

Aprendiendo a restaurar imágenes con poca luz mediante descomposición y mejora

Ke Xu^{1,2} Xin yang^{1,†} Baocai Yin^{1,3} Rynson WH Lau^{2,†}

¹Universidad Tecnológica de Dalian

²Universidad de la ciudad de Hong Kong

³Laboratorio de Pengcheng

Abstracto

Las imágenes con poca luz suelen sufrir dos problemas. Primero, tienen baja visibilidad (es decir, valores de píxel pequeños). En segundo lugar, el ruido se vuelve significativo e interrumpe el contenido de la imagen debido a la baja relación señal-ruido. Sin embargo, la mayoría de los métodos existentes de mejora de imágenes con poca luz aprenden de conjuntos de datos con ruido insignificante. Confían en que los usuarios tengan buenas habilidades fotográficas para tomar imágenes con poco ruido. Desafortunadamente, este no es el caso para la mayoría de las imágenes con poca luz. Si bien mejorar una imagen con poca luz y eliminar el ruido al mismo tiempo no está bien planteado, observamos que el ruido exhibe diferentes niveles de contraste en diferentes capas de frecuencia, y es mucho más fácil detectar el ruido en la capa de baja frecuencia que en la alta. Inspirándonos en esta observación, proponemos un modelo de mejora y descomposición basado en la frecuencia para la mejora de imágenes con poca luz. Con base en este modelo, presentamos una red novedosa que primero aprende a recuperar objetos de imagen en la capa de baja frecuencia y luego mejora los detalles de alta frecuencia en función de los objetos de imagen recuperados. Además, hemos preparado un nuevo conjunto de datos de imágenes con poca luz y ruido real para facilitar el aprendizaje. Finalmente, hemos llevado a cabo extensos experimentos para demostrar que el método propuesto supera los enfoques más avanzados en la mejora de imágenes prácticas ruidosas con poca luz.

1. Introducción

Las imágenes con poca luz son muy populares, para varios propósitos, por ejemplo, vigilancia nocturna e imágenes de paisajes personales al atardecer. Sin embargo, la visibilidad de las imágenes con poca luz en el espacio RGB estándar (sRGB, 24 bits/píxel) no coincide con la percepción humana debido a la cuantificación. Esta baja visibilidad dificulta las tareas de visión (*p.ej.*, detección de objetos [31] y seguimiento [8]), o tareas de edición de imágenes (*p.ej.*, imagen mate [45]). Por lo tanto, la recuperación de imágenes con poca luz es fundamental.

Métodos típicos de mejora de imágenes [46,51,24,7,40,34,48,4] proponen recuperar imágenes con poca luz para que coincidan con la percepción humana. Estos métodos se basan en que los usuarios tengan buenas habilidades fotográficas para tomar imágenes con poco ruido, por lo que estos métodos pueden enfocarse en aprender a manipular

[†]Xin Yang y Rynson Lau son los autores correspondientes. Rynson Lau dirigió este proyecto.

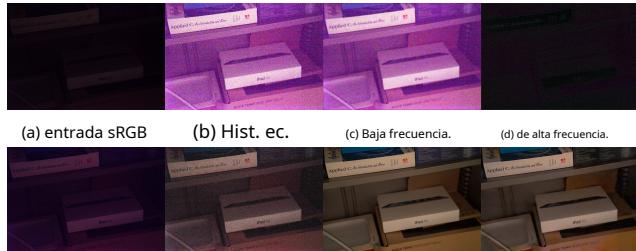


Figura 1. Dada una imagen sRGB con poca luz y profundidad de color de 24 bits (a), los métodos de mejora típicos no pueden producir una imagen agradable con detalles recuperados y ruido suprimido (b, e, f). Para ilustrar nuestra idea, aplicamos un filtro gaussiano para descomponer (b) en una capa de baja frecuencia (c) y una capa de alta frecuencia (d), y observamos que la capa de baja frecuencia conserva suficiente información para recuperar objetos y colores, que luego se pueden usar para mejorar los detalles de alta frecuencia. Esto nos inspira a aprender un método de descomposición y mejora para imágenes con poca luz (h).

los tonos, colores o contrastes de las imágenes. Como tal, no se pueden utilizar para mejorar la mayoría de las imágenes prácticas con poca luz con ruido, que son tomadas por usuarios ocasionales. Cifra1 muestra un ejemplo, donde el contenido de la imagen no solo queda oculto por los valores bajos de intensidad de píxeles, sino que también se ve interrumpido por el ruido, debido a la baja relación señal-ruido (SNR) inherente con poca luz [6]. Los métodos de mejora existentes pueden mejorar tanto el ruido como los detalles de la escena (Figura1(b, f)), o no logran recuperar la baja visibilidad de las imágenes con poca luz (Figura 1(mi)). Además, estas imágenes mejoradas aún tienen SNR bajas, lo que proporciona información contextual útil limitada para detectar el ruido de los detalles de la escena. Por lo tanto, fallan en los métodos existentes de eliminación de ruido [11,49,50,27,37,32,19].

En este documento, abordamos el problema de mejora de imagen sRGB con poca luz, que involucra dos cuestiones: mejora de imagen y eliminación de ruido. Nuestra motivación se basa en dos observaciones. Primero, la capa de baja frecuencia de la imagen conserva más información, *p.ej.*, objetos y colores, y se ve menos afectado por el ruido (Figura1(c)) que la capa de alta frecuencia de la imagen (Figura1(d)). Esto sugiere que es más fácil mejorar la capa de imagen de baja frecuencia que mejorar directamente la imagen completa. En segundo lugar, la dimensionalidad intrínseca muy baja de las primitivas de imagen hace posible que las redes neuronales aprendan un conocimiento completo de las primitivas de imagen [29,41]. Por lo tanto, dada la información de baja frecuencia

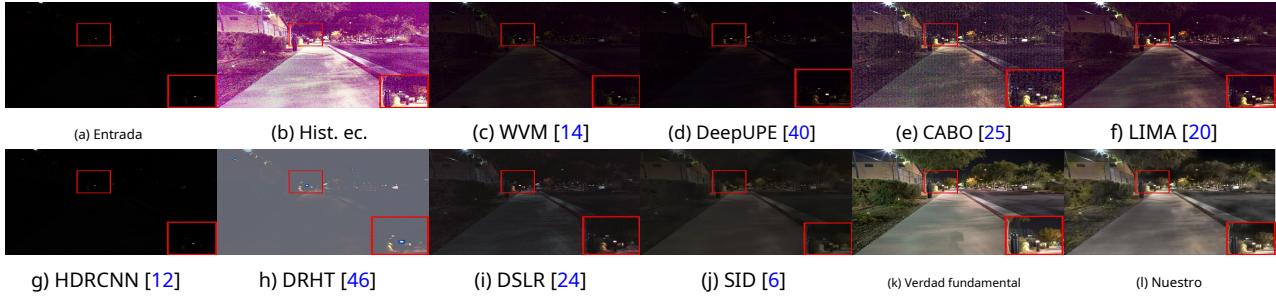


Figura 2. Si bien los métodos existentes ((c) a (j)) generalmente no logran mejorar la imagen de entrada con poca luz y ruido (a), nuestro método produce un resultado más nítido y claro con objetos y detalles recuperados (l).

ción de primitivas, es posible que una red reconstruya todas las primitivas infiriendo la información de alta frecuencia correspondiente. Con esa información previa, podemos aprender a mejorar los detalles de alta frecuencia a partir de la capa de baja frecuencia recuperada.

Estos dos conocimientos nos inspiran a aprender un modelo de mejora y descomposición de imágenes con poca luz basado en la frecuencia. Con este fin, proponemos una nueva red neuronal que aprovecha un módulo de codificación de atención al contexto (ACE) para seleccionar de forma adaptativa información de baja frecuencia para recuperar la capa de baja frecuencia y la eliminación de ruido en la primera etapa, y seleccionar información de alta frecuencia para mejorar los detalles en la segunda etapa. También proponemos un módulo de transformación de dominios cruzados (CDT) para aprovechar las funciones basadas en frecuencias de múltiples escalas para la supresión de ruido y la mejora de detalles en las dos etapas. Como se muestra en la figura 2, nuestro método puede mejorar la imagen sRGB ruidosa con poca luz con contenido/detalles recuperados y supresión de ruido.

En resumen, las principales contribuciones de este trabajo son:

1. Proponemos un nuevo modelo de mejora y descomposición basado en frecuencias para mejorar imágenes con poca luz. Primero recupera el contenido de la imagen en la capa de baja frecuencia mientras suprime el ruido y luego recupera los detalles de la imagen de alta frecuencia.
2. Proponemos una red, con un módulo de codificación de atención al contexto (ACE) para descomponer la imagen de entrada para mejorar de forma adaptativa las capas de alta/baja frecuencia y un módulo de transformación de dominio cruzado (CDT) para la supresión de ruido y mejora de detalles.
3. Preparamos un conjunto de datos de imágenes con poca luz con ruido real e imágenes reales correspondientes, para facilitar el proceso de aprendizaje.

Extensos experimentos verifican el rendimiento superior de el método propuesto sobre los enfoques del estado del arte.

2. Trabajo relacionado

Mejora de la imagen con poca luz. Una línea de métodos mejora las imágenes con poca luz utilizando diferentes funciones de regresión de imagen a imagen. Representado por la ecualización del histograma [36] y corrección gamma, contraste global y local

Se proponen operadores de mejora basados en la detección de regiones semánticas (*p.ej.*, cara y cielo) [25], plantillas de regiones coincidentes [23] o estadísticas de contraste en los límites de la imagen y las regiones texturizadas [38]. Los métodos avanzados basados en el aprendizaje profundo aprenden las funciones de mapeo a partir de imágenes retocadas por el usuario de alta calidad o imágenes tomadas con cámaras de alta gama, utilizando el aprendizaje bilateral [15], supervisión HDR intermedia [46], aprendizaje contradictorio [24,7], o aprendizaje por refuerzo [34,48]. Otra línea de trabajo son los métodos de mejora de imágenes basados en retinex [20,14,51,5,40,47], que descomponen la imagen de entrada con poca luz en iluminación y reflectancia, y luego mejoran la iluminación de la imagen.

Sin embargo, los métodos de mejora existentes pueden fallar al recuperar imágenes con poca luz, debido a sus bajas SNR, como se muestra en la Figura 2. La razón clave es que estos métodos [24,34,7,48,46] generalmente supone que las imágenes serán tomadas por expertos en fotografía con niveles de ruido insignificantes. Por lo tanto, no pueden mejorar las imágenes ruidosas con poca luz.

Recientemente, también hay algunos métodos de mejora [6,22] propuso retocar directamente los datos sin procesar de la cámara en imágenes de salida de alta calidad. En particular, Cheny otros. [6] propuso aprender modelos raw-to-image para generar imágenes mejoradas y suprimidas de ruido a partir de imágenes ruidosas en bruto. Sin embargo, los modelos entrenados en el dominio sin formato no se pueden aplicar a imágenes sRGB normales, que es el espacio de color más adoptado [10], ya que los datos brutos lineales son significativamente diferentes de los datos sRGB no lineales [44]. Además, los datos sin procesar generalmente no están disponibles debido a la falta de experiencia o protocolos desconocidos. En este documento, nos enfocamos en mejorar las imágenes sRGB ruidosas con poca luz.

Eliminación de ruido de la imagen. La eliminación de ruido de una sola imagen es un tema de investigación activo en la visión por computadora y, a menudo, funciona como pre o posprocesamiento para otras tareas de visión. Se han desarrollado muchos métodos basados en imágenes previas, como la autosimilitud [3,11], escasez [13,30], y rango bajo [18,43]. El aprendizaje profundo también se ha aplicado ampliamente al problema de eliminación de ruido [33,49,50,27,37,32]. Estos eliminadores de ruido generalmente aprendieron de conjuntos de datos sintéticos que asumieron ruido aditivo, blanco o gaussiano. A menudo no eliminan el ruido real, que muestra patrones diferentes. Trabajos recientes intentaron mejorar el rendimiento de los eliminadores de ruido en la eliminación de ruido real.

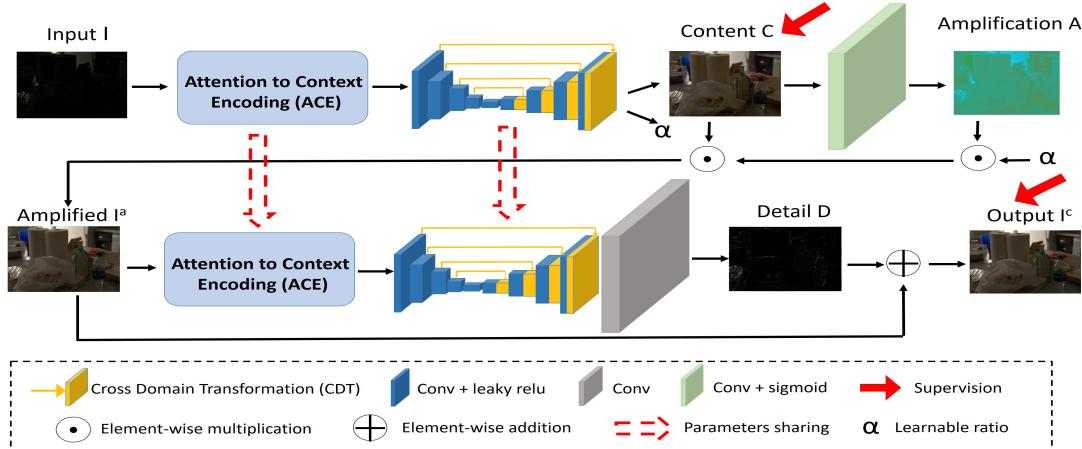


Figura 3. Visión general del modelo propuesto. En la primera etapa, la red mejora los contenidos de baja frecuencia de la imagen de entrada con el ruido suprimido, y luego lo amplifica para producir I^a . En la segunda etapa, la red infiere los detalles de alta frecuencia de I^a para producir la imagen mejorada de salida I^c .

imágenes, sintetizando el ruido en el dominio de datos sin procesar [2], construyendo un conjunto de datos de imagen real [1], desarrollando una estrategia conjunta de entrenamiento de imágenes tanto sintéticas como reales [19], o aprendizaje no supervisado [28].

Sin embargo, no es trivial eliminar el ruido de las imágenes con poca luz simplemente preprocesando o posprocesando con los métodos de eliminación de ruido existentes. Por un lado, los valores de píxel bajos dificultan proporcionar suficiente información contextual para detectar/eliminar el ruido antes de mejorar las imágenes con poca luz. Por otro lado, el ruido puede amplificarse de manera impredecible después de aplicar los métodos de mejora existentes, lo que produce imágenes que aún tienen SNR bajas y, por lo tanto, es difícil eliminar más el ruido. Para abordar esta limitación, proponemos en este documento aprender un modelo de mejora profunda para mejorar las imágenes con poca luz mientras elimina el ruido, de manera recurrente de extremo a extremo.

3. Modelo Propuesto

Nuestro método está inspirado en dos observaciones. En primer lugar, es más fácil mejorar la capa de baja frecuencia de una imagen ruidosa con poca luz, en comparación con mejorar directamente toda la imagen. Esto se debe a que el ruido en la capa de baja frecuencia es más fácil de detectar y luego suprimir. La iluminación/los colores de la imagen se pueden estimar correctamente analizando las propiedades globales de la capa de baja frecuencia de la imagen. En segundo lugar, se sabe que las partes primitivas de las imágenes naturales, *p.ej.*, bordes y esquinas, tienen una dimensionalidad intrínseca muy baja [29]. Tal baja dimensionalidad implica que un pequeño número de ejemplos de imágenes son suficientes para representar bien las primitivas de la imagen [41]. Por lo tanto, dada la información de baja frecuencia de las primitivas, podemos inferir la correspondiente información de alta frecuencia.

Con base en estas dos observaciones, nuestro modelo propuesto, como se muestra en la Figura 3, tiene dos etapas principales. En la primera etapa, proponemos aprender una mejora de imagen de baja frecuencia

función $C(\cdot)$, y luego una función de amplificación $A(\cdot)$ para la recuperación del color. Al modelar conjuntamente el mapeo de $C(\cdot)$ a $A(\cdot)$, la red no tiene que aprender tanto la información global (*p.ej.*, iluminación) e información local (*p.ej.*, color) al mismo tiempo, lo que resulta en una mejora más efectiva. Formalmente, dada una imagen sRGB con poca luz I , la mejora de la primera etapa se puede escribir como:

$$I_a = \alpha A(C(I)) \cdot C(y_o), \quad (1)$$

dónde I_a es la capa amplificada de baja frecuencia. Tenga en cuenta que A es diferente del mapa de iluminación en métodos basados en retinex, ya que estimamos un mapa de amplificación relativa a una proporción global aprendible α del contenido mejorado C . En otras palabras, $\alpha A(\cdot)$ puede interpretarse como un mapa de errores que mejoran C en la forma de autoatención.

En la segunda etapa, proponemos aprender la función de mejora de detalles de alta frecuencia $D(\cdot)$, residencia en I_a desde la primera etapa, en lugar de restaurar directamente los detalles de alta frecuencia de la imagen de entrada original I , que es ruidoso. $D(\cdot)$ luego se modela de manera residual, y la imagen mejorada final se puede obtener como:

$$I_c = I_a + D(y_o). \quad (2)$$

Cifra 4 visualiza la salida de cada paso de nuestro modelo.

Nuestro modelo utiliza dos módulos novedosos, el módulo Atención a la codificación contextual (ACE) y el módulo Transformación entre dominios (CDT). Se explican a continuación.

3.1. Módulo ACE

El objetivo del módulo ACE es aprender funciones sensibles a la frecuencia para la descomposición de imágenes. Para hacer esto, extendemos la operación no local [42], propuesto originalmente para codificar relaciones de largo alcance, para seleccionar información contextual adaptativa de frecuencia. Cifra 5 muestra el diagrama de bloques.

Usamos el primer módulo ACE en la Figura 3 para la explicación. Dadas las características de entrada $X_{\text{en}} \in \mathbb{R}^{H \times W \times C}$, primero usamos dos



Figura 4. La visualización interna (dh) verifica la efectividad del modelo propuesto, frente a la regresión ingenua de imagen a imagen (c).

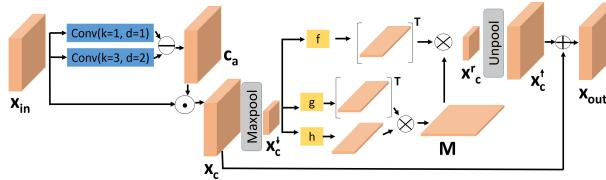


Figura 5. Descripción general del módulo ACE propuesto. Su objetivo es descomponer la imagen en capas basadas en frecuencia para la mejora adaptativa en las dos etapas.

grupos de circunvoluciones dilatadas (con tamaño de núcleo/tasa de dilatación de 1/1y3/2),denotado como F_{d1} y F_{d2} , para extraer características en diferentes campos receptivos. Luego calculamos un mapa de atención consciente del contraste C_a entre estas dos características como:

$$C_a = \text{sigmoide} (f_{d1}(X_{en}) - F_{d2}(X_{en})). \quad (3)$$

C_a indica la información de contraste relativo por píxel, donde los píxeles de alto contraste se consideran pertenecientes a la capa de alta frecuencia. Luego calculamos el mapa inverso $C_a=1 - C_a$ para seleccionar características de X_{en} para representar los contenidos de baja frecuencia como: $X_c = C_a \cdot X_{en}$. Reducimos aún más las características seleccionadas X_c a través de max-pooling para obtener compacto características X_c^t para reducir la memoria GPU y el cálculos para establecer la dependencia no local de píxel a píxel. formalmente dado $X_c \in R^{H \times W \times C}$, el contexto no local proceso de codificación se puede escribir como:

$$X_c = g(X_{en}, C_a) \times h(X_{en}, C_a) \times f(X_{en}, C_a)^T, \quad (4)$$

dónde g , h , f representar grupos de operaciones (convolución, remodelación y transposición de matriz) que primero calculan una tabla de afinidad de píxeles $METRO \in R^{H \times W \times H \times W}$ luego calcular características no mejoradas localmente X_r considerando la relaciones de cada píxel a todos los demás píxeles. Finalmente, obtenemos las características mejoradas no localmente conscientes de la frecuencia $X_{afuera}=Desagrupar(x_c)+X_{de}$ manera residual para facilitar el aprendizaje. proceso de ing Tenga en cuenta que los dos módulos ACE en la Figura 3 compartir sus pesos. El segundo módulo ACE utiliza el mapa de atención sensible al contraste C_a , en lugar del mapa inverso C_a , para conocer los detalles de la imagen a partir de las entidades que representan la capa de alta frecuencia. Cifra 6muestra dos mapas de atención ACE (C_a desde la primera etapa y C_{ade} de la segunda etapa) y sus correspondientes mapas de características descompuestas (X_c desde la primera etapa y X_{de} de la segunda etapa).

3.2. Módulo CDT

Una buena comprensión de las propiedades globales de las imágenes con poca luz puede ayudar a recuperar la iluminación y el contenido de la imagen. Para ello, proponemos el módulo CDT, como se muestra

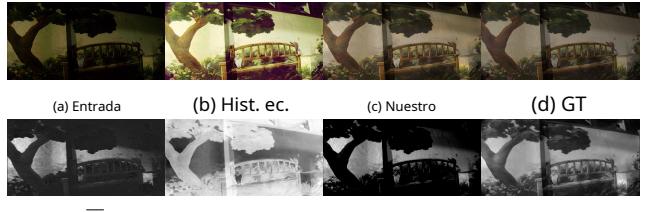


Figura 6. Ejemplo visual de mapas de atención en el módulo ACE de dos etapas y los mapas de características descompuestos. C_a (1etapa) tiende a resaltar las regiones de fondo, mientras que C_a (2etapa) presta más atención a los objetos de primer plano para reconstruir detalles de alta frecuencia.

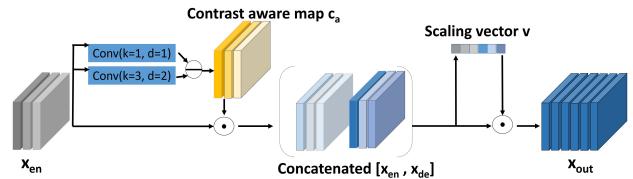


Figura 7. Vista general del módulo CDT propuesto. Su objetivo es aumentar los campos receptivos mientras se cierra la brecha entre el dominio de poca luz y el dominio mejorado.

En figura 7, para aumentar los campos receptivos mientras se cierra la brecha entre las características en el dominio de poca luz y en el dominio mejorado. Compartiendo un espíritu similar a [39] al aumentar los campos receptivos para una información más global, el módulo CDT está especialmente diseñado para ad- resolver el problema de la brecha de dominio,*es decir*, función consciente de la frecuencia turas extraídas en el dominio ruidoso con poca luz versus aquellas en el dominio mejorado.

Específicamente, en la primera etapa, las características ruidosas del codificador X_{es} primero se vuelven a pesar espacialmente a través del mapa de contraste inverso autoderivado C_a para filtrar información de alto contraste, antes de concatenar con características $X_{Delaware}$ del decodificador correspondiente. Luego calculamos vectores de escala global de las características concatenadas $[X_{es}, X_{Delaware}]$, para volver a escalar de forma adaptativa las características de diferentes dominios en forma de canal. En la segunda etapa, utilizamos el mapa de atención consciente del contraste C_a , en lugar del mapa inverso C_a , para conocer los detalles de la imagen, similar al módulo ACE.

3.3. Conjunto de datos propuesto

Para facilitar el aprendizaje del modelo propuesto, hemos preparado un nuevo conjunto de datos con poca luz de pares de imágenes sRGB reales con poca luz y con ruido real.

Ruido en condiciones de poca luz.Preparamos nuestros datos de entrenamiento basados en el conjunto de datos SID [6], que consta de datos sin procesar y pares de imágenes reales. Estos datos sin procesar fueron recopilados

cuando se toman imágenes con poca luz con un tiempo de exposición corto (normalmente 0,1 s o 0,04 s). Sus correspondientes imágenes reales del terreno se tomaron con un tiempo de exposición prolongado (típicamente 10 s o 30 s), donde el ruido es insignificante. Sin embargo, los datos sin procesar de la cámara lineal son significativamente diferentes de los datos sRGB no lineales, particularmente en términos de ruido [2] e intensidad de la imagen [46]. Como resultado, los modelos entrenados con datos sin procesar no se pueden aplicar directamente a las imágenes sRGB. Para abordar este problema, hemos considerado varios pasos clave (*es decir*, compensación de exposición, balance de blancos y deslinealización) en el canal de formación de imágenes, y manipulado sus operaciones para modelar imágenes sRGB ruidosas con poca luz del mundo real tomadas de diferentes cámaras.

Compensación de exposición. Los algoritmos de exposición automática tienen como objetivo determinar automáticamente el tiempo de exposición y la ganancia de la cámara en función de la intensidad de la luz percibida por el sensor. Por lo general, son cajas negras y varían según las cámaras. Para aumentar la diversidad de este tiempo de exposición, muestreamos aleatoriamente el valor de compensación de exposición del rango de [0 vehículo eléctrico, 2VE] a intervalos de 0.5VE

Balance de blancos. Los algoritmos de balance de blancos tienen como objetivo corregir moldes poco realistas mediante la estimación de la ganancia por canal [dieciséis]. También son desconocidos y varían entre cámaras. Lo aumentamos eligiendo aleatoriamente la temperatura de color del rango de [2100K, 4000K], que representa las temperaturas de color de la iluminación doméstica típica y la iluminación del amanecer/atardecer, de acuerdo con la tabla de colores de temperatura Kelvin [9].

Delinealización. Como la no linealidad introducida por la función de respuesta de la cámara varía entre cámaras y es difícil aplicar ingeniería inversa [17], aplicamos la función gamma como función de deslinealización, como se sugiere en [12].

Usando la configuración anterior, hemos producido un total de 4198 pares de imágenes para entrenamiento y 1196 pares de imágenes para prueba. Resultados experimentales en cifras 9 y 10 muestran que la red propuesta entrenada en nuestros datos puede generalizarse bien en imágenes de otras tuberías de formación de imágenes.

3.4. Capacitación

Función de pérdida. Usamos la pérdida de L2 para medir la precisión de la reconstrucción en el proceso de entrenamiento en dos etapas. Específicamente, en la primera etapa, para alentar a nuestra red a enfocarse en predecir los componentes de baja frecuencia de la imagen de entrada, preparamos la verdad de campo correspondiente, denotada como \mathbf{y}_{gt} usando el filtro guiado [21] para filtrar la alta detallado de frecuencia manteniendo las principales estructuras y contenidos de la imagen de verdad del terreno. Normalmente, el reconocimiento

La pérdida de construcción se puede escribir como:

$$L_{\text{cuenta}} = \lambda_1 \|\mathbf{C} - \mathbf{y}_{\text{gt}}\|_2 + \lambda_2 \|\mathbf{I}_{\text{c}} - \mathbf{I}_{\text{gt}}\|_2 \quad (5)$$

dónde \mathbf{C} , \mathbf{y}_{gt} , \mathbf{I}_{c} , \mathbf{I}_{gt} son el contenido de la imagen reconstruida, la imagen recuperada, verdad fundamental de la capa de baja frecuencia

er, y verdad básica de la imagen mejorada, respectivamente. λ_1 y λ_2 son parámetros de equilibrio.

También incorporamos la pérdida de percepción al comparar las distancias de características VGG de \mathbf{I}_{c} y \mathbf{I}_{gt} , utilizando la pérdida L1, como:

$$\| \mathbf{L}_{\text{vgg}} = \lambda_3 \|\Phi(\mathbf{I}_{\text{c}}) - \Phi(\mathbf{I}_{\text{gt}})\|_1 \| \quad (6)$$

donde Φ es la red VGG y λ_3 es un parámetro de equilibrio.

4. Experimentos

Hemos implementado el modelo propuesto en el framework Pytorch [35], y lo probé en una PC con una CPU i7 de 4 GHz y una GPU GTX 1080Ti. A medida que entrenamos nuestro modelo desde cero, los parámetros de la red se inicializan aleatoriamente, excepto la relación de amplificación aprendible a , que se inicializa en 1. Estrategias de aumento estándar, *es decir*, se adoptan la escala, el recorte y el volteo horizontal. Durante el entrenamiento, recortamos aleatoriamente parches de resoluciones 512×384 de las imágenes escaladas de resolución 2048×1536. Para la minimización de pérdidas, adoptamos el optimizador ADAM [26] durante 400 épocas, con una tasa de aprendizaje inicial de 3 mi-4 dividido por 10 en la época 250. λ_1 , λ_2 y λ_3 se establecen en 1, 1 y 0,1, respectivamente. Se necesita 0.33 para que la red propuesta procese una imagen de resoluciones 1024×768.

Para evaluar el rendimiento del método propuesto en la mejora de imágenes con poca luz, comparamos cuantitativa y visualmente nuestro método con 9 métodos de mejora de última generación con códigos disponibles, incluido JieP [5], CAL [20], WVM [14], reflex digital [24], CAPA [25], DRHT [46], DeepUPE [40], HDRCCNN [12] y SID [6]. Usamos PSNR y SSIM para la medición cuantitativa.

4.1. Comparando con el Estado de las Artes

Comparaciones visuales. Primero comparamos visualmente los resultados del método propuesto con los métodos de mejora de imágenes de última generación. Cifra 8 muestra los resultados de diferentes métodos en tres imágenes de entrada con poca luz (a, m, A), que fueron tomadas por una cámara Sony. Podemos ver que WVM [14] y DeepUPE [40] no mejoran estas imágenes (c, d, o, p, C, D). Dado que se basan en descomponer la imagen de entrada en reflectancia e iluminación, cuando una imagen de entrada tiene poca luz, no pueden descomponerla con precisión. CAL [20] puede mejorar las imágenes (f, r, F), ya que estima directamente el mapa de iluminación sin descomponer la imagen de entrada. Sin embargo, mejora tanto los detalles como el ruido juntos. De manera similar, el método basado en la corrección gamma CAPE [25] también mejora conjuntamente los detalles y el ruido juntos (e, q, E). DRHT [46] no logra mejorar las imágenes ruidosas con poca luz (h, t, H), ya que el ruido puede deteriorar tanto la reconstrucción HDR como los procesos de mapeo de tonos. DSLR [24] está entrenado para retroceder una imagen de baja calidad a una de alta calidad. Si bien puede mejorar un poco las imágenes, no elimina el ruido (i, u, I). Dado que el SID original [6] modelo

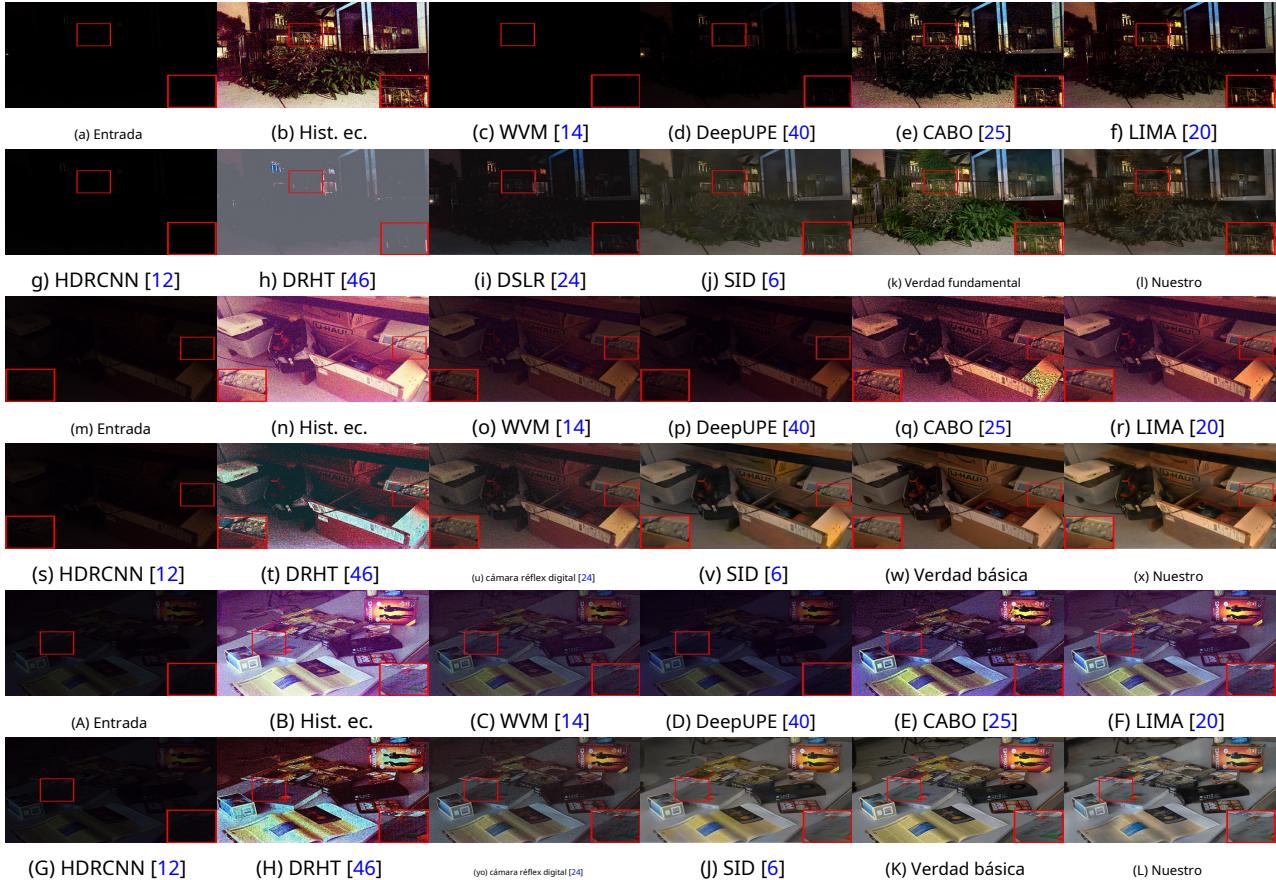


Figura 8. Resultados visuales de métodos de última generación y los nuestros en imágenes de entrada con poca luz (a, m, A). Los cuadros rojos indican las regiones ruidosas donde fallan la mayoría de los métodos existentes. Las imágenes de entrada fueron tomadas por una cámara Sony.

(entrenado en el dominio sin procesar) no se puede aplicar directamente a las imágenes sRGB, lo volvemos a entrenar en las imágenes sRGB. Podemos ver que el modelo SID tiende a eliminar el ruido y los detalles, lo que da como resultado imágenes borrosas (j, v, J). Por el contrario, nuestros resultados (l, x, L) muestran que el método propuesto puede mejorar con éxito el contenido y los detalles de la imagen mientras suprime el ruido.

Cifra 9 muestra los resultados de otras tres imágenes de entrada con poca luz (tomadas con la cámara de un iPhone). Si bien los métodos más avanzados generalmente no logran eliminar el ruido y mejorar los contenidos/detalles al mismo tiempo, nuestro método produce resultados visualmente más convincentes, incluso para las imágenes con texturas más desafiantes (l, x). Cifras 8 y 9 demuestran la buena capacidad de generalización del modelo/conjunto de datos propuesto en imágenes tomadas por diferentes tipos de cámaras.

Comparaciones cuantitativas. También hemos comparado cuantitativamente nuestro método con los métodos de mejora más avanzados. Como se muestra en la Tabla 1, el método propuesto supera a estos métodos de mejora existentes por un amplio margen. Tenga en cuenta que también hemos preprocesado las imágenes de entrada antes de enviarlas a dos métodos [14, 5], amplificando estas intensidades de píxeles de imagen con proporciones predefinidas como en [6] o aplicando la ecualización de histogramas. Sin embargo, los resultados son los mismos que aquellos sin preprocesamiento. Este

indica que la mejora de imágenes ruidosas con poca luz mediante la descomposición de imágenes en reflectancia e iluminación no es adecuada. Por el contrario, nuestra descomposición y mejora basada en la frecuencia puede desacoplar con éxito el problema de mejora y eliminación de ruido de la imagen.

También comparamos nuestro método con SID [6], que se propuso originalmente para mejorar imágenes con poca luz en el dominio sin procesar, tanto en dominios sRGB como sin procesar. Específicamente, en el dominio sRGB, aplicamos dos estrategias: usar directamente el modelo SID original entrenado en imágenes sin procesar (indicado como SID) y usar un modelo SID reentrenado en imágenes sRGB en nuestro conjunto de entrenamiento (indicado como SID+). En el dominio sin procesar, volvemos a entrenar nuestro modelo utilizando los datos sin procesar. Podemos ver que nuestro método supera a SID [6] en dominios sRGB y raw. Además, comparamos nuestro método con el método más nuevo [40] tanto en sRGB (reentrenado en nuestro conjunto de datos) como en dominios sin procesar. Estos resultados muestran que nuestro modelo es más eficaz en la mejora de imágenes con poca luz con ruido, que en el aprendizaje directo de imagen a imagen [6] o imagen a iluminación [40] modelos de regresión.

Finalmente, comparamos nuestro método con diferentes combinaciones de métodos existentes de mejora y eliminación de ruido. Específicamente, elegimos un método clásico de eliminación de ruido B-

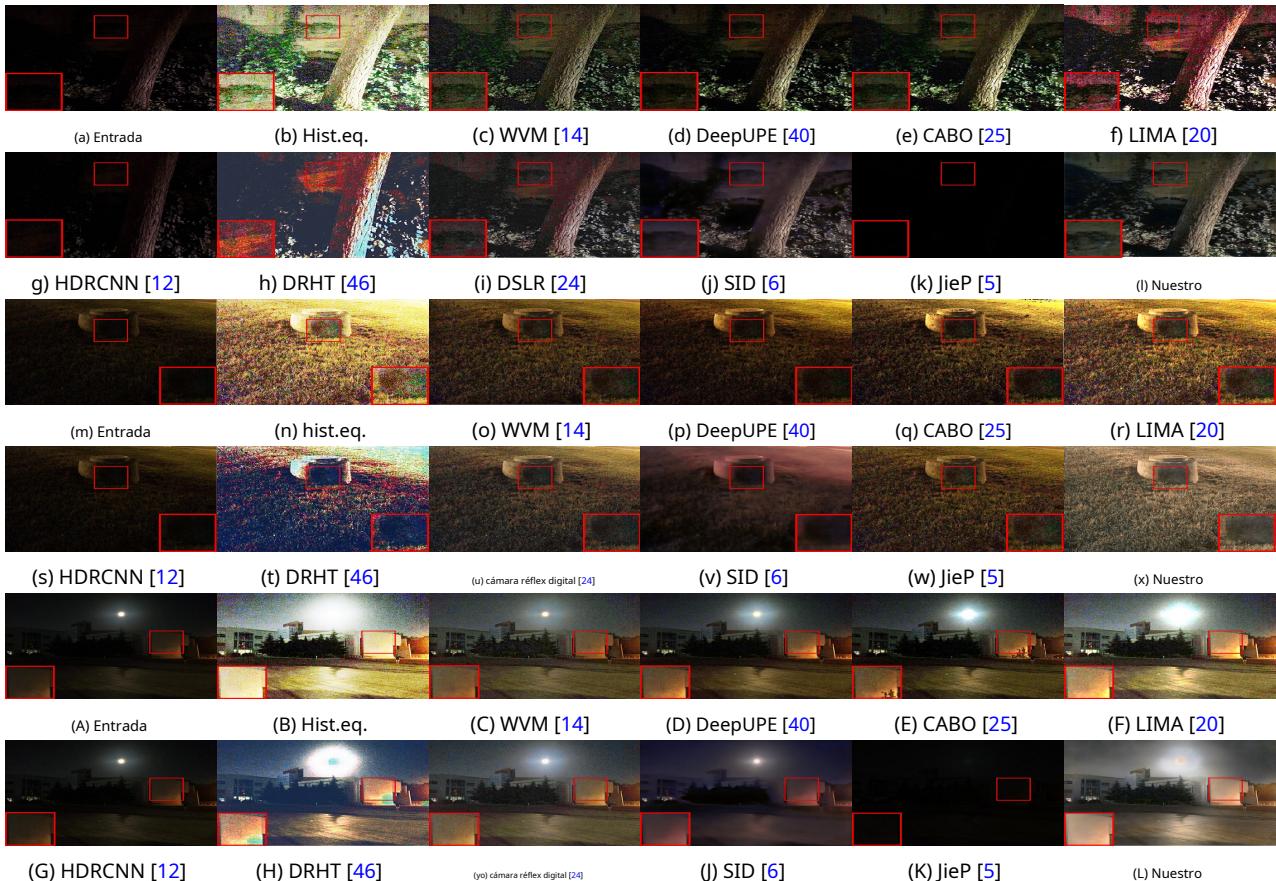


Figura 9. Resultados visuales de métodos de última generación y los nuestros en imágenes de entrada con poca luz (a, m, A). Los cuadros rojos indican las regiones ruidosas donde fallan la mayoría de los métodos existentes. Las imágenes de entrada fueron tomadas por una cámara de iPhone. Resultados de nuestro método aquí y en la Figura 8 demostrar la capacidad de generalización del método en diferentes tipos de cámaras.

M3D [11] y un método reciente de eliminación de ruido basado en aprendizaje profundo xDnCNN [27] para procesar previamente/ posteriormente las imágenes con poca luz (en el conjunto de prueba) antes/ después de que sean procesadas por el método de mejora LIMA [20]. Elegimos LIMA [20] ya que tiene el tercer mejor rendimiento entre los métodos existentes en la Tabla 1. Aunque SID [6] y DeepUPE [40] tienen un mejor rendimiento, ya están capacitados en nuestro conjunto de datos para eliminar el ruido. Por lo tanto, no los usamos aquí. Mesa 2 muestra los resultados. Podemos ver que la aplicación directa de los métodos de eliminación de ruido existentes como un paso de procesamiento previo/ posterior a los métodos de mejora no funciona bien. Como el ruido ya está profundamente enterrado en el contenido y los detalles de la imagen en las imágenes con poca luz, mejorar y eliminar el ruido por separado de estas imágenes no funciona bien. En su lugar, suprimimos el ruido en la capa de baja frecuencia y luego mejoramos los contenidos y detalles de forma adaptativa, produciendo mejores actuaciones. Cifra 10 muestra algunos ejemplos visuales de la combinación de métodos existentes de mejora y eliminación de ruido. Podemos ver que la eliminación de ruido seguida de la mejora produce resultados borrosos (e, f), debido a la eliminación significativa de detalles de la imagen en el paso de eliminación de ruido. Aunque la mejora seguida de la eliminación de ruido puede producir resultados relativamente más nítidos (g, h) en comparación

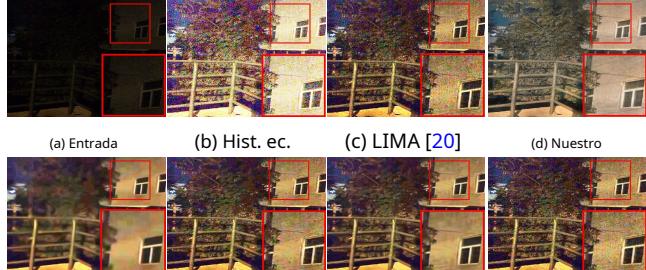


Figura 10. Comparación con diferentes combinaciones de LIME [20] y dos métodos de eliminación de ruido (BM3D [11] y xDnCNN [27]). “X” indica el uso de LIME para el procesamiento posterior, mientras que “+X” indica el uso de LIME para el procesamiento previo. Los cuadros rojos indican las regiones ruidosas donde fallan la mayoría de los métodos existentes.

Son a (e, f), respectivamente, los resultados son más ruidosos ya que tanto el ruido como los detalles se mejoran en el paso de mejora. También es interesante notar que ninguno de estos métodos puede recuperar bien los colores (causados por el ruido), p.ej., el color púrpura del árbol. Por el contrario, nuestro método puede producir una imagen nítida (d), con el ruido suprimido y el color recuperado.

Apunte	Método	PSNR↑	SSIM↑
sRGB	hist. ec.	12,08	0,2236
	CAPA [25]	15,05	0,2306
	JieP [5]	11,93	0,0381
	WVM [14]	11,95	0,0382
	DeepUPE [40]	14,44	0,2208
	DeepUPE* [40]	21,55	0,6531
	DRHT [46]	11,85	0,0969
	HDRCNN [12]	12,64	0,1102
	DSLR [24]	17,25	0,4229
	CAL [20]	17,76	0,3506
	SID [6]	15,35	0,2418
	S.I.D.-[6]	21,16	0,6398
	Nuestro	22,13	0,7172
CRUDO	SID [6]	28,88	0,7870
	DeepUPE [40]	29,13	0,7915
	Nuestro	29,56	0,7991

Tabla 1. Comparación con los métodos de mejora de última generación. El mejor rendimiento está marcado en negrita. Tenga en cuenta que un · indica que el modelo se vuelve a entrenar en nuestro conjunto de entrenamiento sRGB.

Apunte	Método	PSNR↑	SSIM↑
sRGB	CAL [20]	17,76	0,3506
	CAL [20] + BM3D [11]	17,90	0,3610
	CAL [20] + xDnCNN [27]	17,75	0,3511
	BM3D [11] + LIMA [20]	17,41	0,3273
	xDnCNN [27] + LIMA [20]	17,75	0,3511
	Nuestro	22,13	0,7172

Tabla 2. Comparación con diferentes combinaciones de métodos de mejora y eliminación de ruido. El mejor rendimiento está marcado en negrita.

4.2. Análisis interno

Comenzamos estudiando la efectividad del módulo ACE propuesto. Las dos primeras filas de la tabla 3 muestran que quitar el módulo ACE o reemplazarlo por un bloque no local [42] provoca una caída del rendimiento, ya que el ruido ya no se puede filtrar a través de la descomposición de la imagen. Esto verifica la efectividad del módulo ACE propuesto para aprender a seleccionar características beneficiosas y suprimir características dañinas antes de codificar los contextos no locales. Del mismo modo, la eliminación del módulo CDT también provoca una caída del rendimiento, lo que demuestra la importancia de tener grandes campos receptivos mientras se cierra la brecha entre los dominios de poca luz y mejorados. También notamos una caída en el rendimiento causada por la sustitución del mapa con reconocimiento de contraste de los módulos CDT con el módulo ACE, que verifica la necesidad de modelar información multinivel con reconocimiento de contraste para imágenes ruidosas con poca luz. También podemos ver que la incorporación de la pérdida de percepción conduce a mejores resultados, ya que proporciona regularización en el espacio de funciones.

Finalmente, estudiamos las opciones de tubería. Entrenamos a nuestro modelo para que aprenda a mejorar imágenes usando solo una etapa (denominada Disparo único). También entrenamos nuestro modelo usando directamente imágenes reales del terreno para supervisar la salida del primer etapa (denominada como $I_{gt} \rightarrow I_{gt}$), en lugar de usar el suelo verdad de la capa de baja frecuencia. Los resultados se muestran en la

Apunte	Método	PSNR↑	SSIM↑
sRGB	sin ACE	21,34	0,6439
	AS→NL [42]	21,49	0,6477
	sin CDT	21,47	0,6410
	$C_{CDT} \rightarrow C_{AS}$	21,84	0,7006
	sin pérdida de percepción	22,03	0,7033
	Un solo tiro $I_{gt} \rightarrow I_{gt}$	21,63	0,6713
	Nuestro	21,76	0,6874
		22,13	0,7172

Tabla 3. Análisis interno del método propuesto.



Figura 11. Un caso de fracaso. Cuando todos los objetos de la imagen están lejos, es posible que nuestro método, así como los métodos existentes, no puedan seleccionar contextos útiles de las áreas circundantes.

6ely7elflas Muestra la ventaja de aprender un modelo de dos etapas sobre Single Shot. También podemos ver que el uso de datos reales de la capa de baja frecuencia para supervisar la primera etapa produce mejores resultados que el uso de imágenes de datos reales, lo que verifica la importancia de aprender el modelo de descomposición y mejora.

5. Conclusión y Trabajo Futuro

En este artículo, hemos estudiado el problema de mejora de imágenes ruidosas con poca luz. Hemos observado que el ruido afecta a las imágenes de manera diferente en diferentes capas de frecuencia. Sobre la base de esta observación, proponemos un modelo novedoso de descomposición y mejora de imágenes basado en la frecuencia para mejorar de forma adaptativa los contenidos y detalles de la imagen en diferentes capas de frecuencia, mientras que al mismo tiempo suprime el ruido. También presentamos una red con el módulo de codificación de atención al contexto (ACE) propuesto para mejorar de forma adaptativa las capas de alta y baja frecuencia, y el módulo Cross Domain Transformation (CDT) para la supresión de ruido y la mejora de detalles. Para entrenar nuestro modelo, hemos preparado un nuevo conjunto de datos de imágenes con poca luz. Finalmente, hemos realizado extensos experimentos para verificar la efectividad de nuestro método frente a los métodos más avanzados.

Nuestro método tiene limitaciones. Puede fallar en escenas con objetos pequeños, en las que nuestra red puede no ser capaz de extraer información contextual significativa de las áreas circundantes para recuperar los contenidos, como se muestra en la Figura 11. Como trabajo futuro, estamos interesados en ampliar nuestro modelo de mejora para considerar los diseños semánticos de las escenas y usar el aprendizaje generativo antagónico para sintetizar los detalles de la imagen.

Reconocimiento. Este trabajo fue apoyado en parte por subvenciones NNSFC 91748104, 61972067, 61632006, U1811463, U1908214, 61751203; y el Programa Nacional de Investigación y Desarrollo Clave de China, Subvención 2018AAA0102003.

Referencias

- [1] Abdelrahman Abdelhamed, Stephen Lin y Michael Brown. Un conjunto de datos de eliminación de ruido de alta calidad para cámaras de teléfonos inteligentes. En *CVPR*, 2018.3
- [2] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet y Jonathan T. Barron. Desprocesamiento de imágenes para eliminación de ruido en bruto aprendida. En *CVPR*, 2019.3,5
- [3] Antoni Buades, Bartomeu Coll y JM Morel. Un algoritmo no local para la eliminación de ruido de imágenes. En *CVPR*, 2005.2
- [4] Vladimir Bychkovsky, Sylvain Paris, Eric Chan y Frédéric Durand. Aprendizaje del ajuste tonal global fotográfico con una base de datos de pares de imágenes de entrada/salida. En *CVPR*, 2011.1
- [5] Bolun Cai, Xianming Xu, Kailing Guo, Kui Jia, Bin Hu y Dacheng Tao. Un modelo previo conjunto intrínseco-extrínseco para retinex. En *ICCV*, 2017.2,5,6,7,8
- [6] Chen Chen, Qifeng Chen, Jia Xu y Vladlen Koltun. Aprendiendo a ver en la oscuridad. En *CVPR*, 2018.1,2,4,5,6,7,8
- [7] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao y Yung-Yu Chuang. Mejorador de fotos profundas: aprendizaje no emparejado para la mejora de imágenes a partir de fotografías con gans. En *CVPR*, 2018.1,2
- [8] Qi Chu, Wanli Ouyang, Hongsheng Li, Xiaogang Wang, Bin Liu y Nenghai Yu. Seguimiento en línea de múltiples objetos utilizando un rastreador de un solo objeto basado en CNN con mecanismo de atención espacio-temporal. En *ICCV*, 2017.1
- [9] Colaboradores de Wikipedia. Temperatura del color. Disponible de: https://en.wikipedia.org/wiki/Color_temperatura.5
- [10] Colaboradores de Wikipedia. sRGB. Disponible de: <https://en.wikipedia.org/wiki/SRGB.2>
- [11] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik y Karen Egiazarian. Eliminación de ruido de imágenes con coincidencia de bloques y filtrado 3D. En *proc. SPIE*, volumen 6064, 2006.1,2,7,8
- [12] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafa Mantiuk y Jonas Unger. Reconstrucción de imágenes HDR a partir de una sola exposición usando CNN profundas. *TOG ACM*, 2017.2,5,6,7,8
- [13] Michael Elad y Michal Aharon. Eliminación de ruido de imágenes a través de representaciones dispersas y redundantes sobre diccionarios aprendidos. *CONSEJO IEEE*, 2006.2
- [14] Xueyang Fu, Delu Zeng, Yue Huang, Xiaoping Zhang y Xinghao Ding. Un modelo variacional ponderado para la estimación simultánea de reflectancia e iluminación. En *CVPR*, 2016. 2,5,6,7,8
- [15] Michaël Gharbi, Jiawen Chen, Jonathan Barron, Samuel Hasinoff y Frédéric Durand. Aprendizaje bilateral profundo para la mejora de imágenes en tiempo real. En *SIGGRAFÁ*, 2017.2
- [16] A. Gijssenij, T. Gevers y J. van de Weijer. Constancia de color computacional: Encuesta y experimentos. *CONSEJO IEEE*, 2011. 5
- [17] M. Grossberg y S. Nayar. ¿Cuál es el espacio de las funciones de respuesta de la cámara? En *CVPR*, 2003.5
- [18] Shuhang Gu, Lei Zhang, Wangmeng Zuo y Xiangchu Feng. Minimización de la norma nuclear ponderada con aplicación a la eliminación de ruido de la imagen. En *CVPR*, 2014.2
- [19] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo y Lei Zhang. Hacia la eliminación de ruido ciego convolucional de fotografías reales. En *CVPR*, 2019.1,3
- [20] Xiaojie Guo, Yu Li y Haibin Ling. Lime: mejora de la imagen con poca luz a través de la estimación del mapa de iluminación. *CONSEJO IEEE*, 2017.2,5,6,7,8
- [21] Kaiming He, Jian Sun y Xiaou Tang. Filtrado guiado de imágenes. *IEEE TPAMI*, 2013.5
- [22] Yuanming Hu, Hao He, Chenxi Xu, Baoyuan Wang y Stephen Lin. Exposición: un marco de posprocesamiento de fotos de caja blanca. En *SIGGRAFÁ*, 2018.2
- [23] Sung Ju Hwang, Ashish Kapoor y Sing Bing Kang. Mejora de imagen local automática basada en el contexto. En *CE-CV*, 2012.2
- [24] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey y Luc Van Gool. Fotografías con calidad DSLR en dispositivos móviles con redes convolucionales profundas. En *ICCV*, 2017.1,2,5,6,7,8
- [25] Liad Kaufman, Dani Lischinski y Michael Werman. Mejora automática de fotografías según el contenido. *Foro de gráficos por computadora*, 2012.2,5,6,7,8
- [26] P. Kingma y J. Ba. Adam: Un método para la optimización estocástica. *arXiv:1412.6980*, 2014.5
- [27] Idan Kligvasser, Tamar Rott Shaham y Tomer Michaeli. xunit: aprendizaje de una función de activación espacial para la restauración eficiente de imágenes. En *CVPR*, 2018.1,2,7,8
- [28] Alexander Krull, Tim-Oliver Buchholz y Florian Jug. Noise2void: aprendizaje de eliminación de ruido a partir de imágenes ruidosas individuales. En *CVPR*, 2019.3
- [29] Ann Lee, Kim Pedersen y David Mumford. Las estadísticas complejas de parches de alto contraste en imágenes naturales. *SCTV*, 2001.1,3
- [30] Jianwei Li, Xiaowu Chen, Dongqing Zou, Bo Gao y Wei Teng. Representación escasa conforme y de bajo rango para la restauración de imágenes. En *ICCV*, 2015.2
- [31] Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan y Serge Belongie. Cuenta con redes piramidales para la detección de objetos. En *CVPR*, 2017.1
- [32] Ding Liu, Bihang Wen, Yuchen Fan, Chen Change Loy y Thomas Huang. Red recurrente no local para la restauración de imágenes. En *NeurIPS*. 2018.1,2
- [33] Seonghyeon Nam, Youngbae Hwang, Yasuyuki Matsushita y Seonwoo Kim. Un enfoque holístico para el modelado de ruido de imagen de canales cruzados y su aplicación para eliminar el ruido de la imagen. En *CVPR*, 2016.2
- [34] Jongchan Park, Joon-Young Lee, Donggeun Yoo e In So Kweon. Distorsionar y recuperar: mejora del color mediante el aprendizaje de refuerzo profundo. En *CVPR*, 2018.1,2
- [35] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga y Adam Lerer. Diferenciación automática en pytorch. En *Taller NeurIPS*, 2017.5
- [36] Stephen Pizer, E. Philip Amburn, John Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart Ter Haar Romeny y John Zimmerman. Ecualización de histograma adaptativo y sus variaciones. *Visión artificial, gráficos y procesamiento de imágenes*, 1987.2

- [37] Tobias Plötz y Stefan Roth. Redes neuronales de vecinos más cercanos. En *NeurIPS*, 2018.[1,2](#)
- [38] Ramírez Rivera, Byungyong Ryu y O Chae. Mejora de imágenes oscuras según el contenido a través de la división de canales. *CONSEJO IEEE*, 2012.[2](#)
- [39] Olaf Ronneberger, Philipp Fischer y Thomas Brox. U-Net: Redes convolucionales para segmentación de imágenes biomédicas. En *miccai*, 2015.[4](#)
- [40] Wang Ruixing, Zhang Qing, Fu Chiwing, Shen Xiaoyong, Zheng Weishi y Jiaya Jia. Mejora de fotografías subexpuestas utilizando estimación de iluminación profunda. En *CVPR*, 2019.[1, 2,5,6,7,8](#)
- [41] Jian Sun, Nan-Ning Zheng, Hai Tao y Heung-Yeung Shum. Alucinación de imágenes con bocetos primarios previos. En *CVPR*, 2003.[1,3](#)
- [42] Xiaolong Wang, Ross Girshick, Abhinav Gupta y Kaiming He. Redes neuronales no locales. En *CVPR*, 2018.[3,8](#)
- [43] Jun Xu, Lei Zhang, David Zhang y Xiangchu Feng. Minimización de la norma nuclear ponderada multicanal para eliminar el ruido de la imagen en color real. En *ICCV*, 2017.[2](#)
- [44] Xiangyu Xu, Yongrui Ma y Wenxiu Sun. Hacia la superresolución de la escena real con imágenes en bruto. En *CVPR*, 2019.[2](#)
- [45] Xin Yang, Ke Xu, Shaozhe Chen, Shengfeng He, Baocai Yin Yin y Rynson Lau. Esteras activas. 2018.[1](#)
- [46] Xin Yang, Ke Xu, Yibing Song, Qiang Zhang, Xiaopeng Wei y Rynson Lau. Corrección de imagen a través de una profunda transformación HDR recíproca. En *CVPR*, 2018.[1,2,5,6,7, 8](#)
- [47] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang y Wenmin Wang. Un nuevo algoritmo de mejora de imágenes con poca luz que utiliza el modelo de respuesta de la cámara. En *Talleres ICCV*, 2017.[2](#)
- [48] Runsheng Yu, Wenyu Liu, Yasen Zhang, Zhi Qu, Deli Zhao y Bo Zhang. Exposición profunda: aprender a exponer fotos con aprendizaje antagónico reforzado de forma asíncrona. En *NeurIPS*, 2018.[1,2](#)
- [49] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng y Lei Zhang. Más allá de un eliminador de ruido gaussiano: aprendizaje residual de CNN profunda para eliminación de ruido de imágenes. *CONSEJO IEEE*, 2017.[1,2](#)
- [50] Kai Zhang, Wangmeng Zuo, Shuhang Gu y Lei Zhang. Aprendizaje profundo del eliminador de ruido de CNN antes de la restauración de la imagen. En *CVPR*, 2017.[1,2](#)
- [51] Qing Zhang, Ganzhao Yuan, Chunxia Xiao, Lei Zhu y Wei-Shi Zheng. Corrección de exposición de alta calidad de fotos subexpuestas. En *ACM milímetro*, 2018.[1,2](#)