtros de emisión de la muestra (22) y producir la imagen microscópica teñida digitalmente de la muestra que es
- 15 sustancialmente equivalente a una imagen de campo claro correspondiente de la misma muestra que se ha teñido químicamente.

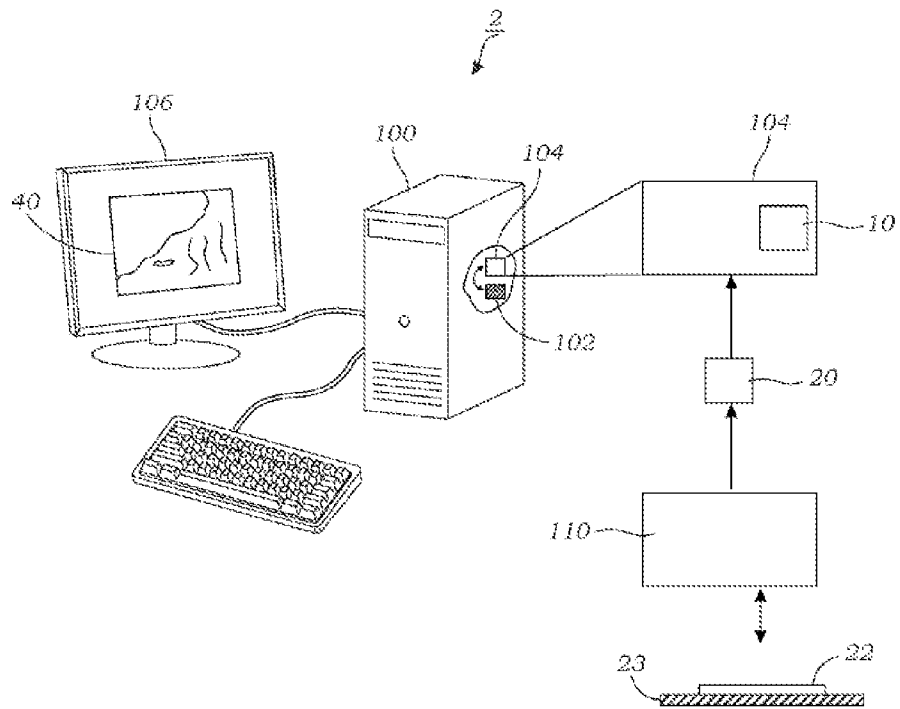


FIG. 1

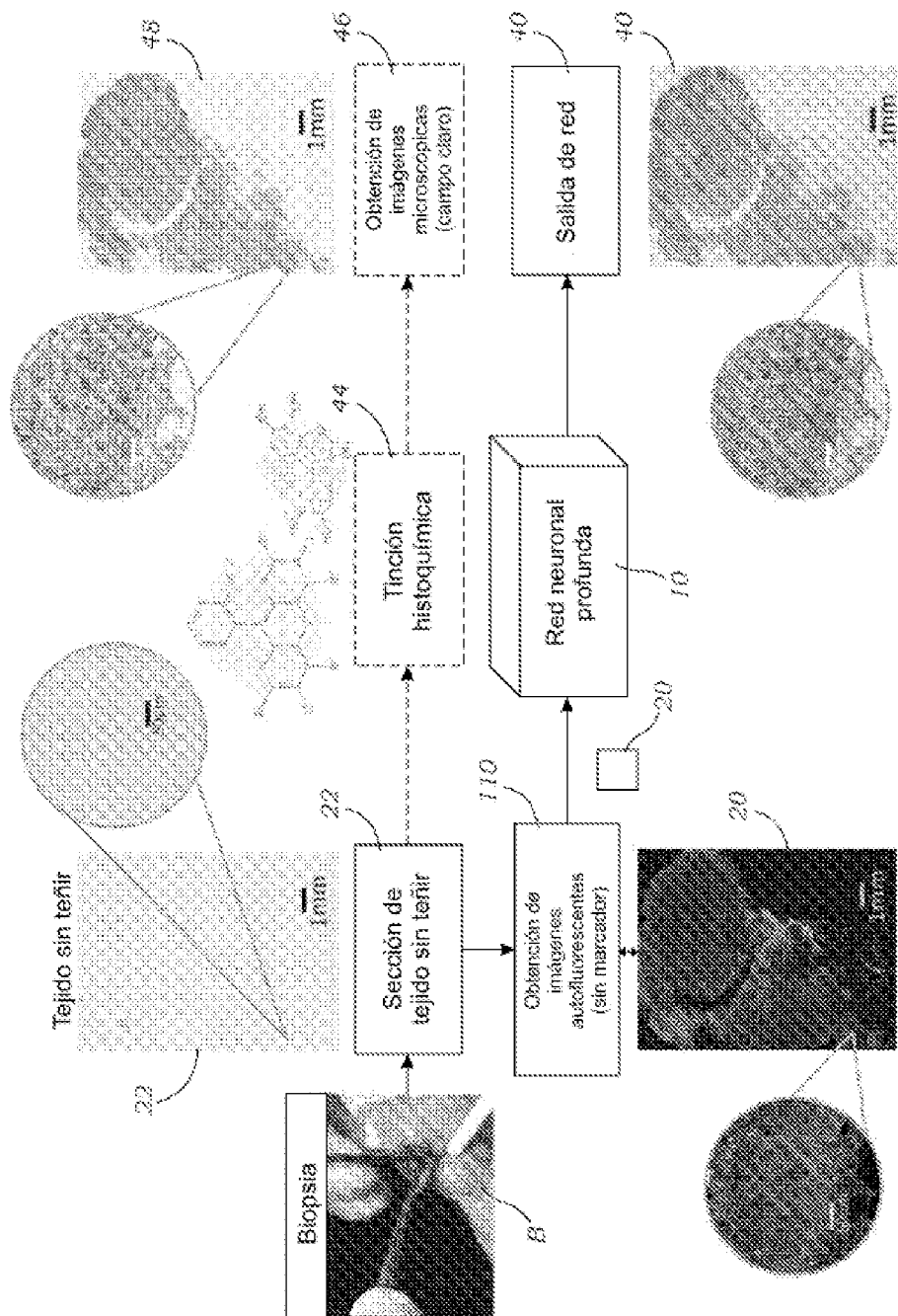
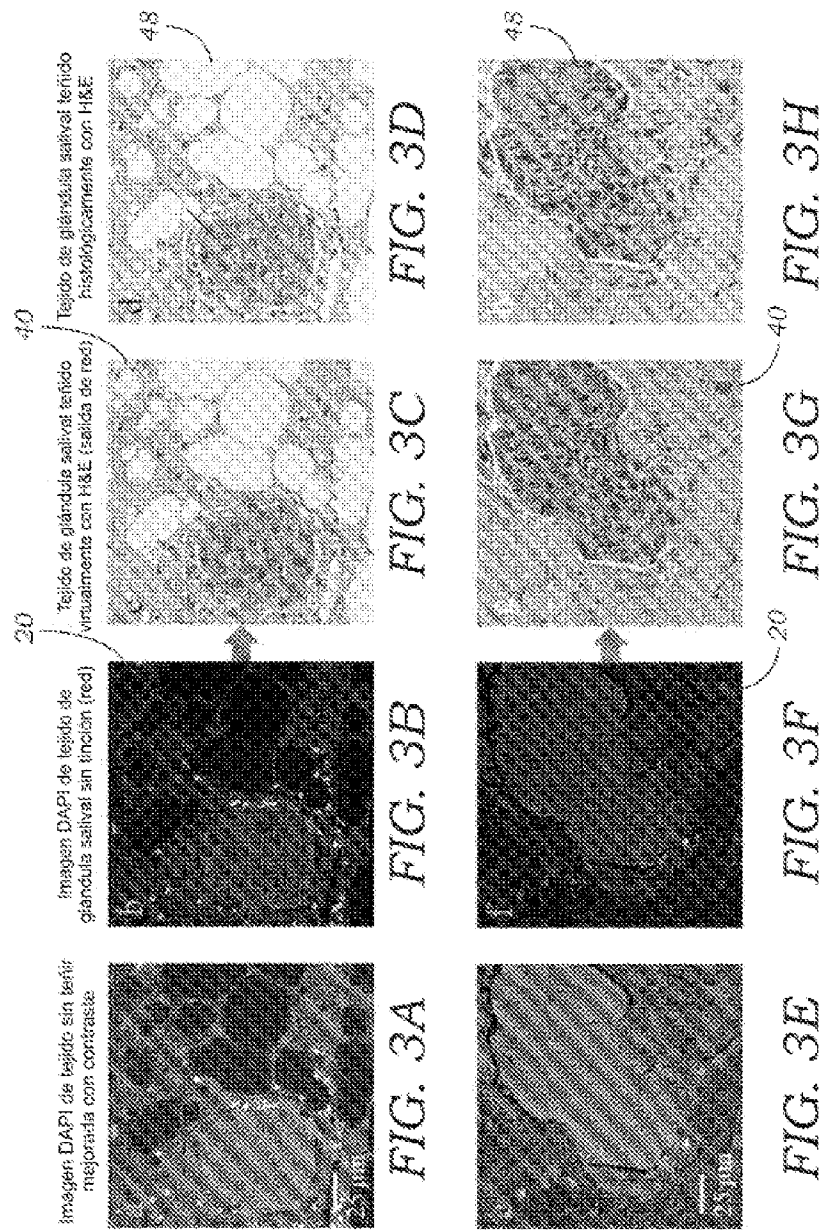
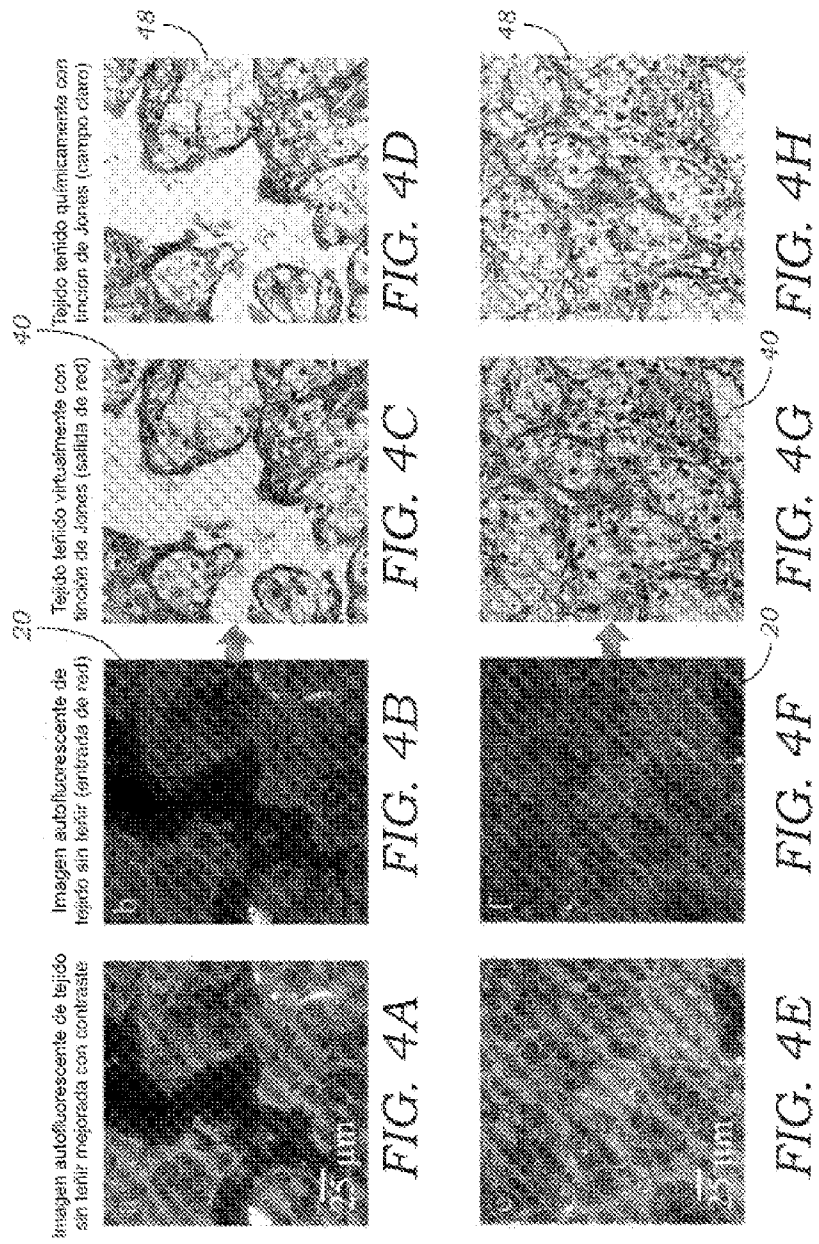
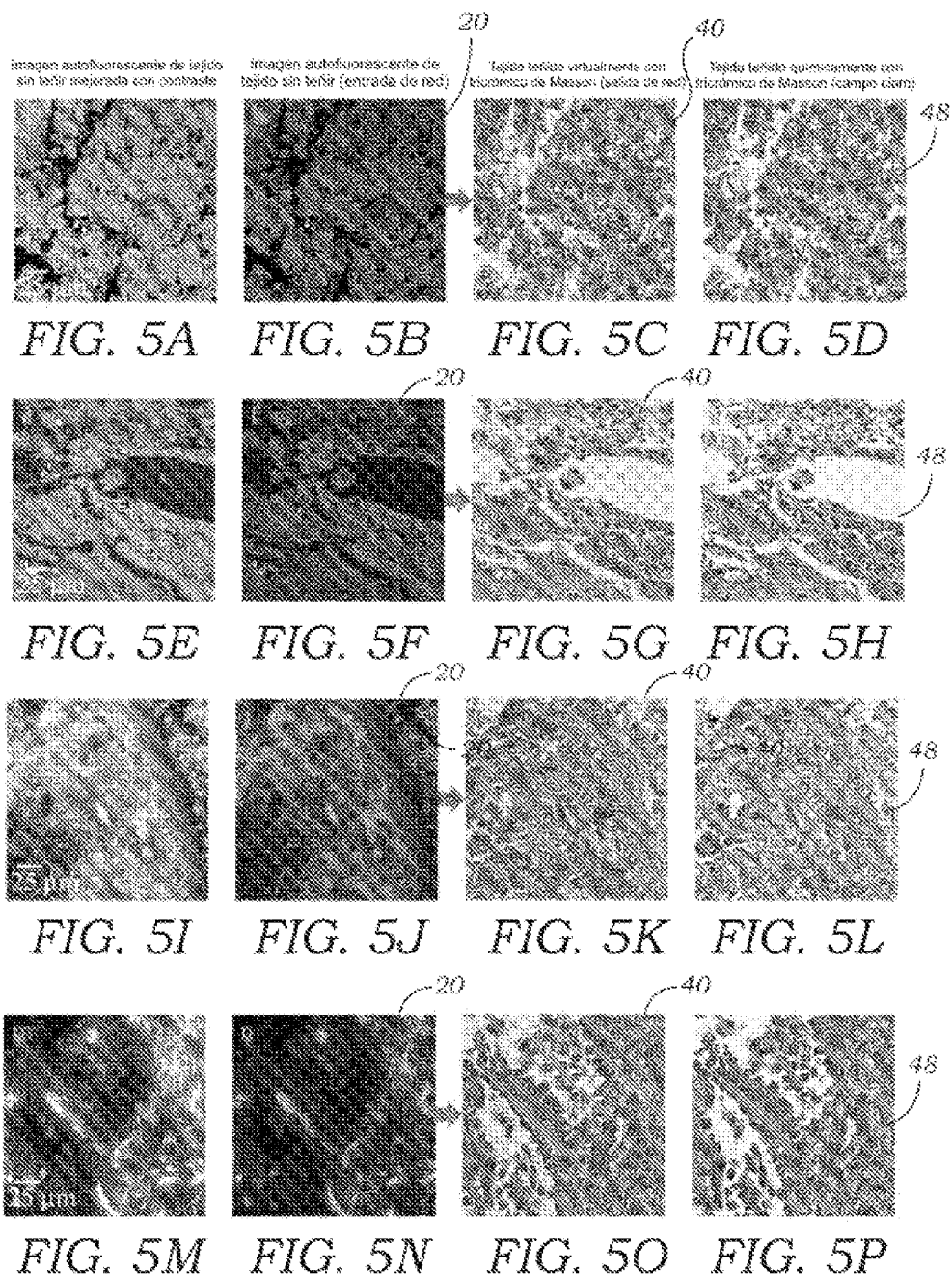


FIG. 2







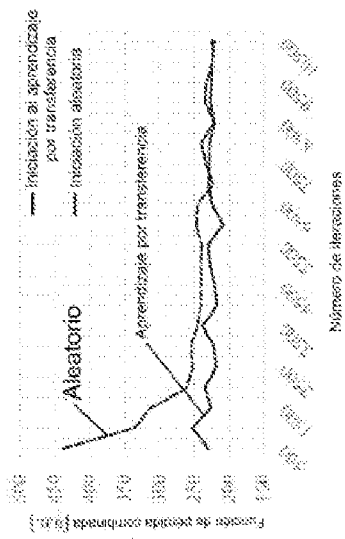
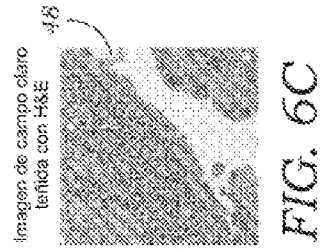
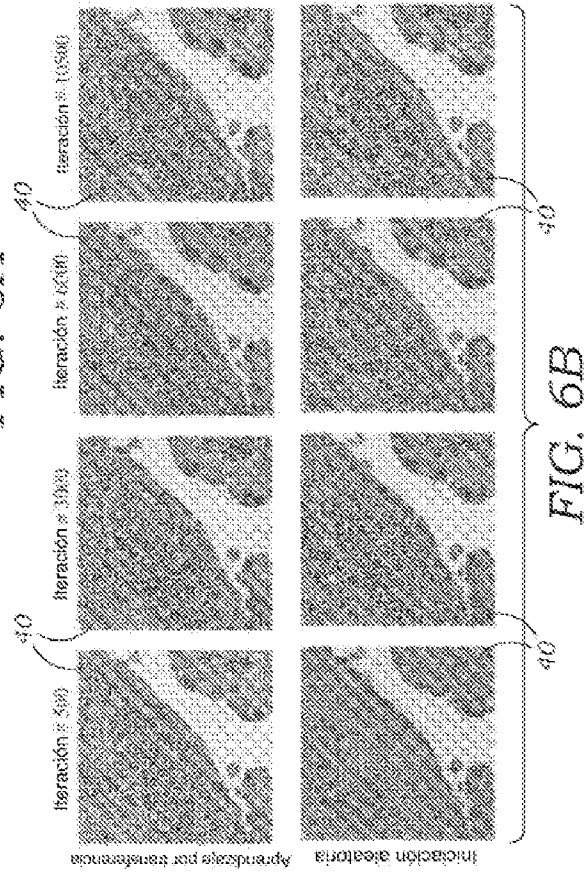
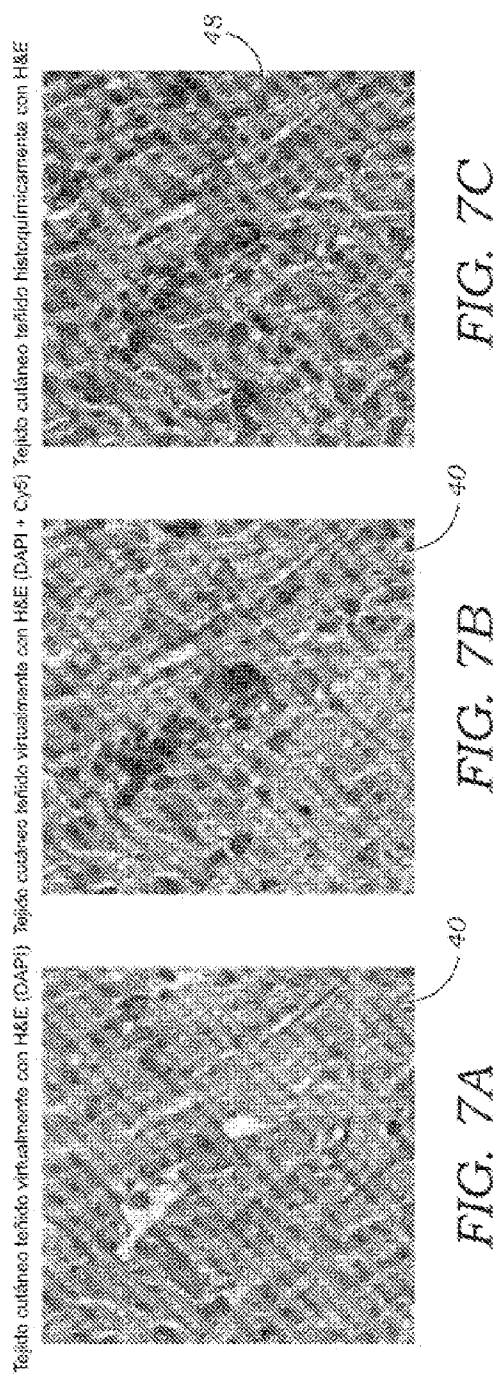


FIG. 6A





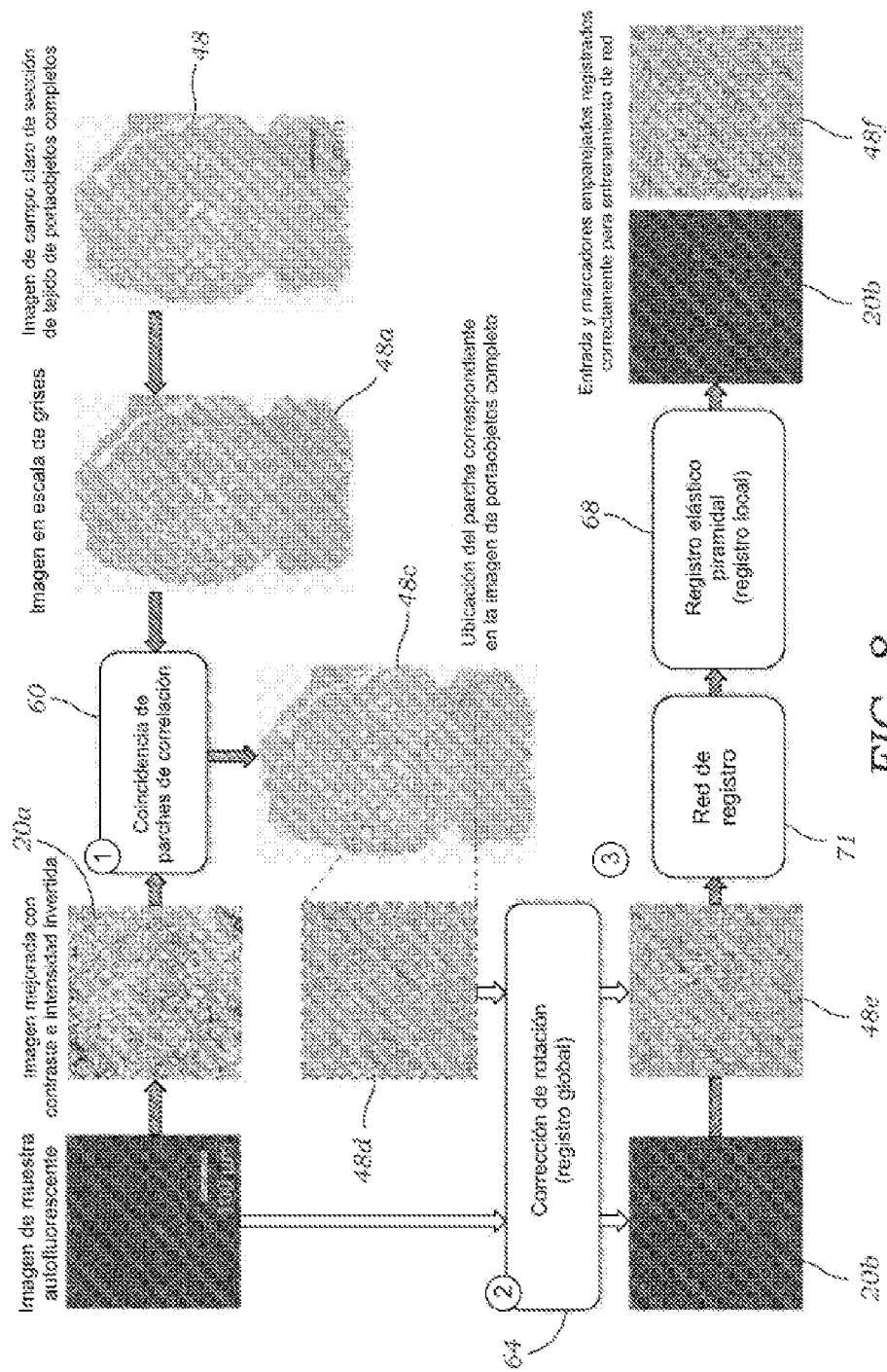


FIG. 8

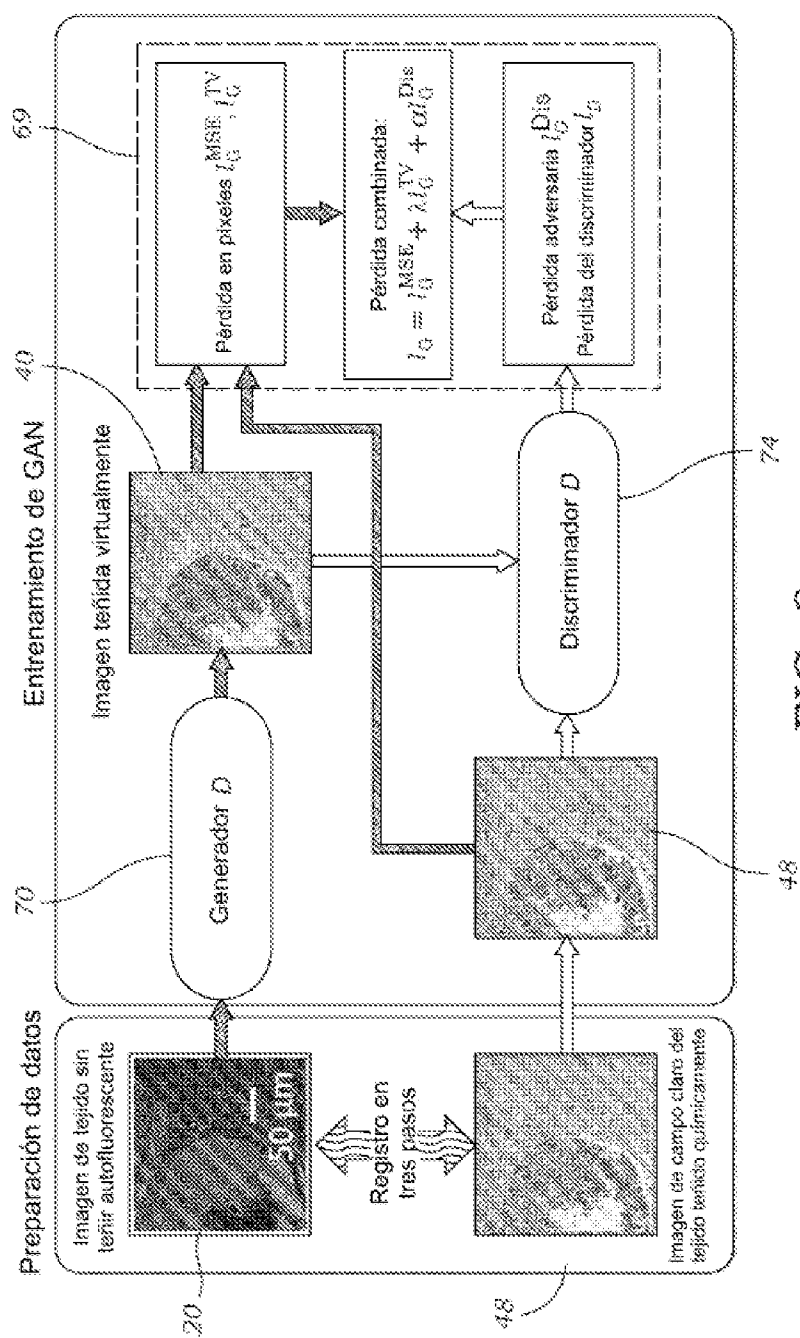


FIG. 9

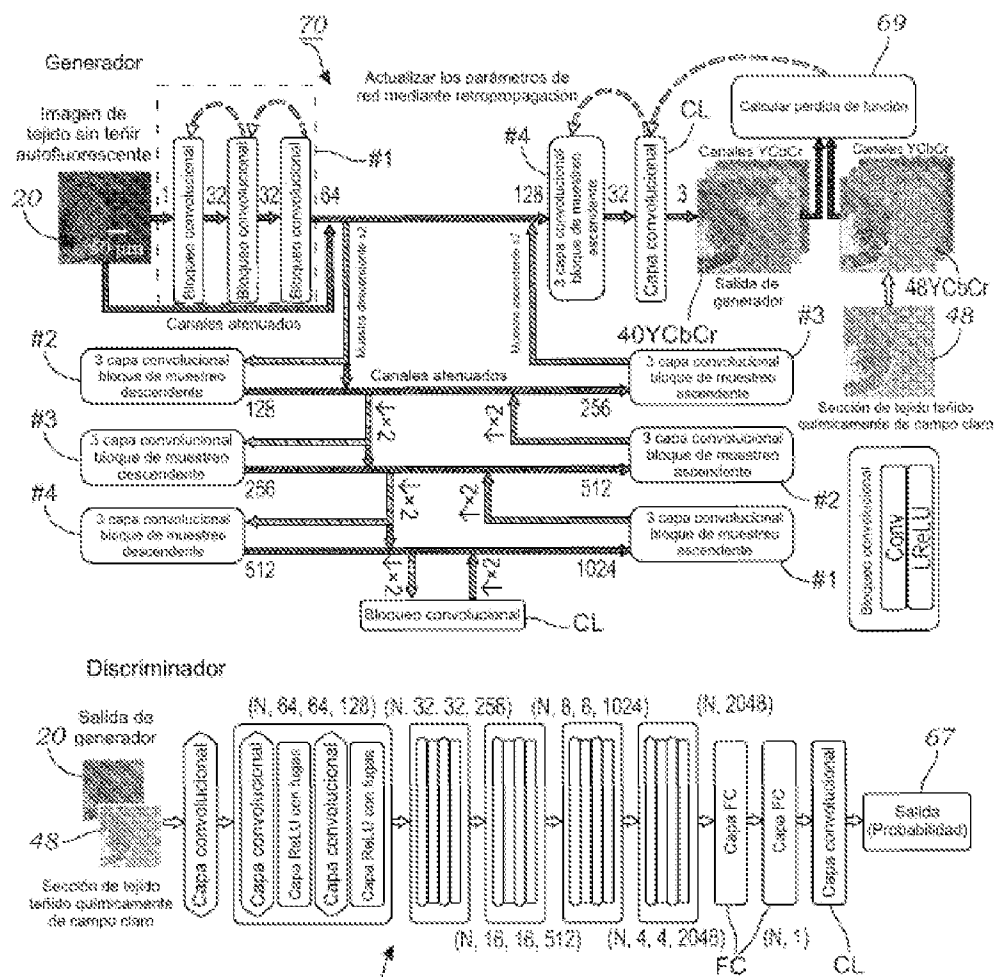


FIG. 10