Reiluminación de imagen con Deep Learning

David Durán Barberán

Resumen– La reiluminación de una imagen es una tarea compleja que actualmente no se ha resuelto a la perfección. En este proyecto nos adentramos en este mundo con la investigación de diferentes proyectos y datasets existentes, poniendo los proyectos a prueba para comprobar su efectividad con diferentes imagenes y comparando resultados para ver qué proyecto ofrece mejores resultados. Para profundizar todavía más, se estudia la estructura interna de uno de los proyectos para ver cómo funciona realmente y probar de modificar el proyecto con el objetivo de mejorar los resultados originales.

Palabras clave– Deep Learning, Datasets, Red convolucional, Red neuronal, GAN, Generador, Discriminador, Codificador, Decodificador, Skip connections, Loss Function.

Abstract– Relighting an image is a complex task that has currently not been perfectly solved. In this project we enter this world with the investigation of different existing projects and datasets, putting the projects to the test to check their effectiveness with different images and comparing results to see which project offers better results. To delve even deeper, the internal structure of one of the projects is studied to see how it really works and try to modify the project in order to improve the original results.

Keywords– Deep Learning, Datasets, Convolutional network, Neuronal network, GAN, Generator, Discriminator, Encoder, Decoder, Skip connections, Loss Function.

1 Introducción

La reiluminación de una imagen consiste en cambiar la dirección de su iluminación, y hacer este proceso con inteligencia artificial sigue siendo una tarea compleja de resolver. Es un tema que últimamente está cobrando mayor interés debido a que automatizar este problema con Deep Learning [1] podría provocar el avance en varios campos.

En el ámbito de la fotografía puede ayudar en la edición de una imagen, mejora de iluminación, montajes fotográficos etc. sin ayuda del trabajo de un profesional. Por otro lado, en la investigación de visión por computador, puede ofrecer el aumento de datos de entrada para el entrenamiento de un modelo, adaptación del dominio, normalización de datos etc. cosa que sería de gran utilidad para conseguir mejores resultados en las redes neuronales.

En base al aumento de interés que se estaba produciendo en este tema, en el AIM2020 [9], una conferencia virtual especializada en visión por computador, se propuso el reto de "Scene Relighting and Illumination Estimation Challenge"[4] en el que diferentes grupos de personas profesionales competían con su propio proyecto con el objetivo de obtener mejores resultados. Lo mismo ha pasado este año 2021 en NTIRE 2021 [8] con "NTIRE 2021 Depth Guided Image Relighting Challenge"[5].

Dentro de este mismo problema hay diversos factores a tener en cuenta, es una tarea que tiene más subtareas en su interior que se deben resolver: reflectancia de los objetos en escena, eliminación y formación de sombras, cambios de color por la iluminación y estimación y transporte de la iluminación.

Por otra parte, un factor muy importante son los datos de entrenamiento que se utiliza, el dataset. Hay variedad de datasets que buscan obtener un mejor resultado en un determinado problema. Hay datasets basados en fotografias de personas, animales, objetos, naturaleza...

Tanto en el Challange del AIM2020 [4] como el de NTIRE [8] se propone el uso de un Dataset creado recientemente llamado Virtual Image Dataset for Illumination Transfer (VIDIT) [3], en el cual resumidamente, hay 390 imágenes de diferentes escenas, de cada una se forman 40 imágenes diferentes, por la combinación entre 5 colores de temperatura y 8 direcciones, en total 15,600 imagenes.

• E-mail de contacte: skkiper2000@gmail.com
• Menció realitzada: Computació
• Treball tutoritzat per: Ramón Baldrich Caselles (CVC)
• Curs 2021/22

Pero, estos proyectos obtienen buenos resultados en otros datasets? En otros objetivos que no sean escenas?

El objetivo principal de este trabajo es estudiar, analizar y experimentar con el estado del arte actual sobre este mismo tema.

Los objetivos propuestos por orden prioritario son:

- Ejecución, análisis i comparación: ¿Qué resultados nos ofrecen los proyectos con diferentes datasets? ¿Qué proyecto da mejores resultados? ¿En qué destaca cada proyecto?
- Modificación y mejora : ¿Se pueden modificar los proyectos de alguna manera?¿En caso afirmativo, qué posibles cambios se puede hacer? ¿Qué resultados nos otorga estas modificaciones?

2 METODOLOGIA I PLANIFICACIÓN

El proyecto está planificado para llevarse a cabo en un plazo de 18 semanas. Cada 4 semanas se hará un informe del progreso y una reunión del equipo para la planificación de tareas a hacer en las próximas 4 semanas, dividiendo así el proyecto en 5 partes (la última parte de 2 semanas), haciéndolo más flexible para poder adaptarlo a cualquier contratiempo y/o cambios de idea.

Cada semana se hará una evaluación interna del trabajo completado (durante esa misma semana) y una conclusión que determinará cómo va el progreso, si va según lo planeado, avanzado o retrasado. Depende de la evaluación habrá que tomar medidas para buscar el éxito en el desarrollo.

Resumidamente, la parte de desarrollo del proyecto consta de 4 periodos base, como se muestra en la figura **??**. La planificación completa se muestra en el apéndice, incluido una descripción de cada tarea, su tiempo estimado y la evaluación del resultado (completado o fallido)

Primer periodo	Investigación del tema Definición de objetivos Planificación general
Segundo periodo	Selección de proyectos Selección de datasets Primer contacto proyecto 1 Primer contacto proyecto 2 Primer contacto proyecto 3
Tercer periodo	Pruebas con proyectos Análisis de resultados Estudio profundo de un proyecto
Cuarto periodo	Brain storming de posibles cambios Modificación de código Pruebas i análisis de resultados

TAULA 1: TABLA DE PERIODOS Y TAREAS DE PLANIFICACIÓN

Para llevar a cabo el trabajo, en base a los objetivos que se proponen en la planificación, se hará una revisión bibliográfica para la investigación de la variedad de proyectos y datasets relacionados con el tema de reiluminación de imagen. Por otra parte, experimentación con proyectos y datasets, seguido de un análisis, observación y comparación de los resultados obtenidos.

3 PROYECTOS ESTUDIADOS

En este apartado se hará una descripción de los proyectos seleccionados con los que se realizaran las pruebas comparativas. La selección de los proyectos es en base al resultado que se ha obtenido en los retos propuestos del AIM2020 i de NTIRE2021. Un proyecto es seleccionado por la diferencia respecto al resto por utilizar GAN.

3.1. Deep Relighting Networks for Image Light Source Manipulation

Deep Relighting Network [10] consta de tres partes: reconversión de la escena, estimación previa de la sombra y rerenderización. En primer lugar, la imagen de entrada pasa por la red de reconversión de escenas para eliminar los efectos de la iluminación, que extrae estructuras inherentes de la imagen de entrada. Al mismo tiempo, otra rama (estimación previa de la sombra) se centra en el cambio del efecto de iluminación, que vuelve a proyectar las sombras de acuerdo con la fuente de luz objetivo. A continuación, la parte del rerenderizador percibe el efecto de iluminación y vuelve a pintar la imagen con el apoyo de la información de la estructura. Las partes de reconversión de escena y la estimación de sombra. tienen una estructura de codificación automática profunda similar, que consiste en una variación mejorada de la red “Pix2Pix”. Es un proyecto que no destaca por obtener los mejores resultados, pero al utilizar GAN se obtienen resultados más realistas visualmente. El modelo tiene una la estructura que se muestra en la figura 1

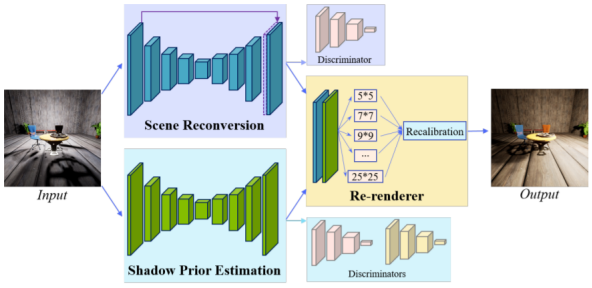

Fig. 1: Esquema de modelo DRN. Existen 3 partes: reconversión de escena, estimación de sombra y rerenderizado, utilizando discriminadores en las dos primeras para obtener un resultado más realista.

3.2. Multi-scale Self-calibrated Network for Image Light Source Transfer

Multi-scale Self-calibrated Network for Image Light Source Transfer [11] consta de tres partes: subred de reconversión de escenas, subred de estimación de sombras y subred de reproducción de imágenes. En primer lugar, la imagen de entrada se procesa en la subred de reconversión de escenas para extraer estructuras de escenas primarias eliminando los efectos de luz. Al mismo tiempo, la subred

de estimación de sombras tiene como objetivo el cambio de efecto de iluminación, que vuelve a proyectar las sombras de acuerdo con la configuración de la fuente de luz de destino. Finalmente, la subred de reproducción de imágenes aprende la temperatura de color objetivo y percibe el efecto de fuente de luz global, y vuelve a generar la imagen con el apoyo de la información de la estructura de la escena primaria y la sombra predicha. La arquitectura general del método se muestra en la figura 2. Es uno de los proyectos que mejores resultados ha obtenido en NTIRE2021.

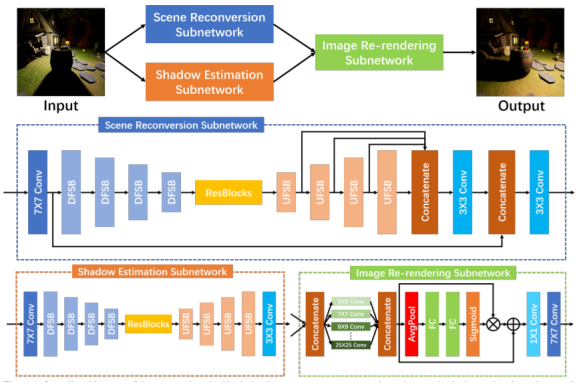

Fig. 2: Esquema de modelo Multi-scale Self-calibrated Network for Image Light Source Transfer. La primera fila muestra un esquema general del modelo, la segunda fila muestra la reconversión de escena de manera más detallada y la última fila muestra la estimación de sombra y la reproducción de imagen.

3.3. YorkU: Norm-Relighting-U-Net (NRU-Net)

El modelo consta de dos redes: (1) la red de normalización, que es responsable de producir imágenes con balance de blancos iluminadas uniformemente, y (2) la red de reiluminación. La red de reiluminación se alimenta de la imagen de entrada y de la imagen uniformemente iluminada producida por la red de normalización. Se utiliza un aumentador de balance de blancos en para aumentar los datos de entrenamiento y obtener mejores resultados. A pesar de no tener paper de explicación, es uno de los proyectos con mejores resultados en los diferentes tracks del AIM2020. El modelo tiene la estructura de la figura 3

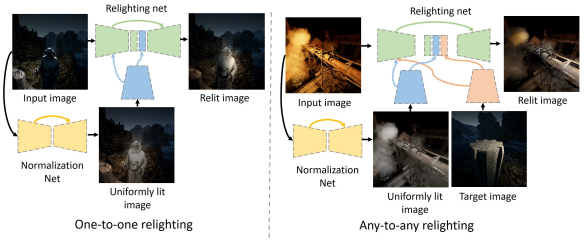

Fig. 3: Esquema de modelo Norm-Relighting-U-Net. El modelo consta de 2 partes: iluminación uniforme y reiluminación de imagen. Se muestra la estructura del modelo para resolver los 2 challanges que se proponen en NTIRE2021

3.4. Resultados métricos

En este apartado se muestran los resultados conseguidos por los proyectos en los challange.

Team	MPS (higher)	SSIM (higher)	LPIPS (lower)	Platform
MCN	0.6347	0.6091	0.3659	PyTorch
YorkU	0.6216	0.5928	0.4144	PyTorch
DeepRelight	0.5892	0.6387	0.3794	PyTorch

TAULA 2: TABLA DE RESULTADOS DE LOS PROYECTOS ESTUDIADOS EN LOS CHALLANGES

Donde se evalúan los resultados utilizando las métricas PSNR, SSIM y LPIPS. SSIM calcula la similitud entre dos imágenes teniendo en cuenta la luminancia, el contraste y la estructura. LPIPS es una red neuronal capaz de predecir un valor de similitud, basado en la similitud visual humana, entrenado con un conjunto de imágenes distorsionadas.

Para la clasificación final, se define Mean Percentual Score (MPS)

$$MPS = 0{,}5 * (SSIM + (1 - LPIPS)) \qquad (1)$$

como el promedio de los puntajes SSIM y LPIPS normalizados, obteniendo así un valor que tenga en cuenta la similitud numérica y la similitud visual humana de las imágenes.

4 DATASETS

En esta sección se presentarán los datasets más destacados que se utilizan para entrenar redes en el campo de la reiluminación de imagen.

4.1. A Dataset of Multi-Illumination Images in the Wild

El conjunto de datos consta de 1016 escenas interiores, cada una fotografiada bajo 25 direcciones de iluminación predeterminadas. Las escenas representan entornos típicos de oficina y domésticos. Para maximizar el número de escenas y materiales, se les añade objetos en diferentes ubicaciones. Las diferentes iluminaciones se consiguen enfocando un foco hacia un punto en concreto y obteniendo así el reflejo de luz como la fuente de luz dominante. La intensidad, la nitidez y la dirección precisas de la iluminación resultante dependen de la geometría de la habitación y sus materiales. Por otra parte, también contiene una imagen etiquetada con el material para cada escena, donde cada material se simboliza con un color. Un conjunto de ejemplo se muestra en la figura 4.

Fig. 4: Imagenes de ejemplo de MIL Dataset con una imagen de una esfera en la esquina superior derecha describiendo los datos de iluminación y la imagen de materiales. Una muestra más completa se muestra en el apéndice.

4.2. VIDIT: Virtual Image Dataset for Illumination Transfer

VIDIT incluye en total 390 escenas diferentes, cada una de las cuales se captura con 40 configuraciones de iluminación predeterminadas, lo que da como resultado 15,600 imágenes. Los ajustes de iluminación son todas las combinaciones de 5 temperaturas de color (2500K, 3500K, 4500K, 5500K y 6500K) y 8 direcciones de luz (N, NE, E, SE, S, SW, W, NW). El conjunto de datos incluye escenas interiores y exteriores, y objetos diversos con diferentes superficies y materiales. Todas las escenas se renderizan con una resolución de 1024×1024 píxeles, contienen diferentes materiales como metal, madera, piedra, agua, plantas, telas, humo, fuego, plástico, etc. Los ajustes de iluminación así como la información de profundidad se registran con cada una de las imágenes renderizadas. VIDIT se divide en diferentes conjuntos; train (300 escenas), validation (45 escenas) y test (45 escenas). Del mismo dataset aparecen dos versiones distintas:

- ECCV 2020 AIM data
- CVPR 2021 NTIRE data

con la diferencia de que en AIM2020 se utiliza el dataset original mientras que en NTIRE2021 se añade un fichero *numpy* que representa un mapa de profundidad de la escena. Un conjunto de ejemplo se muestra en la figura 5

Fig. 5: Imagenes de ejemplo VIDIT con iluminación aleatoria

4.3. Deep Single Image Portrait Relighting

Deep Single Image Portrait Relighting dataset [12] contiene un total de 138.135 imágenes en resolución de hasta 1024x1024. Estas se generan con un algoritmo de reiluminación de un solo retrato. En este trabajo, diseñamos un algoritmo de reiluminación de un solo retrato. Primero se aplica un método de reiluminación de retratos basado en la física para generar un conjunto de datos de reiluminación de retratos (DPR) a gran escala y de alta calidad. Luego, se entrena una red neuronal convolucional (CNN) profunda utilizando este conjunto de datos para generar una imagen de retrato iluminada mediante el uso de una imagen de origen y una iluminación de destino como entrada.

5 PRIMEROS RESULTADOS

En este apartado se experimentará con los proyectos con la finalidad de comprobar que tan buenos son los resultados que nos ofrecen utilizando diferentes datasets. Como imágenes de entrada se forma un conjunto de 2 imágenes de VIDIT, un retrato de DPR, imagen captada con cámara y una imagen de MIL. Los proyectos de DeepRelight y MCN ofrecen un modelo entrenado de muestra, que debido

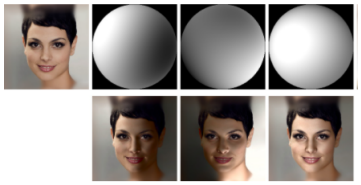

Fig. 6: Muestra de imágenes del dataset de DPR. La primera imagen es un retrato con iluminación uniforme y posteriormente se aplica

al alto coste de recursos que conlleva entrenar los modelos con un dataset completo, se utilizarán estos proyectos con sus respectivos modelos entrenados para realizar la experimentación. Los resultados se obtienen de ejecutar los proyectos en la plataforma de Google Colaboratory, que ofrece los recursos necesarios para ejecutar los repositorios con los modelos entrenados.

De manera general ambos modelos hacen un buen trabajo. Podemos ver como en la figura 7, las imágenes del dataset de VIDIT destacan del resto, son las imágenes con mejores resultados, el motivo es que los modelos entrenados se entrenan con este mismo dataset. Por otra parte, las imágenes que no son de este dataset dan resultados insatisfactorios, vemos que en global las imágenes pierden el color. MCN deja la imagen mejor estructurada que DeepRelight, es decir, visualmente se puede diferenciar la zona iluminada de la zona sombreada, cosa que en DeepRelight los resultados son más confusos.

La imagen de Obama (DPR dataset) tiene el sombreado bien definido pero el color de piel tiene muy poca saturación. Este color tan brillante debería ser únicamente en puntos específicos de la superficie donde la luz impacta directamente. La imagen del dataset de MIL pierde prácticamente todo el color y queda totalmente sombreada. La imagen captada en cámara es realmente complicada de modificar, contiene reflejos, brillos y colores muy vivos, los resultados pierden bastante color, sobre todo el resultado de DeepRelight, donde la imagen queda oscurecida completamente, el resultado de MCN no pierde tanto color pero no se aprecia casi el cambio.

Las imágenes que mejores resultados tienen en la figura 7 las del dataset de VIDIT. La definición de las superficies iluminadas y sombreadas es muy buena, no obstante, las zonas en las que originalmente había sombra no se colorean de la manera correcta, quedan con partes distorsionadas. Comparados con el resto de imágenes, el objetivo está bastante bien conseguido. No obstante, los resultados no son del todo satisfactorios y el motivo puede ser la complejidad de las imágenes. Por esto, se realiza otra prueba utilizando un conjunto de imágenes con iluminación más uniforme.

En la figura 8 vemos como las imágenes del dataset de VIDIT dan resultados muy buenos. La iluminación es correcta, a pesar de perder algo de color, y el sombreado está bien representado. La primera imagen de VIDIT obtiene mejores resultados que la segunda. Esta segunda pierde más color y queda mucho más difuminada. La imagen de DPR, a pesar de ser más uniforme en cuestión de iluminación, los resultados son peores comparados con los de la figura 7. Vemos como para DeepRelight la imagen queda prácticamente oscurecida al completo y el

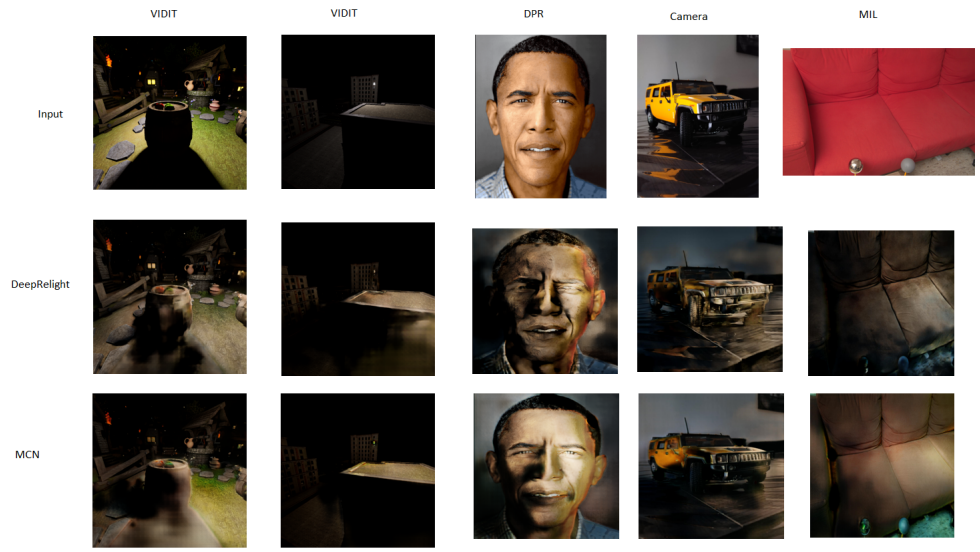

Fig. 7: Primera comparativa de resultados. En la primera fila tenemos las imágenes de input y a continuación los resultados de los proyectos DeepRelight y MCN.

Fig. 8: Segunda comparativa de resultados. En la primera fila tenemos las imágenes de input (con iluminación más uniforme) y a continuación los resultados de los proyectos DeepRelight y MCN.

color se pierde totalmente. Para MCN la sombra está mal representada y se pierden detalles de la imagen original (la sombra es opaca). En cambio, para la imagen de MIL se obtienen buenos resultados. El color no se pierde y la iluminación, a pesar de ser demasiado brillante, está bien representada. Se puede ver una ligera zona oscurecida en la zona donde no debería dar iluminación, mostrando así que no hay ningún objeto que tape por completo la luz. Los resultados de la imagen del dataset de MIL son mejores de lo esperado, teniendo como referencia el resultado anterior.

En conclusión, generalmente los resultados cumplen con el objetivo, aunque no al detalle. Las imágenes quedan más o menos bien iluminadas, con una buena definición de sombras. Por otra parte, algunas pierden mucha tonalidad de color y quedan totalmente irreconocibles (en comparación con la original). La reiluminación es más correcta en imágenes con iluminación uniforme debido a que tiene más información de la escena y es capaz de modificar la imagen sin perder tanto detalle.

6 Estudio profundo de DeepRelight

Como se ha comentado anteriormente, DeepRelight consta de tres partes: reconversión de escena, estimación previa de sombra y rerenderización. En este apartado veremos de una manera más detallada cómo funciona cada parte.

6.1. Reconversión de escena

El objetivo de la reconversión de escena es extraer la información de la estructura inherente de la imagen con los efectos de iluminación eliminados. La red adopta una

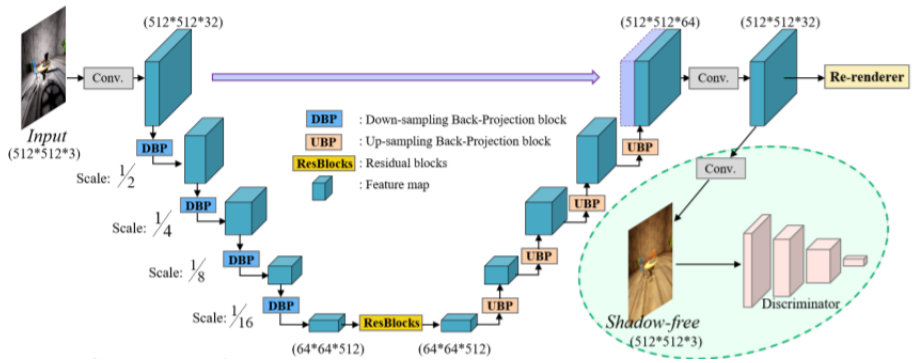

Fig. 9: Estructura de la red de reconversión de escena. La imagen de entrada pasa por 3 bloques llamados ”DBP”, por 9 bloques ResNetz finalmente por 4 bloques ÜBPçon skip connection en la primera capa y un discriminador. El circulo verde sólo existe en la parte de entrenamiento

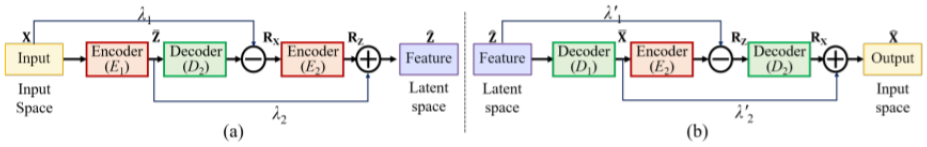

Fig. 10: Estructura de (a) Donw-sampling Back-Projection(DBP) i (b) Up-sampling Back-Projection(UBP)

estructura encoder-decoder con una única *skip connection* que va del principio al final. Primeramente, la imagen pasa por la fase de *downsample* pasando por 4 bloques llamados ”DBP“(Down-sampling Back-Projection) donde el tamaño se reduce a la mitad, a la vez que los canales se doblan para mantener el máximo de información posible. Una vez hemos pasado por la fase de downsample, la información pasa por 9 bloques llamados “ResBlocks”que se encargan de eliminar los efectos de iluminación. A continuación, se pasa por la fase de *upsample* que son 4 bloques “UBP”(Upsample Back-Projection) para devolver la imagen a su tamaño original. Como ground-truth se toma una imagen resultante de un método de fusión [7] donde se combinan todas las imágenes de una escena con diferente iluminación para conseguir una imagen con iluminación uniforme. El discriminador toma la estructura de [6] con 4 capas convolucionales para extraer las representaciones globales.

6.1.1. Back-Projection Blocks

En lugar de realizar un down-sampling de caracteristicas, se adopta el bloque para compensar la pérdida de información a través de residuos. Tanto DBP como UBP consiste en operaciones con codificación y decodificación. Por ejemplo para el bloque DBP, primero se mapea la imagen de entrada (**X**) al espacio latente ($\bar{\mathbf{Z}}$) a través de un encoder. Después se pasa por un decoder para mapear de vuelta al espacio inical ($\hat{\mathbf{X}}$) y se calcula la diferencia resiudal (**Rx = X - $\hat{\mathbf{X}}$**). Este residuo pasa por un encoder para pasarlo al espacio latente **Rz** y finalmente $\hat{\mathbf{Z}}$ **= $\bar{\mathbf{Z}}$ + Rz**. Se puede definir como resultado de cada bloque:

$$\hat{Z} = \lambda_2 E_1(\mathbf{X}) + E_2(D_2(E_1(\mathbf{X})) - \lambda_1 \mathbf{X}) \quad (2)$$

$$\hat{X} = \lambda_2 D_1(\hat{\mathbf{Z}}) + D_2(E_2(D_1(\hat{\mathbf{Z}})) - \lambda_1 \hat{\mathbf{Z}}) \quad (3)$$

Donde **X** es el input, $\hat{\mathbf{Z}}$ es el resultado de los bloques DBP y $\hat{\mathbf{X}}$ el resultado de UBP. Por otra parte tenemos:

- El encoder es una convolución con filter size 3 × 3, stride = 2, padding = 1
- El decoder es una deconvolución con filter size 4 × 4, stride = 2, padding = 1

6.2. Estimación de sombra

Esta red se encarga de producir los efectos de iluminación y sombra deseados. Su estructura es similar a la de la reconversión de escena como podemos ver en la figura 11. Hay 3 diferencias respecto a la reconversión de escena: 1) Se elimina la skip connection, debido a que la red se tiene que centrar en efectos de luz globales. 2) Se añade otro discriminador que se centra en las regiones de sombra, donde se valorarán los pixeles que no superen el threshold de 15/255. 3) La imagen que se toma como ground-truth es la imagen del dataset que tiene la iluminación esperada.

6.3. Re-rendering

Después de procesar la reconversión de escena y la estimación de sombra, se deben fusionar para obtener el resultado final. La fase de rerenderizado tiene 3 partes llamadas como: percepción multiescala, recalibración por canal i proceso de coloreado. La percepción multiescala consiste en pasar por un bloque donde se extraen diferentes características mediante filtros de diferentes tamaños (3x3,5x5...). Después de procesar la percepción multiescala, las características se fusionan en un único mapa de características. Finalmente una convolución (filter size 7x7, padding=3, stride=1 i tanh de capa de activación) colorea la estimación de la imagen.

6.4. Loss function

La función de perdida para el rerenderizado consiste en el error de reconstrucción de imagen por pixel y la diferencia perceptual, La reconstrucción por pixel se mide por

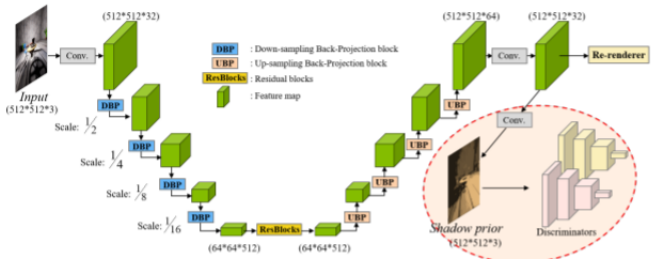

Fig. 11: Estructura de la red de estimación de sombra. Muy similar a la estructura de reconversión de escena pero sin skip connection y con 2 discriminadores. La estructura del circulo rojo desaparece después del entrenamiento

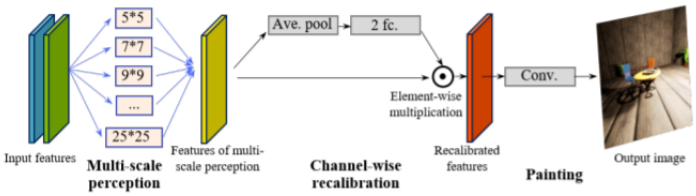

Fig. 12: Estructura de rerenderizado. Consta de 3 partes como: percepción multiescala, recalibración i coloreado.

L1 normalizado y la diferencia perceptual se calcula por las características extraídas por la red VGG-19. VGG-19 está entrenada previamente con el dataset ImageNet[2] para la clasificación de imagen. La fórmula queda definida como:

$$L(Y,T) = ||Y - T|| + \lambda||feat(Y) - feat(T)|| \quad (4)$$

Donde Y es el ground-truth y T el resultado estimado, feat() determina el mapa de características extraídas de VGG-19 y λ es 0.01.

7 Modificación y mejora

Anteriormente se ha visto cómo funcionaba este proyecto con diferentes datasets y hemos visto que los resultados perdían tonalidad de color y detalle de imagen. En este apartado se modifica la estructura interna con la intención de corregir estos errores.

Cambio de estructura con skip connections: Este cambio puede provocar que las imágenes resultantes estén más detalladas y no contengan partes difuminadas. Se pueden aplicar skip connections a las diferentes estructuras que contiene el proyecto, pero se aplicará solamente a la reconversión de escena que es el punto más importante y que más cambio puede ofrecer.

Modificación de loss function - de L1 a MSE: MSE frente a L1 cuadra el error antes de tomar un promedio, de esta manera, se vuelve muy alto si las imágenes tienen valores atípicos. Se puede conseguir que las imágenes no pierdan tanto color.

Modificación de loss function - cambio de estructura de VGG-19: La red VGG-19 devuelve un resultado teniendo en cuenta los datos de una serie de capas específicas. Se pueden modificar y/o añadir capas y observar como evoluciona el resultado. En este caso se añadirán más capas.

8 Nuevos Resultados

En este apartado se muestran los resultados de los cambios comentados en el apartado anterior. Debido al coste de recursos cada modificación de modelo se entrena con el dataset de VIDIT reducido, con un total de 4000 imágenes. En la figura 13 podemos ver los resultados. Como era de esperar, al utilizar un dataset con una reducción de 3 respecto al original, los resultados no muestran ningún tipo de mejora pero si podemos ver diferencias entre ellos. En la modificación de la red VGG-19, las imágenes pierden parte de color y la zona sombreada contiene .ªgujeros", mezclándose así iluminación y sombra en una misma superficie. En cambio, utilizando MSE la zona de sombra queda mejor representada respecto con la modificación de VGG, por otra parte podemos observar como las imágenes que no son del dataset de VIDIT no pierden tanto color. Por ultimo el mejor resultado que se obtiene es con las skip connections, las imágenes no pierden tanto color y la sombra queda mejor representada. Es posible que con un entrenamiento más extenso consiga mejorar el resultado original.

9 Conclusiones

Finalmente nos hemos adentrado en el mundo de Deep Learning para la reiluminación, pasando por el estudio de proyectos (incluido un estudio profundo de uno de ellos), datasets, métricas utilizadas para comparación de imágenes y la modificación de uno de los proyectos. Hemos podido ver que los diferentes proyectos utilizaban ideas similares en la estructura de la red neuronal, a pesar de que más internamente no tengan tanta similitud, utilizan la idea de dividir el problema en tres partes: iluminar uniformemente, crear sombras según el objetivo y finalmente fusionar resultados. Los datasets basados en la multi iluminación de una escena son escasos, pero los existentes son muy completos y contienen un gran numero de imágenes, teniendo en cuenta que el procesado de imágenes es complejo y requiere una gran cantidad de recursos. El proyecto modificado dedicaba más de 80h para entrenarse con el dataset completo de VIDIT, utilizando al menos 8 GPU's simultáneamente. Por esto mismo el dataset utilizado para el entrenamiento del modelo modificado es reducido consideradamente. Por otra

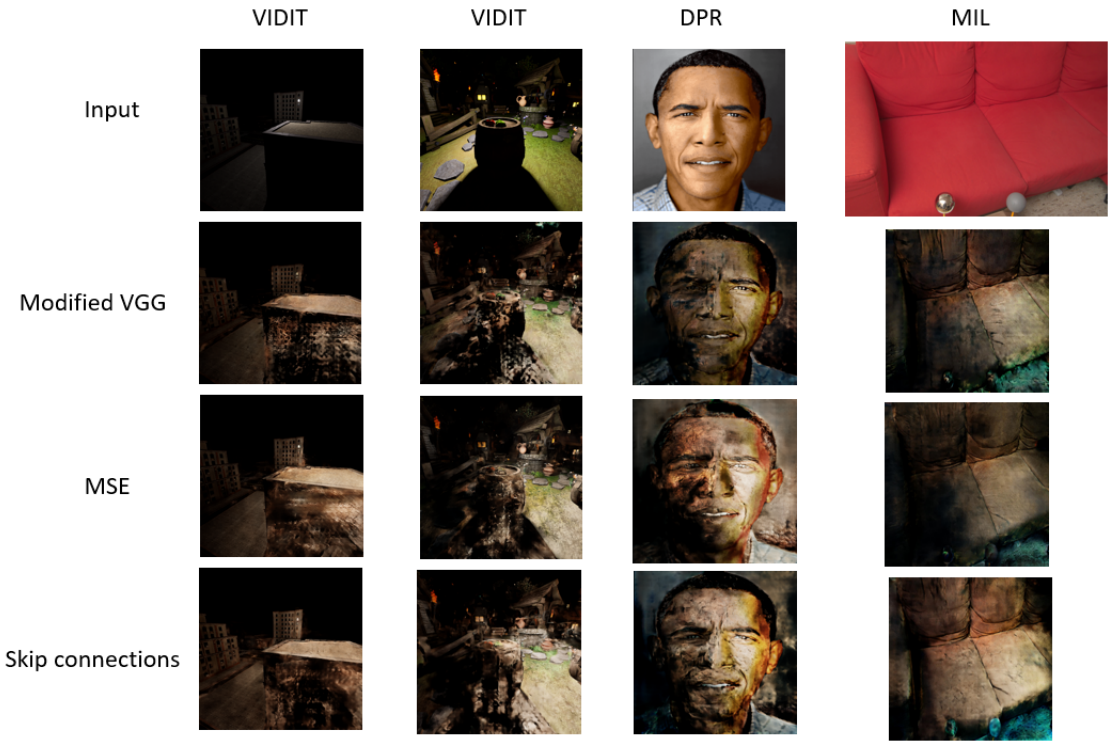

Fig. 13: Imagenes de input-output de la modificación de capas de VGG-19

parte, un contratiempo totalmente inesperado ha sido el ciberataque a la UAB que ha inutilizó temporalmente sus servidores y para este trabajo no se ha podido acceder en gran parte del tiempo de desarrollo a sus servidores de GPU's, este hecho se añade al motivo por el que se utiliza un dataset reducido y al motivo por el que no se ha podido comparar el resultado del proyecto de YorkU.

No obstante, queda pendiente como trabajo futuro el entrenamiento de las modificaciones del proyecto con el dataset completo, para comprobar si realmente se ha sido capaz de realizar alguna mejora.

REFERENCIAS

[1] Raúl Arrabales. Deep learning: Qué es y por qué va a ser una tecnología clave en el futuro de la inteligencia artificial, Oct 2016.

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[3] Majed El Helou, Ruofan Zhou, Johan Barthas, and Sabine Süsstrunk. VIDIT: virtual image dataset for illumination transfer. *CoRR*, abs/2005.05460, 2020.

[4] Majed El Helou, Ruofan Zhou, et al. AIM 2020: Scene relighting and illumination estimation challenge. *CoRR*, abs/2009.12798, 2020.

[5] Majed El Helou, Ruofan Zhou, Sabine Süsstrunk, and Radu Timofte. NTIRE 2021 depth guided image relighting challenge. *CoRR*, abs/2104.13365, 2021.

[6] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks, 2018.

[7] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. In *Computer graphics forum*, volume 28, pages 161–171. Wiley Online Library, 2009.

[8] Radu Timofte. Ntire 2021: New trends in image restoration and enhancement.

[9] Vivaco. Aim2020: Advances in image manipulation workshop and challenges on image and video manipulation.

[10] Li-Wen Wang, Wan-Chi Siu, Zhi-Song Liu, Chu-Tak Li, and Daniel Pak-Kong Lun. Deep relighting networks for image light source manipulation. *CoRR*, abs/2008.08298, 2020.

[11] Yuanzhi Wang, Tao Lu, Yanduo Zhang, and Yuntao Wu. Multi-scale self-calibrated network for image light source transfer. *CoRR*, abs/2104.08838, 2021.

[12] Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David W Jacobs. Deep single-image portrait relighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7194–7202, 2019.

APÉNDICE

A.1. Planificación detallada

A.1.1. Primer periodo

El primer periodo de 4 semanas esta dedicado exclusivamente a la investigación de proyectos, definición de objetivos y planificación general. Durante este periodo nos hemos adentrado en el mundo de la reiluminación de imagen, estudiando papers de varios proyectos, descubriendo las conferencias de visión por computador con retos y observando los proyectos presentados. Por otra parte, a medida que se estudiaban los proyectos surgían ideas que finalmente se han presentado como los objetivos de este trabajo. Una vez tenemos objetivos, preparamos una planificación para poder cumplirlos en el plazo de tiempo correspondiente.

A.1.2. Segundo periodo

El segundo periodo está dedicado a la elección de los proyectos, primeras pruebas de los proyectos y selección de datasets.

- Selección de proyectos (2-3 días): estudio (básico), selección y descarga de proyectos que despierten interés. [Completado]
- Selección de datasets (2-3 días): selección y descarga de datasets destacados, [Completado]
- Prueba primer proyecto (5 días): Preparar y ejecutar el proyecto con la intención de tener el primer contacto, seguido de un pequeño test con los primeros resultados. [Completado]
- Prueba segundo proyecto (5 días): Preparar y ejecutar el proyecto con la intención de tener el primer contacto, seguido de un pequeño test con los primeros resultados. [Completado]
- Prueba tercer proyecto (5 días): Preparar y ejecutar el proyecto con la intención de tener el primer contacto, seguido de un pequeño test con los primeros resultados. [Fallido]
- Primeros resultados (5 días): Análisis y comparación de los resultados. [Completado]
- Tiempo restante para informe, contratiempos o avanzar.

Tanto la selección de proyectos y datasets, se escogen de la plataforma Github.

A.1.3. Tercer periodo

El tercer periodo consiste en experimentar con los diferentes proyectos y datasets y explayar el análisis de resultados. Por otra parte se hará un análisis profundo de uno de los proyectos para preparar una estrategia de modificación de cara al siguiente periodo. La idea es alternar estas dos tareas de manera que podamos dividir la mitad del periodo para cada una, pero ir avanzando de manera paralela.

- Pruebas y análisis de resultados (2 semanas): La idea es escoger un conjunto de imágenes y ver que resultados ofrecen los diferentes proyectos estudiados, seguido de un análisis de estos. [Completado parcialmente]
- Estudio profundo de un modelo (2 semanas): Estudio de documentación y código de uno de los proyectos. [Completado]

Las primeras pruebas se harán como se ha dicho anteriormente en Colab, debido al tiempo de ejecución y la potencia necesaria para un dataset completo. Una vez preparadas las pruebas a realizar, se harán con acceso remoto a un ordenador más potente.

A.1.4. Cuarto periodo

El cuarto periodo está dedicado a la modificación i/o mejora de uno de los proyectos estudiados.

- Estudio profundo (5 días): Continuación de la tarea del anterior periodo. [Completado]
- Brain Storming de cambios (2 días): Lluvia de ideas que se puedan llevar a cabo de modificación [Completado]
- Modificación y análisis de resultados (3 semanas): Modificar el modelo, entrenarlo y mostrar los resultados que devuelve con un análisis. [Completado]

A.1.5. Quinto periodo

Finalmente el quinto y último periodo servirá para terminar de pulir el informe final y preparar una presentación del trabajo de manera paralela

- Continuación de periodo anterior (si hace falta) (2 semanas) [Completado]
- Pulir informe final (1 semana) i Preparación de presentación (1 semana) [Completado]

A.2. Muestra más completa Dataset of Multi-Illumination Images in the Wild

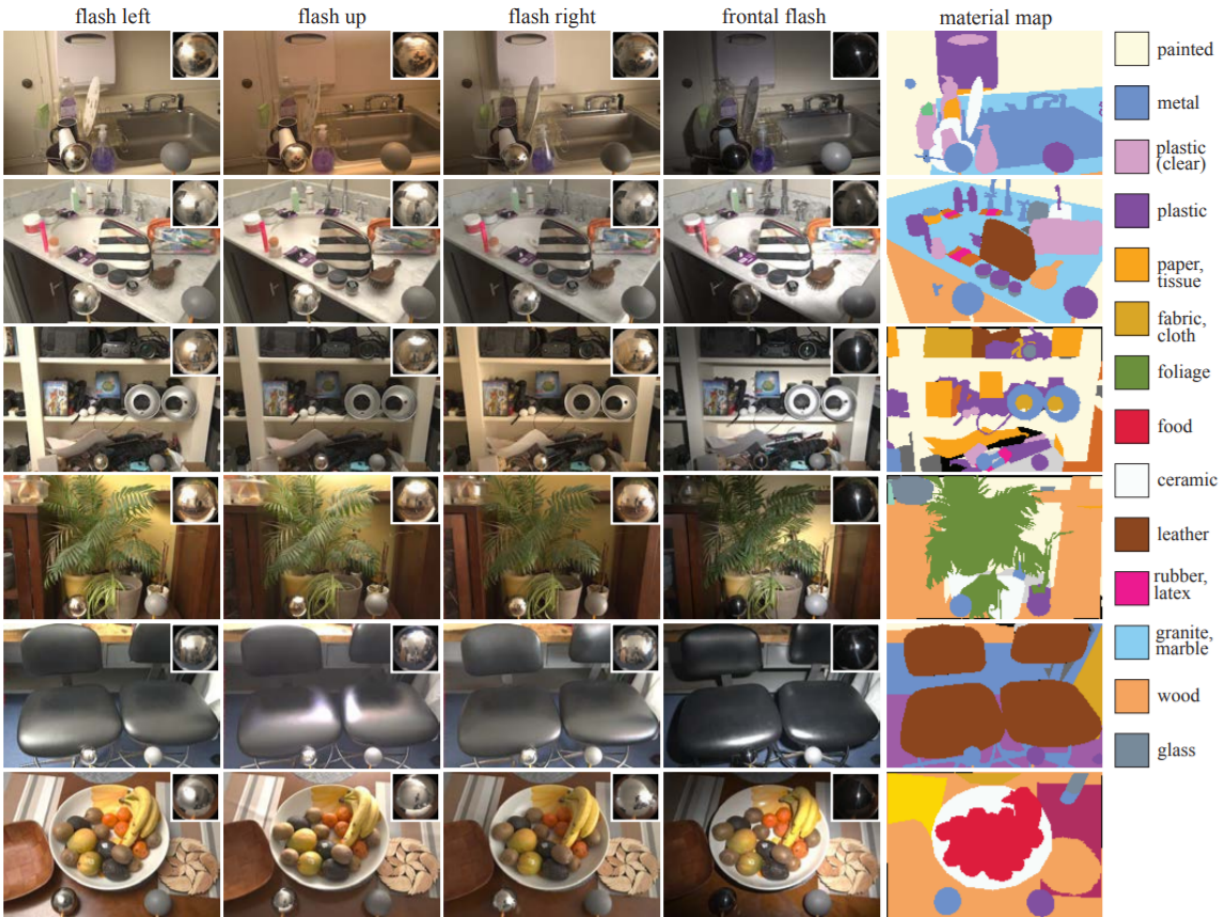

Fig. 14: Muestra de MIL dataset

A.3. Muestra más completa VIDIT

Fig. 15: Muestra de VIDIT dataset

A.4. Muestra más completa Deep Single Image Portrait Relighting

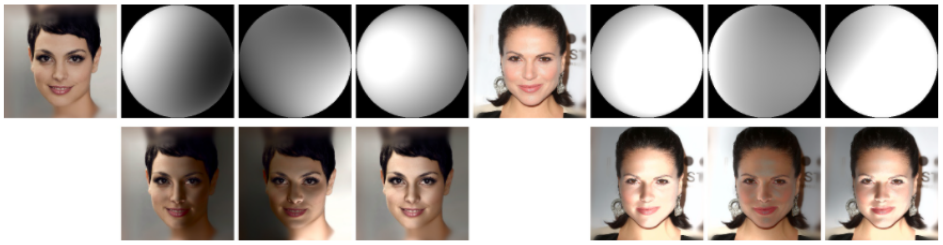

Fig. 16: Muestra de DPR dataset