

A. RESUMEN

Las escenas virtuales hiperrealistas están cada vez más presentes en nuestra sociedad, con una amplia variedad de aplicaciones en áreas como la fabricación, la arquitectura, el diseño de moda y el entretenimiento, incluyendo películas, videojuegos y realidad aumentada y virtual. Generar imágenes realistas de tales escenas requiere modelos de iluminación, geometría y materiales altamente precisos, lo que puede ser costoso de obtener en recursos y tiempo. Tradicionalmente, estos modelos a menudo han sido creados manualmente por artistas entrenados, pero este proceso puede ser prohibitivamente ineficiente. Alternativamente, se pueden capturar ejemplos del mundo real, pero este enfoque presenta desafíos adicionales en términos de precisión y escalabilidad. Además, aunque el realismo y la precisión son cruciales en esos procesos, la eficiencia en el renderizado también es un requisito clave, para que se puedan generar imágenes realistas con la velocidad requerida en muchas aplicaciones del mundo real. Uno de los desafíos más significativos en este sentido es la adquisición y representación de materiales, que son un componente crítico de nuestro mundo visual y, por extensión, de las representaciones virtuales del mismo. Sin embargo, los enfoques existentes para la adquisición y representación de materiales están limitados en términos de eficiencia y precisión, lo que restringe su impacto en el mundo real. Para abordar estos desafíos, los enfoques basados en datos que aprovechan el aprendizaje automático pueden proporcionar soluciones viables. Sin embargo, el diseño y entrenamiento de modelos de redes neuronales que satisfagan todos estos requisitos sigue siendo un desafío, que requiere una cuidadosa consideración de los compromisos entre calidad y eficiencia.

En esta tesis, proponemos soluciones novedosas basadas en aprendizaje automático para abordar varios desafíos clave en la renderización de escenas virtuales. Nuestros métodos hacen uso de diversas formas de redes neuronales para introducir algoritmos innovadores para la codificación de la radiancia, la generación digital de materiales, su edición y estimación. En primer lugar, presentamos un sistema de transferencia de atributos visuales que puede generalizar eficazmente a nuevas condiciones de iluminación y distorsiones geométricas, que ampliamos para la adquisición de materiales de alta resolución mediante un dispositivo diseñado para tal propósito. También, proponemos un modelo generativo capaz de sintetizar texturas teselables – que se puede concatenar consigo misma sin que aparezcan artefactos de repetición – a partir de una sola imagen de entrada, lo que ayuda a mejorar la calidad de la renderización de materiales. Esta tesis también presenta una representación de materiales que codifica con precisión la reflectancia del material y ofrece potentes capacidades de edición y propagación. Para lograr este objetivo, nos basamos en trabajos recientes en campos neuronales. Además de la reflectancia, presentamos un método novedoso para la codificación de iluminación

global que aprovecha modelos generativos cuidadosamente diseñados para lograr un muestreo significativamente más rápido y de mayor calidad que los trabajos disponibles en la literatura. Por último, proponemos dos métodos innovadores para la digitalización escalable de materiales. Usando escáneres como dispositivo de captura, presentamos un modelo generativo que puede proporcionar estimaciones de reflectancia de materiales de alta resolución utilizando una sola imagen como entrada. Adicionalmente, introducimos el primer algoritmo de cuantificación de incertidumbre que aumenta su fiabilidad y eficiencia. Además, presentamos un método neuronal para digitalizar las propiedades mecánicas de tejidos utilizando imágenes de profundidad como entrada, que extendemos con una métrica de similitud perceptualmente validada mediante un estudio de usuario. Globalmente, las contribuciones de esta tesis representan avances significativos en los campos de la codificación de la radiancia y la adquisición y edición digital de materiales, lo que permite mejorar la calidad, escalabilidad y eficiencia de los sistemas de renderización.

A.1 Antecedentes

Generación Digital de Materiales

La digitalización precisa y de alta resolución de la reflectancia de los materiales es un problema crucial para generar entornos virtuales atractivos y realistas. Este proceso de digitalización se realiza típicamente utilizando un *goniorrefleatómetro*, una máquina altamente compleja diseñada para medir la respuesta fotométrica de un material utilizando fuentes de luz y cámaras colocadas en diferentes posiciones alrededor de su semiesfera. Los dispositivos existentes para medir la apariencia del material en muestras espacialmente variables se limitan a una sola escala, ya sea micro o mesoscópica. Esta es una limitación práctica cuando el material tiene una estructura compleja a múltiples escalas, como es común en muchos materiales del mundo real. Para ellos, medir con precisión su reflectancia espacialmente variable requiere imágenes microscópicas de alta resolución. Por ejemplo, el comportamiento óptico de los textiles depende mucho de sus fibras o microestructuras [CLA19]. Además, muchos materiales también presentan variaciones *mesoscópicas* a gran escala, como patrones de impresión o los tartanes, que no se pueden capturar utilizando únicamente fotografías microscópicas. Además, muchos de estos dispositivos no pueden capturar propiedades ópticas importantes, como la transmitancia del material, que es clave para la representación realista de objetos como los tejidos. No hay dispositivos de captura ni algoritmos disponibles para la digitalización precisa y de alta resolución de *Spatially Varying Bidirectional Scattering Distribution Function* (SVBSDF) para materiales heterogéneos. Para introducir tales materiales, que son muy comunes en muchas industrias como la moda, la textil o la fabricación de cuero, en los sistemas de representación, es necesario

crear estas máquinas y desarrollar algoritmos precisos para manejar los datos que capturan. Finalmente, dicho dispositivo puede servir como fuente de generación de datos para entrenar modelos que operen en escenarios de datos más limitados, como en la digitalización a partir de una sola imagen.

Edición y Síntesis

SVBRDFs Teselables Incluso si la reflectancia puede ser medida con precisión para una gran superficie de un material, en entornos de renderización, también es importante que el material digital sea *teselable*. Una textura teselable es una imagen de textura que puede ser repetida infinitamente en todas las direcciones sin presentar discontinuidades visibles. En otras palabras, una textura teselable está diseñada de tal manera que se conecta perfectamente con copias idénticas de sí misma, que pueden ser dispuestas una al lado de la otra o apiladas una encima de la otra para crear un patrón más grande y continuo sin ninguna repetición visible. Esto se utiliza a menudo en gráficos por ordenador, especialmente para crear texturas en modelos 3D, o incluso para la colocación de fondos en sitios web. En los últimos años, se han propuesto diferentes métodos para la generación de texturas teselables a partir de un solo ejemplo de entrada [Mor+17; Rod+19; Mor+20]; sin embargo, estos trabajos presentan varias limitaciones para la generación de *Spatially Varying Bidirectional Reflectance Distribution Functions* (SVBRDFs) teselables. En primer lugar, todos suponen un cierto nivel de regularidad en las imágenes de entrada, o bien las texturas generadas pierden una cantidad significativa de fidelidad visual. Además, estos métodos se limitan a generar un solo mapa de textura (normalmente, el albedo del material). Esto es problemático para la síntesis de SVBRDFs teselables, que requieren transformar cada mapa de textura en el SVBRDF en un mapa teselable, al mismo tiempo que se preserva la coherencia píxel a píxel entre los mapas. En resumen, en la actualidad no es posible generar SVBRDFs teselables de alta calidad a partir de un solo ejemplo de entrada con los métodos propuestos en trabajos previos.

Transferencia de Atributos Visuales Las representaciones efectivas de materiales requieren comprender las propiedades que los definen de manera única, a las que nos referimos como *atributos visuales*. Estos atributos son parámetros que varían espacialmente y mantienen la coherencia espacial con respecto a la estructura del material, mientras que permanecen in-variantes ante los cambios en la iluminación de la escena o la geometría del objeto subyacente. Por ejemplo, pueden representar propiedades ópticas de un SVBRDF, estilizaciones artísticas o máscaras de segmentación semántica. Obtener estos atributos para muestras grandes de materiales es un problema que presenta numerosas dificultades. Una forma de abordarlo es obtener estos atributos en un pequeño ejemplo del material y *transferirlos* a porciones más grandes del mismo, utilizando una imagen de entrada grande que sirve como guía. En los últimos años, se han propuesto diferentes métodos

de transferencia de atributos visuales, sin embargo, están limitados en su eficiencia, robustez, generalización o rendimiento. Idealmente, un método de transferencia de atributos visuales para materiales debería ser robusto ante nuevas condiciones de iluminación del material, degradaciones de la cámara o distorsiones de entrada; además de ser eficiente (proporcionando tiempos de transferencia interactivos incluso en presupuestos computacionales bajos), controlable y predecible, y capaz de generar salidas de alta resolución, que son necesarias para SVBRDFs realistas. Desarrollar un método con estas características puede resultar útil para muchas tareas posteriores, como sistemas eficientes de generación de datos o digitalización de materiales.

Digitalización Escalable

Estimación de la Reflectancia de Materiales Como ya hemos mencionado, la digitalización precisa de los materiales típicamente requiere dispositivos costosos que utilizan cantidades significativas de tiempo e intervención manual para operar. Estos dispositivos proporcionan representaciones digitales altamente realistas de la reflectancia del material a costa de la escalabilidad, lo que limita su aplicabilidad en muchos entornos del mundo real. Por ejemplo, muchos procesos de fabricación, como los de la industria textil, generan una amplia variedad de diferentes materiales que no pueden ser digitalizados fácil ni económicamente con la cadencia requerida en estas industrias. Además, la disponibilidad limitada de estos dispositivos introduce ineficiencias adicionales y costes económicos y ecológicos, ya que las muestras físicas de los materiales deben ser enviadas desde el cliente a la localización del dispositivo de captura.

La estimación de la reflectancia de materiales a partir de dispositivos de bajo coste es un problema que se ha estudiado ampliamente en la literatura [AWL13; AWL15; Des+18; Hen+21; Guo+20b; VPS21]. Estos métodos prometen proporcionar una digitalización de materiales escalable, eficiente y económica. Normalmente, se basan en una o varias imágenes de un material iluminadas con flash tomadas con una cámara de smartphone, y en un modelo de aprendizaje profundo entrenado en materiales sintéticos que proporciona estimaciones plausibles. Sin embargo, estos enfoques presentan varias limitaciones que los hacen inadecuados para flujos de trabajo en los que se requiere digitalización precisa. En primer lugar, muchos de estos métodos se basan en modelos generativos que tienden a producir artefactos o utilizan grandes redes neuronales que limitan la resolución de salida, lo que reduce la calidad de los materiales estimados. Además, las imágenes tomadas con smartphones presentan importantes desafíos de calibración, mientras que la mayoría de los conjuntos de datos existentes utilizados para entrenar estos modelos son puramente sintéticos, lo que limita aún más la calidad de estas estimaciones. Además, muchos de estos métodos utilizan funciones de coste perceptuales para el entrenamiento o la optimización en tiempo de evaluación,

lo que, si bien ayuda a digitalizar materiales estocásticos, presenta desafíos adicionales para garantizar la repetibilidad y consistencia necesarias para construir un inventario digital. Además de estas limitaciones, la estimación de reflectancia de materiales a partir de una sola imagen sigue siendo un problema mal planteado y, por lo tanto, muchos resultados diferentes pueden ser plausibles dados los mismos datos de entrada. Hasta ahora, no existe una forma de cuantificar la incertidumbre para este problema, lo que dificulta la aplicabilidad de estos sistemas en escenarios del mundo real, y limita su fiabilidad y eficiencia.

Comportamiento Mecánico de Materiales Textiles Además de la reflectancia de los materiales, otro componente importante de las escenas virtuales es la geometría de los objetos. Mientras que muchos objetos del mundo real tienen formas estáticas, que pueden ser escaneadas con dispositivos relativamente accesibles como el sensor *LiDAR* en *iPhones*, existen materiales que cambian de forma cuando se les aplican fuerzas externas, como la gravedad. En este sentido, los materiales deformables requieren modelos computacionales y sistemas de captura específicos. Uno de los materiales más comunes de este tipo son las telas, que son ubicuas en el mundo real y se utilizan en nuestra ropa, muebles o vehículos. Para ellas, capturar con precisión su forma estática no es suficiente: Es necesario poder simular su comportamiento en nuevos escenarios, como nuevas prendas o fuerzas dinámicas, como el viento [Ber+23], o el movimiento humano [SOC22]. Además, las telas son increíblemente variadas en sus patrones de fabricación (por ejemplo, diferentes patrones de tejido o estructuras de punto), composiciones y acabados, lo que no solo determina su apariencia óptica [CLA19] sino también su comportamiento mecánico. El problema de capturar las propiedades mecánicas de las telas para que puedan ser re-simuladas en entornos virtuales ha sido ampliamente estudiado en la literatura; sin embargo, las soluciones actuales requieren una intervención humana tediosa o dispositivos de captura costosos, lo que dificulta su escalabilidad. Además, existe una falta de entendimiento sobre la similitud perceptual del comportamiento mecánico de las telas, lo que crea desafíos adicionales para diseñar sistemas precisos de estimación de parámetros mecánicos de tejidos.

Codificación Neuronal de Radiancia

Reflectancia Las *Bidirectional Texture Functions (BTFs)* proporcionan representaciones densas y de gran exactitud de la reflectancia de los materiales, sin embargo, son típicamente prohibitivas en términos computacionales y de memoria. Trabajo reciente [Rai+19; Rai+20; Kuz+21] presenta representaciones neuronales para BTFs, consiguiendo resultados de muy alta calidad y eficiencia. Los métodos neuronales de codificación de reflectancia entrenados en BTF sintéticos o medidos han mostrado resultados prometedores para aumentar el realismo de los materiales renderizados. Sin embargo, estas

codificaciones son inmutables, lo que significa que su salida para una consulta determinada de UV, cámara y vector de luz queda fija una vez que se entrenan. Esto puede ser muy limitante cuando el fragmento del material utilizado para el entrenamiento es demasiado pequeño o no se puede teselar, lo que ocurre con frecuencia cuando el material ha sido escaneado con un dispositivo de captura. Por lo tanto, las codificaciones de reflectancia neuronal actuales, aunque precisas y eficientes, tienen importantes limitaciones en cuanto a las capacidades de edición, lo que dificulta su aplicabilidad y utilidad en entornos de renderizado del mundo real.

Iluminación Existen numerosas aproximaciones para la representación de iluminación en escenas virtuales, incluyendo *Spherical Harmonics*, *Spherical Gaussians* o *Haar Wavelets*. Estas, sin embargo, tienen dificultades para representar de manera eficiente y precisa condiciones de iluminación no difusas, como fuentes de luz de alta frecuencia, como el sol o las bombillas. Trabajo reciente sobre representaciones neuronales de la iluminación natural [GES22] proporciona mejores aproximaciones para estos casos, pero tiene dificultades con la iluminación interior o las escenas nocturnas con múltiples fuentes de luz. Además, estas aproximaciones neuronales no proporcionan capacidades de muestreo y evaluación de PDF, que son esenciales para el renderizado de Monte Carlo en el trazado de rayos. No existen métodos de aproximación de iluminación que proporcionen estas capacidades de muestreo, a la vez que funcionen con precisión y eficiencia en cualquier tipo de mapa de entorno de entrada. Trabajo reciente sobre modelos generativos invertibles y representaciones neuronales implícitas puede proporcionar un camino prometedor para resolver estos problemas.

A.2 Objetivos

El objetivo principal de esta tesis es crear algoritmos innovadores para la representación computacional y la captura de materiales y la radiancia de escenas. Esto es importante por varios motivos. Estas representaciones digitales pueden ser utilizadas en una amplia variedad de aplicaciones y problemas, incluyendo gráficos por ordenador, videojuegos, producción de películas, realidad virtual y aumentada, o diseño computacional. Tienen el potencial de aumentar significativamente el realismo visual y la calidad de los entornos virtuales creados en estos sistemas. Además, la precisión y la eficiencia de estos métodos tienen un impacto directo en la calidad y los recursos computacionales requeridos para estas tareas. Por lo tanto, obtener representaciones y algoritmos de digitalización precisos y eficientes para materiales y otros componentes de las escenas virtuales puede ayudar a reducir el costo computacional de la renderización, el diseño, la captura y la modelización, lo que puede llevar a sistemas más rápidos,

eficientes y escalables. Además, pueden permitir la creación procedural eficiente de conjuntos de datos para entrenar modelos para resolver otras tareas, como segmentación o clasificación, o generación de imágenes o videos. Finalmente, además de la eficiencia y la precisión, características adicionales como la diferenciabilidad o la capacidad de cuantificación de la incertidumbre crean oportunidades para aplicaciones de optimización o el aprendizaje activo.

En esta tesis, presentamos nuevos métodos basados en aprendizaje automático para abordar los problemas descritos anteriormente. Nuestras soluciones propuestas buscan lograr la máxima eficiencia computacional y de datos, así como la precisión, el control y la fiabilidad. Además, nuestros algoritmos están diseñados para ser beneficiosos para los usuarios finales al incorporar requisitos humanos como componentes perceptuales, de edición, predictibilidad, robustez y cuantificación de incertidumbre. Todos nuestros métodos propuestos aprovechan las capacidades de las redes neuronales, incorporando los últimos avances en representaciones implícitas, diseño de arquitectura y funciones de coste, y procedimientos de entrenamiento. Todos nuestros modelos son completamente diferenciables y se pueden incorporar en los procesos de optimización para resolver problemas de renderización inversa. Nuestras soluciones para la codificación de la radiancia se pueden incorporar sin problemas en motores de trazado de rayos para aumentar la eficiencia del renderizado o en representaciones neuronales de escenas para aumentar el realismo. Además, nuestros métodos de digitalización casual de alta calidad permiten la captura escalable de materiales, lo que puede ayudar a los usuarios finales a crear sus propios materiales virtuales realistas o crear grandes conjuntos de datos a un menor coste.

Los objetivos de esta tesis se pueden resumir en tres puntos:

- Diseñar nuevos métodos basados en algoritmos de aprendizaje automático para la generación, propagación, edición y síntesis de materiales digitales, con altos estándares de predecibilidad, control y calidad.
- Introducir nuevas representaciones para la radiancia de escenas, que sean eficientes, editables y diferenciables.
- Democratizar la captura digital de alta calidad de materiales mediante el diseño de nuevos métodos de bajo coste, escalables, fiables y validados perceptualmente.

A.3 Metodología

En esta sección detallamos la metodología seguida durante el transcurso de esta tesis.

Revisión Bibliográfica

Una vez definidos los objetivos de la tesis, descritos anteriormente, se revisó la bibliografía de trabajos previos para identificar huecos en la literatura y encontrar líneas de investigación prometedoras. En este proceso de revisión de la bibliografía, encontramos una importante ausencia de métodos capaces de generar datos de alta calidad para el entrenamiento de modelos de captura y representación de materiales; además de la oportunidad de entrenar modelos de captura y edición de materiales más precisos, eficientes y robustos. Es importante recalcar que esta revisión se realizó durante todas las etapas de la tesis, ya que los campos del aprendizaje automático, la informática gráfica y la visión por ordenador se mueven a una velocidad muy elevada. En ellos, nuevas soluciones innovadoras se proponen a diario, por lo que mantener vivo este análisis del estado del arte es fundamental para el desarrollo efectivo de algoritmos. Por ejemplo, nuevas tendencias como los campos neuronales [Mil+20] y las representaciones implícitas [Sit+20] han sido identificadas y han sido claves para el desarrollo de muchos de los algoritmos propuestos en esta tesis.

Generación de Materiales Digitales

Los primeros proyectos desarrollados en el contexto de esta tesis están enfocados en la generación de materiales digitales de alta calidad, los cuales pueden ser utilizados posteriormente para otras tareas, como entrenar representaciones neuronales de materiales o crear grandes conjuntos de datos para modelos de estimación de materiales. Estos métodos toman como entrada un solo material, lo que ayuda a aumentar la predictibilidad de sus resultados.

En el capítulo 2, presentamos un método neuronal para la de propagación de atributos visuales capaz de transferir, para un material dado, muchos tipos de mapas de propiedades visuales a imágenes de nuevos parches del mismo material, o uno similar. Estas nuevas imágenes pueden ser capturadas bajo cualquier condición de iluminación, variaciones de color, distorsiones geométricas o configuración de cámara. Este modelo utiliza redes neuronales convolucionales de imagen a imagen con pocos parámetros, para proporcionar la primera solución capaz de aprovechar el comportamiento óptico del material para el problema de transferencia de atributos visuales, al ser entrenado con un conjunto de datos fotométrico y una política amplia de aumentado sintético de datos. Este método puede ser entrenado en menos de un minuto y es capaz de transferir cualquier tipo de atributo visual, desde las normales de la superficie, hasta estilizaciones artísticas o mapas de segmentación semántica. Sin embargo, está limitado en precisión, sólo puede aprender de un conjunto de datos a la vez y sólo puede aprender a transferir un atributo visual.

Para abordar estas limitaciones, en el capítulo 2.C ampliamos este método, presentando una red neuronal más compleja capaz de generar mapas de SVBSDF de gran calidad y alta resolución para cualquier material, dados uno o varios conjuntos de datos fotométricos que representan pequeñas porciones del material. A través de extensas modificaciones en el procedimiento de entrenamiento y mejoras en la arquitectura del modelo, no sólo habilitamos esta transferencia de atributos visuales con la capacidad de aprender de múltiples conjuntos de datos y transferir varias imágenes simultáneamente, sino que también aumentamos la precisión y nitidez de sus estimaciones. Sin embargo, estas mejoras incrementan el coste computacional del entrenamiento de la red neuronal y limitan su generalidad. Con estos modelos, podemos construir materiales a gran escala, que pueden ser utilizados para generar datos BTF para entrenar otras representaciones (Capítulo 4), o construir grandes conjuntos de datos para entrenar modelos que se generalicen a nuevos materiales (Capítulo 5).

Las texturas utilizadas en aplicaciones de renderización, además de representar una gran parte de las variaciones del material, necesitan ser teselables. Es decir, cuando se concatenan espacialmente consigo mismas, no debe ser posible identificar estas repeticiones. En el capítulo 3, presentamos una Red Generativa Antagónica (GAN) entrenada con una sola imagen, que puede generar múltiples texturas teselables del mismo material. Para lograr esto, ampliamos trabajos anteriores sobre síntesis de texturas para permitir texturas teselables mediante un nuevo algoritmo de manipulación de espacios latentes; y para manejar múltiples mapas de textura simultáneamente, mediante modificaciones en la arquitectura neuronal. Además, aprovechamos el discriminador de la GAN como función de estimación de calidad en tiempo de evaluación, lo que ayuda a distinguir entre generaciones de alta y baja calidad. Este modelo logra texturas teselables de mayor calidad que trabajos anteriores, en tiempos computacionales más bajos.

Estos modelos dependen de un solo material como entrada, lo que, si bien aumenta la pre-dictibilidad, limita sus capacidades de generalización. No obstante, estos métodos proporcionan evidencia de que, bajo ciertas suposiciones y mediante procedimientos de entrenamiento altamente controlados, las redes neuronales pueden ser utilizadas para generar datos de alta calidad. Estos datos pueden ser aprovechados para otras tareas, como se demuestra en otras secciones de esta tesis.

Codificación Neuronal de Radiancia

Uno de los problemas de investigación que exploramos en esta tesis es la representación de radiancia utilizando redes neuronales, lo que ofrece varios beneficios potenciales en comparación con enfoques tradicionales. Las redes

neuronales pueden proporcionar funciones eficientes, continuas y completamente diferenciables que son útiles para la renderización directa y para problemas de renderización inversa. Además, pueden permitir novedosas capacidades de edición o muestreo eficiente para la renderización mediante Monte Carlo. En este trabajo, nos centramos en dos componentes clave de las escenas virtuales: los materiales y la iluminación. Nuestro objetivo es desarrollar enfoques basados en redes neuronales para codificar estos elementos que sean precisos y computacionalmente eficientes, abriendo el camino para nuevas aplicaciones en informática gráfica y visión por ordenador.

Respecto a los materiales en el capítulo 4 presentamos una nueva representación neuronal para codificar su reflectancia. Construyendo sobre trabajos previos sobre materiales neuronales [Kuz+21], nuestro método tiene tres componentes: Una *textura neuronal*, que codifica la reflectancia del material en un espacio latente 2D; un *renderizador neuronal* completamente convolucional y con activaciones sinusoidales, que toma como entrada esta reflectancia latente y los ángulos de luz y visión y genera valores lineales RGB; y un *autoencoder*, que puede estimar la textura neuronal de cualquier imagen de entrada 2D del material. Al igual que en trabajos previos sobre campos neuronales [Mil+20], entrenamos cada modelo desde cero para cada material, tomando como entrada su *Bidirectional Texture Function* (BTF), que puede ser capturada o generada sintéticamente. Nuestro enfoque introduce la primera representación neuronal para BTFs con entrada condicional, lo que hace posible la extrapolación de BTFs, la creación de BTF teselables y la síntesis de nuevos materiales mediante propagación de reflectancia. Abordamos una de las principales limitaciones de los trabajos previos sobre materiales neuronales, que era su falta de capacidades de edición, lo que impedía su aplicabilidad a escenarios de renderización del mundo real que requieren materiales de gran escala y teselables. A pesar de estas mejoras, nuestra representación tiene limitaciones. Por ejemplo, propiedades semánticas como la especularidad o el albedo se codifican en el espacio latente y, por tanto, no son fácilmente editables. Sin embargo, creemos que nuestro método tiene un potencial significativo para trabajos futuros sobre representaciones neuronales de materiales.

En el capítulo 7, abordamos el problema de la codificación de la iluminación, presentando un nuevo método neuronal para el muestreo, la codificación de la PDF y la compresión RGB de los mapas de entorno utilizados para la iluminación global en la renderización. Nuestro método se basa en trabajos anteriores sobre *representaciones neuronales implícitas* para comprimir el mapa de entorno HDRI en una función continua y diferenciable. Diseñamos un *normalizing flow* que puede muestrear direcciones de mapas de entorno y medir su densidad de probabilidad. Ambas funcionalidades son esenciales para el muestreo por importancia en la renderización mediante Monte Carlo. Nuestro diseño de arquitectura neuronal y de sus procedimientos de entrenamiento nos permiten obtener modelos que son hasta dos órdenes

de magnitud más rápidos que los métodos analíticos. Utilizando un conjunto de datos de mapas de entorno con propiedades diversas, demostramos alta generalidad y precisión, superando trabajos anteriores en este campo. Para fomentar futuras investigaciones, proporcionaremos una implementación de código abierto y un conjunto de datos de modelos entrenados. Si bien este es el primer método que aprende a muestrear de mapas de entorno, creemos que existe potencial para mejorar aún más nuestro enfoque mediante la introducción de diseños de redes neuronales más sofisticados o aprendiendo distribuciones a priori sobre la iluminación global.

Digitalización Escalable de Materiales

Uno de los objetivos clave de esta tesis es desarrollar soluciones de digitalización que sean escalables y asequibles, sin depender de hardware costoso o inaccesible. Para lograrlo, hemos realizado dos contribuciones diferentes. En primer lugar, en el capítulo 5, presentamos un método novedoso para estimar las *Spatially-Varying Bidirectional Reflectance Distribution Functions* (SVBRDF) utilizando escáneres como dispositivo de captura. Este método se basa en gran medida en los algoritmos de generación de datos presentados en secciones anteriores de esta tesis. Nuestro enfoque utiliza una red neuronal ligera con mecanismos de atención, lo que resulta en estimaciones con una resolución considerablemente mayor que los métodos anteriores que se utilizaban imágenes tomadas con smartphones como entrada. Observamos empíricamente que la reflectancia del material depende en gran medida de su microgeometría, que los escáneres planos pueden capturar adecuadamente, a la vez que proporcionan imágenes que se pueden utilizar directamente como el *albedo* de la SVBRDF. Una de las contribuciones clave de este trabajo es la introducción del primer algoritmo de cuantificación de incertidumbre para la estimación de materiales, para el cual demostramos aplicaciones en la creación de conjuntos de datos a través del aprendizaje activo. Además, una extensión de este trabajo forma parte de [Textura.ai](https://textura.ai), que es utilizada por usuarios reales, lo que demuestra su robustez y fiabilidad. Tenemos confianza en que las novedades introducidas por este método darán forma a futuros avances sobre el problema de la digitalización de materiales a partir de una sola imagen. Aunque confiamos en las capacidades de nuestro método, este se puede extender para estimar propiedades de reflectancia adicionales como la anisotropía o la transmitancia, para generalizarse a otros materiales o dispositivos, y para mejorar aún más su exactitud.

Además, en el capítulo 6, proponemos un método para la estimación de parámetros mecánicos de materiales textiles utilizando imágenes de profundidad y la densidad del material como entradas. Nuestro método logra una mayor escalabilidad que los trabajos anteriores, que requerían secuencias de vídeo, intervención manual o dispositivos costosos y específicos. Para lograrlo, proponemos

una solución que es agnóstica a la apariencia óptica del tejido y que puede ser utilizada por operadores no expertos. Utilizamos una red neuronal que predice el conjunto completo de parámetros mecánicos de cualquier tejido utilizando datos sintéticos para su entrenamiento, aprovechando una política extensiva de aumento de datos, aprendizaje por transferencia, mecanismos de atención y múltiples formas de agrupamiento de activaciones internas. Una contribución significativa de este trabajo es nuestra propuesta de una métrica de similitud de propiedades mecánicas en tejidos, que mide las distancias entre el comportamiento mecánico de los tejidos utilizando solo imágenes como entrada. Validamos esta métrica con un estudio de usuarios, demostrando que se relaciona en gran medida con la percepción humana. Aunque nuestro método tiene contribuciones significativas para la estimación del comportamiento mecánico de los textiles, todavía está limitado en al menos dos formas: utiliza imágenes de profundidad como entrada, que son relativamente fáciles de obtener pero menos accesibles que las imágenes RGB, y la métrica propuesta de similitud no es fácilmente diferenciable, lo que dificulta su potencial como función de pérdida en problemas de optimización.

A.4 Resultados

Estas son las principales contribuciones de esta tesis:

- Un método basado en aprendizaje profundo para la propagación de atributos visuales de un material a muestras más grandes del mismo material, o de un material similar. Usando una red neuronal convolucional con un número reducido de parámetros, en-trenada mediante un *dataset fotométrico* y una política extensiva de aumentado de datos, el modelo puede generalizar a nuevas condiciones de iluminación, cámara, color o distorsiones geométricas. Enseñamos las capacidades de nuestro método para la transferencia de atributos visuales de varios tipos, como las normales de la superficie del material, mapas de segmentación semántica o ediciones artísticas. (Capítulo 2).
- Una extensión al método anterior, para habilitar un sistema de captura a dos escalas, que permite digitalizar un material a altos niveles de resolución y precisión. Para ello, extendemos el método presentado en el capítulo 2 para permitir la transferencia de varios mapas de propiedades simultáneamente, con un mayor nivel de exactitud. (Capítulo 2.C).
- Un modelo generativo capaz de generar *texturas teselables* usando una sola imagen como entrada. Explotando las capacidades de las redes generativas antagónicas para la síntesis de texturas, un innovador algoritmo de manipulación de espacios latentes, y usando el discriminador como una métrica de calidad, nuestro modelo es capaz de generar texturas teselables a un mayor nivel de calidad perceptual y a un coste computacional que métodos anteriores. Además, proponemos una extensión a este método para la generación de SVBRDFs teselables usando un sólo material como entrada,

y enseñamos las capacidades de nuestro algoritmo para sintetizar texturas de diferentes niveles de regularidad. (Capítulo 3).

- Una representación basada en campos neuronales capaz de codificar y transferir BTFs. Nuestro modelo se construye sobre trabajo anterior en representaciones neuronales de materiales, y proponemos una representación latente ligera que puede ser descodificada por una red neuronal implícita en valores de reflectancia por téxel. Esta representación latente es estimada por una red neuronal del tipo *autoencoder*, que puede ser utilizada para propagar valores de BTF a estructuras nuevas, habilitando nuevas capacidades de edición y transferencia de materiales neuronales, proporcionando nuevas competencias a este tipo de representaciones. (Capítulo 4).
- Un innovador sistema de digitalización de materiales capaz de, a partir de una sola imagen de entrada, generar mapas de SVBRDF a una alta resolución, usando un escáner como dispositivo de captura. Nuestro sistema usa un conjunto de datos generado para este propósito y un novedoso modelo de translación de imagen a imagen con mecanismos de atención entrenado con funciones de coste específicamente diseñadas para este problema. Este modelo es capaz de generar materiales digitales con un alto nivel de exactitud y detalle, usando la microgeometría del material como la principal señal de supervisión. Además, proponemos el primer algoritmo para la cuantificación de incertidumbre en problemas de digitalización de materiales, construyendo sobre métodos de aprendizaje profundo Bayesiano, *dropout* de Monte Carlo y métricas perceptuales de BRDF. Demostramos que nuestra métrica es eficaz para la predicción de error en digitalización en el momento de la evaluación, y para la generación eficiente de datasets mediante aprendizaje activo. (Capítulo 5).
- Un novedoso sistema de digitalización de propiedades mecánicas de tejidos usando un sistema casual de captura e imágenes de profundidad. Entrenamos nuestro modelo usando únicamente datos sintéticos, y lo diseñamos para hacer uso de mecanismos de atención neuronales, aprendizaje por transferencia, y diferentes mecanismos de agregación de características latentes, además de una extensiva política de aumentado de datos. Nuestro modelo es capaz de predecir con exactitud las propiedades mecánicas de fragmentos de materiales textiles sin la necesidad de costosos sistemas de captura. Además, proporcionamos una innovadora métrica perceptual para medir la similitud entre materiales textiles en función de sus parámetros mecánicos, la cual validamos mediante un estudio de usuario. (Capítulo 6).
- Un método neuronal para mapas de entorno, que habilita una representación eficiente para iluminación global en escenas virtuales. Para ello, usamos dos redes neuronales diferentes: Un *normalizing flow*, capaz de aprender la distribución de probabilidad de un mapa de entorno de entrada, y de muestrear direcciones de luz de forma eficiente; y una red neuronal implícita con activaciones sinusoidales, capaz de transformar direcciones de luz a valores de radiancia lineal en RGB. Demostramos la eficacia de nuestro método en aplicaciones de muestreo por importancia para renderización por

Monte Carlo, consiguiendo representaciones de iluminación más precisas y exactas que métodos anteriores de compresión de iluminación; y dos órdenes de magnitud menor coste computacional que métodos tradicionales de muestreo de iluminación. (Capítulo 7).

- Un estudio integral de las capacidades y limitaciones de las redes neuronales profundas para el problema de la descomposición intrínseca. Mediante un estudio exhaustivo del estado del arte, los datasets, funciones de coste, arquitecturas neuronales y heurísticas, proporcionamos una nueva categorización de estos trabajos. Además, proponemos nuevas direcciones de investigación, construyendo sobre artículos recientes en renderización neuronal, inverso o diferenciable; además de los nuevos avances en sistemas de aprendizaje automático. Esta contribución está publicada en [Gar+22], y usamos parte de ella para la sección 1.1.

Estas contribuciones han dado lugar a las siguientes publicaciones:

- *"Neural Photometry-guided Visual Attribute Transfer"*
Carlos Rodríguez-Pardo, Elena Garces; *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 2021 [RG21].
Este journal tiene un factor de impacto de 5.226, su posición en el índice JCR es 13 de 110 (Q1) en la categoría Computer Science, Software Engineering (datos de 2021).
- *"A Survey on Intrinsic Images: Delving Deep into Lambert and Beyond"*
Elena Garces, Carlos Rodríguez-Pardo, Dan Casas, Jorge Lopez-Moreno; *International Journal in Computer Vision (IJCV)*, 2022 [Gar+22].
Este journal tiene un factor de impacto de 13.369, su posición en el índice JCR es 10 de 145 (Q1) en la categoría Computer Science, Artificial Intelligence (datos de 2021).
- *"SeamlessGAN: Self-Supervised Synthesis of Tileable Texture Maps"*
Carlos Rodríguez-Pardo, Elena Garces; *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 2022 [RG22].
Este journal tiene un factor de impacto de 5.226, su posición en el índice JCR es 13 de 110 (Q1) en la categoría Computer Science, Software Engineering (data from 2021)
- *"How Will It Drape Like? Capturing Fabric Mechanics From Depth Images"*
Carlos Rodríguez-Pardo, Melania Prieto-Martin, Dan Casas, Elena Garces; *Computer Graphics Forum (Proceedings of Eurographics 2023)* [Rod+23b].
Este journal tiene un factor de impacto de 2.363, su posición en el índice JCR es 53 de 110 (Q2) en la categoría Computer Science, Software Engineering (datos de 2021).

- *"UMat: Uncertainty-Aware Single Image High Resolution Material Capture"*
Carlos Rodríguez-Pardo, Henar Domínguez, David Pascual, Elena Garces;
Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023 [Rod+23a].
Esta conferencia es el evento de mayor impacto en ciencias de la computación según *Google Scholar*, y está categorizada como CORE A++ conference (datos de 2021)).
- *"Towards Material Digitization with a Dual-scale Optical System"*
Elena Garces, Víctor Arellano, Carlos Rodríguez-Pardo, David Pascual, Sergio Suja, Jorge López-Moreno; *ACM Transactions on Graphics (Proceedings of SIGGRAPH), 2023* [Gar+23].
Este journal tiene un factor de impacto de 7.403, su posición en el índice JCR es 9 de 110 (Q1) en la categoría Computer Science, Software Engineering (datos de 2021).
- *"NeuBTF: Neural Fields for BTF Encoding and Transfer"*
Carlos Rodríguez-Pardo, Konstantinos Kazatzis, Jorge López-Moreno, Elena Garces; 2023
Esta publicación se encuentra actualmente en revisión.
- *"NEnv: Neural Environment Maps for Global Illumination"*
Carlos Rodríguez-Pardo, Javier Fabre, Elena Garces, Jorge López-Moreno; 2023
Esta publicación se encuentra actualmente en revisión.

A.5 Conclusiones

A lo largo de esta tesis, se han abordado diferentes retos en los campos de la informática gráfica, el aprendizaje automático y la visión artificial; desde la generación de representaciones neuronales más eficientes para la codificación de componentes de las escenas, hasta el diseño de algoritmos de aprendizaje profundo para la digitalización de materiales a bajo coste y alta precisión. En concreto, se han propuesto algoritmos para la propagación de atributos visuales (Capítulo 2), para la digitalización realista de la reflectancia de materiales (Capítulo 2.C), su edición y síntesis hacia texturas teselables (Capítulo 3), o su codificación mediante campos neuronales (Capítulo 4). Además, se han propuesto sistemas de captura escalables y de alta calidad mediante dispositivos de bajo coste, incluyendo escáneres para la captura de reflectancia (Capítulo 5) o las cámaras de profundidad para la captura de parámetros mecánicos de tejidos (Capítulo 6). Finalmente, se ha propuesto una representación neuronal capaz de introducir iluminación global a escenas virtuales de forma diferenciable y computacionalmente más eficiente y precisa

que las soluciones disponibles en la literatura (Capítulo 7). Los resultados de esta tesis forman parte de ocho publicaciones, seis de las cuales están aceptadas en revistas académicas o conferencias internacionales, mientras que las demás se encuentran en proceso de revisión.

En conjunto, en esta tesis se introducen avances significativos para la representación y generación de escenas virtuales realistas, proporcionando métodos basados en aprendizaje automático capaces de mejorar el realismo y la eficiencia de estas escenas. El impresionante ritmo de la investigación en inteligencia artificial dificulta prever las soluciones que se desarrollarán en el futuro, sin embargo, creemos que la combinación de soluciones basadas en datos con la renderización basada en la física seguirá definiendo el campo en los años venideros. Esperamos que las ideas presentadas en esta tesis contribuyan a la generación de soluciones digitales más accesibles y sostenibles medioambiental y socialmente.