

Estudio comparativo de algoritmos de súper resolución de una sola imagen basados en aprendizaje profundo

Leandro José Rodríguez Hernández¹, Humberto de Jesús Ochoa Domínguez¹,
Osslan Osiris Vergara Villegas², Juan Humberto Sossa Azuela³

¹ Universidad Autónoma de Ciudad Juárez,
Departamento de Ingeniería Eléctrica y Computación,
México

² Universidad Autónoma de Ciudad Juárez,
Departamento de Ingeniería Industrial y Manufactura,
México

³ Instituto Politécnico Nacional,
Centro de Investigación en Computación,
Laboratorio de Robótica y Mecatrónica,
México

{a1194726, hochoa}@uacj.mx, overgara@uacj.mx,
hsossa@cic.ipn.mx

Resumen. El problema de la súper resolución comprende un conjunto de métodos de procesamiento digital de imágenes; su objetivo es incrementar la resolución de las imágenes para mejorar su calidad visual. En años recientes, con el auge del aprendizaje profundo y el aumento de la capacidad computacional, han surgido una serie de métodos para súper resolución basados en aprendizaje profundo. Este trabajo presenta un estudio comparativo de cuatro métodos recientes del estado del arte de súper resolución utilizando aprendizaje profundo. Se analiza su fundamento teórico y sus arquitecturas. Se comparan y reportan sus rendimientos usando diferentes bases de datos utilizadas en súper resolución.

Palabras claves: Súper resolución, aprendizaje profundo, redes neuronales convolucionales.

Comparative Study of Single-Image Super Resolution Algorithms Based on Deep Learning

Abstract. The super resolution problem involves a set of digital image processing methods; its objective is to increase the resolution of the images to improve their visual quality. In recent years, with the rise of deep learning and increasing computational capacity, a number of methods for super resolution based on deep learning have emerged. This work presents a comparative study of four recent super resolution state-of-the-art methods using deep learning. Its theoretical foundation and architectures are analyzed. Their performances are compared and reported for different databases used in super resolution.

Keywords: Super resolution, deep learning, convolutional neural networks.

1. Introducción

La degradación de las imágenes digitales es resultado de una serie de factores que influyen en el proceso de adquisición y digitalización de las imágenes. Es deseable que las imágenes digitales contengan la mayor cantidad de información posible. En algunas aplicaciones, como por ejemplo las imágenes médicas, puede significar la detección temprana y eficaz de una enfermedad. Con el objetivo de mejorar la calidad de las imágenes digitales, se producen mejoras continuas en el hardware de los dispositivos de adquisición; sin embargo, las mejoras de hardware son altamente costosas y tienen límites físicos para su implementación. Existen áreas, donde las imágenes que se logran adquirir son de baja calidad, ejemplo: las imágenes médicas. Para abordar este problema han sido propuestos no solo mejoras en el hardware de los equipos. También mediante la utilización de técnicas del procesamiento digital de imágenes, mediante las cuales se intenta mejorar la calidad de imágenes ya adquiridas.

La súper resolución de una sola imagen (*SISR*, por sus siglas en inglés) es un problema clásico en el procesamiento de imágenes digitales. Su objetivo es incrementar la resolución de una imagen de baja resolución (*LR*, por sus siglas en inglés) para obtener una imagen de mayor resolución y recuperar información de alta frecuencia [1, 2, 3]. Existen diversas aplicaciones en diferentes áreas como video vigilancia [4, 5], sensado remoto [6] e imágenes médicas [7, 8]. En general, para *SISR* se supone que la imagen de *LR* es una versión submuestreada y ruidosa de la imagen de alta resolución (*HR*, por sus siglas en inglés). Para abordar el problema de la súper resolución, se han propuesto diferentes métodos. Los métodos basados en interpolación han sido ampliamente utilizados, debido al bajo costo y complejidad computacional [9, 10]. Tienen a difuminar los detalles de alta frecuencia, sin embargo, pueden producir escalas muy grandes.

Otro enfoque clásico de *SISR* se fundamenta en el aprendizaje o algoritmos basados en ejemplos [11-15] donde se explota la propiedad de auto similitud para generar imágenes de *HR* a partir de la entrada de *LR*. Los métodos basados en ejemplos utilizan una base de datos, también llamada diccionarios, que contiene pares de parches (porciones de la imagen) de alta y baja resolución para establecer relaciones de mapeo entre los parches de la imagen de *LR* y los parches de la imagen de *HR*; tal relación de mapeo se utiliza posteriormente para estimar nuevas imágenes de *HR* a partir de la entrada *LR*.

De acuerdo con las formulaciones de mapeo, la *SISR* fundamentada en el aprendizaje utiliza métodos que usan vecindades [16, 17], que supone que los pares de parches *LR-HR* tienen geometrías similares; basados en regresión [18, 19], que construyen un modelo de regresión para estimar las relaciones *LR-HR* y, finalmente, los métodos que utilizan la escasez [11,12], en los cuales se calcula el código escaso del parche de *LR* para estimar el parche de *HR*. Estos métodos requieren que se busque en un amplio conjunto de datos de referencia patrones similares.

Recientemente, se han propuesto algunos métodos basados en redes neuronales convolucionales (*CNN*, por sus siglas en inglés) [20-25], los cuales recuperan imágenes de *HR* aprendiendo un mapeo no lineal entre los parches de *LR* y de *HR*.

Para mejorar los resultados de la reconstrucción, las redes se entrenan utilizando conjuntos grandes de datos. En comparación con los métodos basados en interpolación y en ejemplos, los enfoques que utilizan una *CNN* proporcionan un rendimiento superior en calidad de restauración, dando como resultado mejoras en la relación señal a ruido (*PSNR*, por sus siglas en inglés) y en similitud estructural (*SSIM*, por sus siglas en inglés).

El objetivo del presente trabajo es presentar y describir la teoría y el diseño de cuatro algoritmos de súper resolución basados en aprendizaje profundo, así como comparar su rendimiento. El objetivo del estudio comparativo es proporcionar al investigador, elementos que le permitan formar una base teórica y un acercamiento a las principales arquitecturas y tendencias dentro del área, así como sus posibles aplicaciones.

Las arquitecturas seleccionadas fueron elegidas porque dentro de la gran diversidad que existen, han marcado las pautas, representaron hitos y tendencias de las cuales se derivan muchos otros trabajos, además de que hasta donde sabemos, no existe un estudio comparativo de algoritmos de *SISR* basados en aprendizaje profundo previo al que se muestra en el presente trabajo. El resto del artículo se organiza de la siguiente manera. En la sección 2 se describen los algoritmos utilizados en la comparación. En la sección 3 se describe la experimentación y en la sección 4 los resultados obtenidos son presentados. En la sección 5 se discuten los resultados y finalmente en la sección 6 se presentan las conclusiones.

2. Súper resolución de una sola imagen basada en aprendizaje profundo

El aprendizaje automático ha demostrado su eficacia como herramienta para establecer modelos y relaciones entre conjuntos de datos y se ha aplicado a diversos campos de la ciencia [26, 27]. La súper resolución es un problema inverso, también conocido como problema mal planteado, porque no están disponibles todos los elementos para establecer soluciones únicas, es decir, varias imágenes de alta resolución pueden corresponder a la misma imagen de baja resolución. Algunas técnicas de aprendizaje automático, específicamente el aprendizaje profundo, se han utilizado como herramientas para establecer relaciones óptimas entre una imagen de *LR* y su correspondiente imagen de *HR* [20, 22, 23].

La Fig. 1 muestra el esquema general para súper resolución. La imagen x es la imagen de baja resolución de entrada a la red. Los filtros de convolución W_i son los encargados de extraer los mapas de características $F_i(x)$. Opcionalmente, algunos autores incorporan un término de “*bias*” o sesgo, que se denota como B_i . La salida de la red es la imagen \hat{y} , que es la imagen de súper resolución. La imagen y de *HR*, es la imagen ideal o “*ground truth*”, utilizada para el entrenamiento.

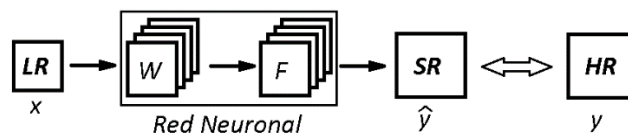


Fig. 1. Esquema general de súper resolución basada en aprendizaje profundo.

2.1. Arquitecturas poco profundas

El término "arquitecturas poco profundas" se utiliza para referirse a redes neuronales con pocas capas. Aunque no existe un límite rígido, una red con menos de diez capas se considera poco profunda. Este tipo de arquitecturas tiene la ventaja de ser más fácil de entrenar, produce resultados superiores a los métodos tradicionales de *SISR* mencionados anteriormente, manteniendo un nivel medio de complejidad computacional. En los trabajos de la Red Neuronal Convolutiva para Súper Resolución (*SRCNN* por sus siglas en inglés) [20] y de la Red Neuronal Convolutiva Rápida para Súper Resolución (*FSRCNN* por sus siglas en inglés) [22], se utilizaron arquitecturas de redes neuronales poco profundas.

Red Neuronal Convolutiva para Súper Resolución (*SRCNN*): La red neuronal *SRCNN* [20] fue el primer trabajo que utilizó una *CNN* para *SISR*, demostrando un rendimiento superior a los métodos anteriores de la literatura, tanto en velocidad como en calidad de restauración. Los autores presentan un método que aprende la relación de mapeo entre las imágenes *LR* y *HR*. El mapeo está representado por una *CNN* que toma la imagen *LR* como entrada y genera la imagen *HR*. La arquitectura se muestra en la Fig. 2. Es una *CNN* de tres capas, donde los tamaños de los filtros son $64 \times 1 \times 9 \times 9$, $32 \times 64 \times 5 \times 5$ y $1 \times 32 \times 5 \times 5$ respectivamente. Las funciones de estas tres capas de transformaciones no lineales son: extracción de características, mapeo no lineal y reconstrucción de la imagen súper resuelta.

1. *Capa de extracción de características.* Antes de ingresar a la red *SRCNN*, la entrada de baja resolución se escala al tamaño deseado mediante interpolación bicúbica. Posteriormente, la primera capa realiza una convolución estándar de la imagen escalada con los coeficientes de los filtros entrenados, seguida de una unidad ReLU para obtener el mapa de características $F_1(x)$:

$$F_1(x) = \max(0, W_1 * x + B_1). \quad (1)$$

2. *Capa de mapeo no lineal.* Es un mapeo del vector n_1 -dimensional al vector n_2 -dimensional. El número de características se reduce de 64 a 32. Esta operación de mapeo no lineal se realiza mediante la ecuación:

$$F_2(x) = \max(0, W_2 * F_1(x) + B_2). \quad (2)$$

3. *Capa de reconstrucción:* Después del mapeo, se obtiene la imagen de súper resolución mediante una capa de convolución final:

$$F(x) = \max(0, W_3 * F_2(x) + B_3). \quad (3)$$

En esta red se utilizó el error cuadrático medio (*MSE* por sus siglas en inglés) como función de pérdida y el algoritmo gradiente descendiente estocástico (*SGD* por sus siglas en inglés) para resolver el problema de optimización:

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n \|F(x_i) - y_i\|^2. \quad (4)$$

La Fig. 2 muestra la arquitectura general de la red *SRCNN*.

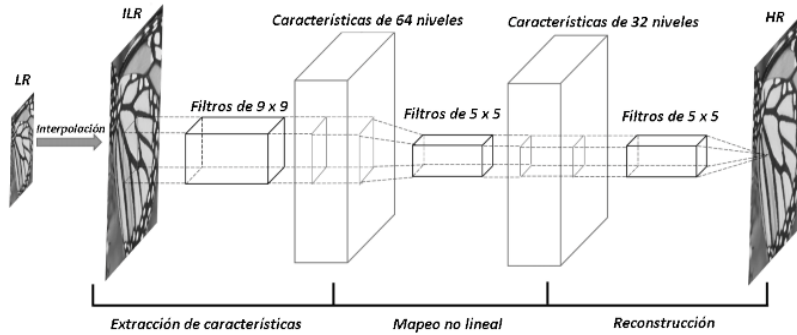


Fig. 2. Red SRCNN [20].

Red Neuronal Convolutiva Rápida para Súper Resolución (FSRCNN): Dong et al. [19] propusieron la red *FSRCNN* para acelerar la red *SRCNN* [17] a través de una *CNN* de estructura más compacta. El procedimiento de *SISR* se logró más rápidamente y obtuvieron mejores resultados. Rediseñaron la estructura *SRCNN* en tres aspectos principales: introdujeron una capa de deconvolución al final de la red, las capas de mapeo se modificaron reduciendo las dimensiones de las características de entrada antes del mapeo y se usaron filtros de menor tamaño, pero se usaron más capas de mapeo. El modelo propuesto alcanzó una velocidad 40 veces mayor que la red *SRCNN*, con una calidad de restauración de la imagen superior a la de los métodos clásicos. La red *FSRCNN* se organiza en cinco fases principales que son:

1. Extracción de características: mediante la utilización de filtros de 5×5 para obtener d mapas de características.
2. Reducción de las dimensiones: se utiliza una convolución 1×1 para reducir la cantidad de mapas de características de d a s .
3. Mapeo no lineal: mediante la utilización de filtros de 3×3 se realiza un mapeo entre las características baja resolución y alta resolución.
4. Expansión: la convolución 1×1 se realiza para aumentar la cantidad de mapas de características de s a d .
5. Deconvolución: se utilizan filtros de 9×9 para reconstruir la imagen de *HR*.

El error cuadrático medio (*MSE*) se usó como la función de pérdida en este método:

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n \|F(x_i) - y_i\|^2. \quad (5)$$

La Fig. 3 muestra la estructura general de la red *FSRCNN*.

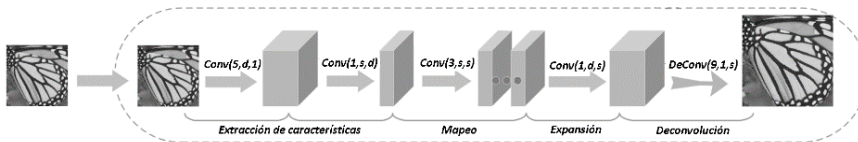


Fig. 3. Red FSRCNN [22].

2.2. Arquitecturas profundas

Una red más profunda tiene una arquitectura más compleja (Tabla 4) y puede lograr mejores resultados, aunque el costo computacional en la etapa de entrenamiento y ejecución aumenta considerablemente. Se considera una red muy profunda si tiene más de diez capas. Los autores de la Red muy profunda para Súper Resolución (“Very Deep Super Resolution” *VDSR* por sus siglas en inglés) [23] y de la Red profunda y recursiva para Súper Resolución (“Deep Recursive Convolutional Network” *DRCN* por sus siglas en inglés) [24] utilizaron arquitecturas de redes neuronales muy profundas para aplicarlas a sus trabajos en *SISR*.

Las dos redes, basan su entrenamiento en el aprendizaje residual, el cual permite pasar información de la imagen original magnificada (a la escala deseada mediante un filtro bicúbico), a las capas finales, lo que contribuye a reducir los efectos adversos del desvanecimiento del gradiente [28]. Generalmente, la información se pasa de la primera capa a la última, donde finalmente se suman la información procesada por la red y la información inicial. Para el caso de *SISR*, la imagen escalada a la entrada se asume que es la información de baja frecuencia, ya que carece de detalles y bordes bien definidos. La red neuronal recupera la información de alta frecuencia y, finalmente, suma las dos informaciones para obtener la imagen de súper resolución.

Red muy Profunda para Súper Resolución (*VDSR*): En [23] se propone la red *VDSR* para *SISR*. El método utiliza una red convolucional muy profunda inspirada en la red *VGG* [29], donde el aumento en la profundidad de la red contribuye a una mejora significativa en la precisión. El modelo usa 20 capas con pequeños filtros en cascada para lograr utilizar la información contextual sobre grandes regiones de la imagen de manera más eficiente. Para resolver el problema de la velocidad de convergencia durante el entrenamiento, se propuso utilizar una arquitectura residual, pasando la información de entrada hasta la última capa, para añadirla con la información procesada por la red y tasas de entrenamiento muy altas (104 veces más altas que la *SRCNN*). Los gradientes se acotan entre $[-\theta/\gamma; \theta/\gamma]$, donde γ denota la tasa de aprendizaje actual; θ se ajusta a un valor pequeño para evitar el problema de la explosión del gradiente durante la etapa de entrenamiento. La arquitectura de la red *VDSR* se organiza de la siguiente manera:

1. La imagen de *LR* se interpola y se obtiene la imagen *ILR* que se ingresa a la red.
2. La imagen *ILR* pasa a través de las $(D-1)$ capas Conv y ReLU.
3. Continúa la D -ésima capa residual (Conv.D (Residual)).
4. Finalmente, la salida se suma con la imagen *ILR* y se obtiene la imagen *HR*.

En la red residual, la imagen de entrada de baja resolución se agrega a la salida de la red para obtener la imagen *HR*. Por lo tanto, la función de pérdida es:

$$\frac{1}{2} \|r - F(x)\|^2, \quad (6)$$

donde $F(x)$ es la salida de la red y $r = Y - X$. Se aprenden los errores residuales entre la salida y la entrada en lugar de aprender la salida *HR* directamente.

La Fig. 4 muestra la arquitectura general de la red *VDSR*.

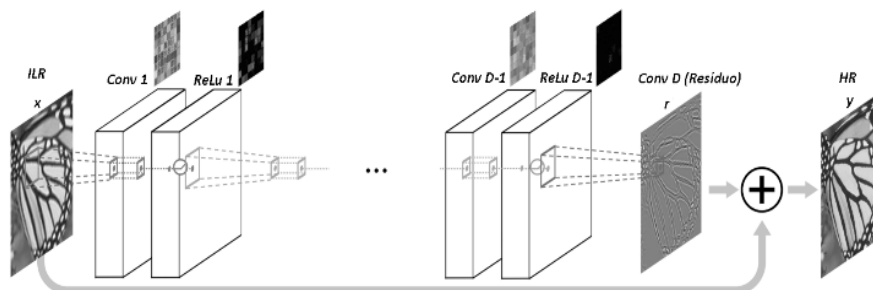


Fig. 4. Red VDSR [23].

Red Profunda y Recursiva para Súper Resolución (DRCN): La Red Profunda y Recursiva para Súper Resolución (DRCN por sus siglas en inglés) [24] consta de tres etapas: de extracción de características, de inferencia y de reconstrucción. Utiliza una arquitectura residual:

1. *Etapas de extracción de características f_1 .* Se toma la imagen de baja resolución interpolada x y se representa como un conjunto de mapas de características F_0 utilizando dos capas convolucionales:

$$F_{-1} = \max(0, W_{-1} * x + B_{-1}), \quad (7)$$

$$F_0 = \max(0, W_0 * F_{-1} + B_0), \quad (8)$$

$$f_1(x) = F_0. \quad (9)$$

2. *Etapas de inferencia f_2 .* Los mapas de características de salida de la red pasan por una sola capa recursiva. Cada recursión aplica la misma convolución seguida de una capa ReLU:

$$F_d = g(F_{d-1}) = \max(0, W * F_{d-1} + B), \quad (10)$$

$$f_2(F) = (g \circ g \circ \dots \circ g)(F) = g^D(F). \quad (11)$$

3. *Etapas de reconstrucción f_3 .* Transforma los mapas de funciones de salida de la red de inferencia en la imagen de alta resolución:

$$F_{D+1} = \max(0, W_{D+1} * F_D + B_{D+1}), \quad (12)$$

$$\hat{y} = \max(0, W_{D+2} * F_{D+1} + B_{D+2}), \quad (13)$$

$$f_3(F) = \hat{y}. \quad (14)$$

La función de pérdida es una combinación de L1 y L2, además de un término de regularización.

Pérdida L1: *MSE* entre las salidas D y la imagen de HR usando supervisión recursiva:

$$L_1(\theta) = \sum_{d=1}^D \sum_{i=1}^N \frac{1}{2DN} \|y^{(i)} - \hat{y}_d^{(i)}\|^2. \quad (15)$$

Pérdida L2: *MSE* entre las salidas D ponderadas y la imagen HR.

$$L_2(\theta) = \sum_{i=1}^N \frac{1}{2N} \left\| y^{(i)} - \sum_{d=1}^D w_d * \hat{y}_d^{(i)} \right\|^2. \quad (16)$$

Resultando finalmente:

$$L(\theta) = \alpha L_1(\theta) + (1 - \alpha)L_2(\theta) + \beta \|\theta\|^2, \quad (17)$$

donde θ son los parámetros de la red.

La Fig. 5 muestra la arquitectura general de la red DRCN.

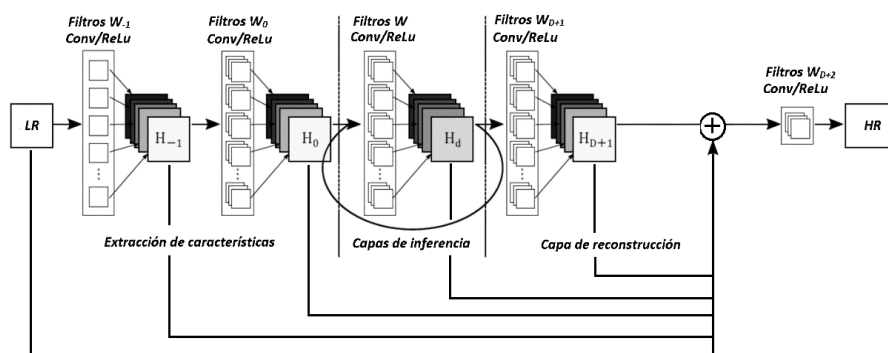


Fig. 5. Red DRCN [23].

3. Estudio comparativo

Los algoritmos utilizados en esta comparación están implementados en Matlab [30] y Pytorch [31], los parámetros de cada método se establecieron de acuerdo con los valores dados en sus reportes correspondientes. Las pruebas se realizaron en una computadora con sistema operativo Windows 10, procesador Intel (R) Core (TM) i7-6700 3.40GHz, con 16 GB de RAM. Para las pruebas, se utilizaron imágenes naturales de bases de datos "Set5" [32], "Set14" [33], "Urban100" [34], y "BSD100"[35, 36] las cuales son ampliamente utilizadas para realizar las pruebas en trabajos de súper resolución [20, 22, 23].

3.1. Métricas

Para comparar el rendimiento de los algoritmos *SISR*, se utilizaron las métricas Relación Señal Ruido ("Peak-Signal to Noise Ratio" *PSNR*) e Índice de Similitud Estructural ("Structural Similarity Index" *SSIM*), que se calculan entre la imagen súper resuelta y su original. El *PSNR* se calcula a partir del error medio cuadrático (*MSE*),

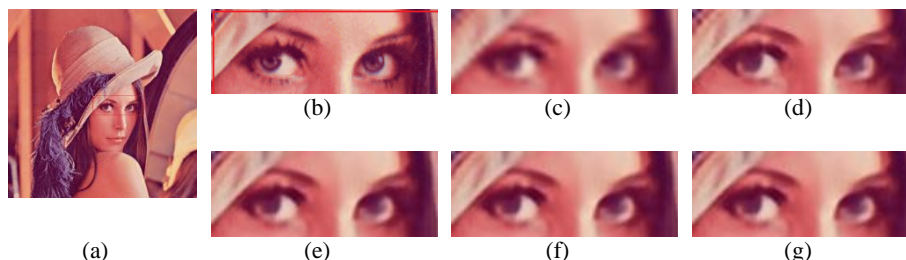


Fig. 6. Comparación visual para una escala de 4x en la imagen de Lena, (a) imagen de alta resolución *HR*, (b) porción de la imagen de *HR*, (c) interpolación bicúbica, (d) método *SRCNN*, (e) método *FSRCNN*, (f) método *VDSR* y (g) método *DRCN*.

que es el error promedio entre la imagen original y la imagen súper resuelta. Dada una imagen súper resuelta $\hat{y}(i, j)$ de dimensiones $m \times n$ y su original $y(i, j)$, *MSE* y *PSNR* se definen como:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [\hat{y}(i, j) - y(i, j)]^2, \quad (18)$$

$$PSNR = 20 \log_{10} \left(\frac{255}{\sqrt{MSE}} \right). \quad (19)$$

El índice *SSIM* [37] es un modelo basado en la percepción y calcula la similitud entre las imágenes originales y súper resueltas, entre más cercano a 1 es este índice, mayor es el parecido entre las imágenes comparadas. El *SSIM* tiene en cuenta la luminancia, el contraste y los cambios estructurales entre las dos imágenes. El índice *SSIM* se define como:

$$SSIM = \frac{(2\mu_y\mu_{\hat{y}} + C_1)(2\sigma_{y\hat{y}} + C_2)}{(\mu_y^2 + \mu_{\hat{y}}^2 + C_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + C_2)}, \quad (20)$$

donde μ_y , $\mu_{\hat{y}}$ son las medias y σ_y , $\sigma_{\hat{y}}$ son las desviaciones estándar de las imágenes original y súper resuelta, $\sigma_{y\hat{y}}$ es la covarianza entre y e \hat{y} . C_1 , C_2 son constantes.

3.2. Experimentación

En la experimentación, se utilizó un total de 219 imágenes de las diferentes bases de datos, con el fin de medir el desempeño promedio de *PSNR* y el *SSIM* de cada uno de los algoritmos. Las pruebas se realizaron en las escalas de aumento de 2x, 3x y 4x, reportando para cada escala los resultados promedios del *PSNR* y el *SSIM* en cada una de las bases de datos de imágenes utilizadas.

El algoritmo de interpolación bicúbica se consideró para la comparación, ya que se usa ampliamente como punto de referencia para la comparación en trabajos recientes del estado del arte.

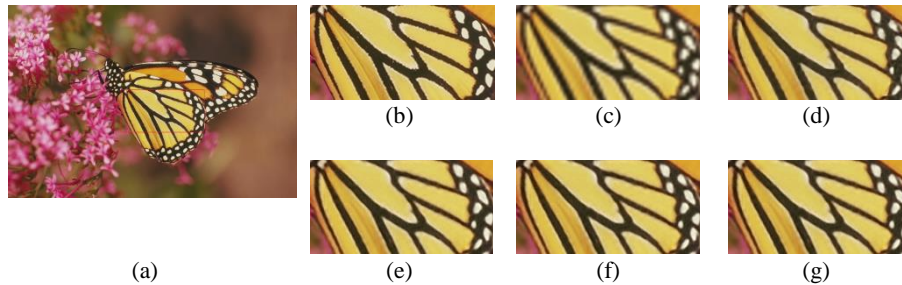


Fig. 7. Comparación visual para una escala de 4x de la imagen Monarch, (a) imagen de alta resolución *HR*, (b) porción de imagen de *HR*, (c) interpolación bicúbica, (d) método *SRCNN*, (e) método *FSRCNN*, (f) método *VDSR* y (g) método *DRCN*.

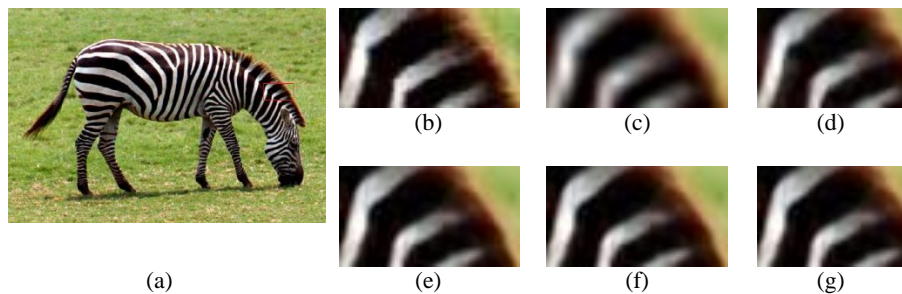


Fig. 8. Comparación visual para una escala de 4x de la imagen Zebra, (a) imagen de alta resolución *HR*, (b) porción de imagen de alta resolución *HR*, (c) interpolación bicúbica, (d) método *SRCNN*, (e) método *FSRCNN*, (f) método *VDSR* y (g) método *DRCN*.

Tabla 1. Promedios de *PSNR* y *SSIM* para un factor de escala de 2x.

Método	Tamaño	Bases de datos							
		Set5		Set14		BSD100		Urban100	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicúbico		33.65	0.965	30.20	0.930	29.57	0.928	26.89	0.856
SRCNN		36.52	0.976	32.48	0.933	31.34	0.938	29.56	0.895
FSRCNN	2x	36.98	0.977	32.64	0.932	31.50	0.937	29.83	0.909
VDSR		37.54	0.981	33.05	0.938	31.88	0.945	30.58	0.911
DRCN		37.62	0.981	33.08	0.937	31.84	0.943	30.64	0.915

Tabla 2. Promedios de *PSNR* y *SSIM* para un factor de escala de 3x.

Método	Tamaño	Bases de datos							
		Set5		Set14		BSD100		Urban100	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicúbico		30.38	0.936	27.57	0.895	27.25	0.853	24.38	0.789
SRCNN		32.59	0.955	29.30	0.911	28.36	0.884	26.29	0.802
FSRCNN	3x	33.10	0.959	29.39	0.911	28.53	0.940	26.41	0.810
VDSR		33.62	0.962	29.68	0.931	28.85	0.940	27.18	0.846
DRCN		33.80	0.962	29.74	0.932	28.85	0.942	27.20	0.851

Tabla 3. Promedios de *PSNR* y *SSIM* para un factor de escala de 4x.

Método	Tamaño	Bases de datos							
		Set5		Set14		BSD100		Urban100	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicúbico		28.41	0.908	26.02	0.880	25.97	0.834	23.06	0.722
SRCNN	4x	30.38	0.934	27.52	0.896	26.85	0.851	24.48	0.789
FSRCNN		30.70	0.936	27.62	0.896	26.91	0.850	24.60	0.834
VDSR		31.31	0.939	28.08	0.901	27.29	0.862	25.14	0.835
DRCN		31.54	0.939	28.05	0.911	27.35	0.868	25.25	0.847

Tabla 4. Número de capas y parámetros en cada uno de los métodos.

Método	Número de capas	Número de parámetros
SRCNN	3	57 mil
FSRCNN	8	12 mil
VDSR	20	665 mil
DRCN	5 (recursivas)	1.77 millones

4. Resultados

Los resultados obtenidos en la experimentación se presentan a continuación. Las figuras 6, 7 y 8 muestran los resultados visuales de la comparación en las imágenes de Lena, Monarch y Zebra de la base de datos BSD100, para un factor de escala de 4x. Para una mejor apreciación visual, se muestran sólo porciones de las imágenes mencionadas.

Las tablas 1, 2 y 3 muestran los promedios de *PSNR* y *SSIM* utilizando las bases de datos Set5, Set14, BSD100 y Urban100 para diferentes factores de escala 2x, 3x y 4x.

5. Discusión

En las tablas 1-3 se observa que cualitativamente con el *SSIM* y cuantitativamente con el *PSNR*, el método de interpolación bicúbica proporciona los resultados más modestos. Los métodos muy profundos obtienen mejores resultados en comparación con las arquitecturas poco profundas. Los métodos muy profundos logran, en promedio, mejorar el *PSNR* al compararlos contra la interpolación bicúbica en 2.39 dB, mientras que los métodos poco profundos aventajan a la interpolación bicúbica en 2.17dB.

Para una escala de magnificación de 4x, los métodos profundos aventajan por 2.13 dB a la interpolación bicúbica, mientras que los poco profundos elevan en promedio al *PSNR* en 1.51 dB. Para la escala de 3x, las mejoras son de 2.47 dB para los muy profundos, contra 1.85 dB de los poco profundos y para la escala 2x, los muy profundos logran mejoras de 3.20 dB, contra 2.52 dB de los poco profundos. Podemos ver que a medida en que la escala aumenta, la calidad de las imágenes recuperadas se degrada. La recursividad en la estructura *DRCN* permite que la red *VDSR* sea más compacta,

reduciendo el número de capas y conexiones. Aunque el uso de la recursividad requiere una gran cantidad de memoria RAM, lo que hace que el método sea potencialmente inestable.

Los métodos *VDSR* y *DRCN* tienen rendimientos similares, superiores a los métodos poco profundo, la arquitectura *DRCN* es más compacta que *VDSR*. Los menos profundos tienen una menor cantidad de capas como se aprecia en la tabla 4, esto implica una menor complejidad computacional, ya que el número de operaciones y la cantidad de valores a almacenar será menor.

Por lo tanto, son buenos candidatos cuando se requieren métodos computacionalmente livianos, sacrificando calidad de imagen. Estos métodos son los candidatos ideales para aplicaciones de tiempo real, donde se necesitan obtener los resultados en un período relativamente corto de tiempo. Las redes muy profundas son aplicables donde la complejidad no es un factor determinante, por ejemplo, en la súper resolución de imágenes médicas donde la calidad de la imagen, puede ayudar a un mejor diagnóstico.

Finalmente, se puede observar que los valores más altos de *PSNR* no siempre corresponden a una mejor percepción visual de la calidad de imagen, esto indica que esta métrica, a pesar de ser ampliamente utilizada en la literatura, no es suficiente para establecer la calidad absoluta del método de súper resolución. Debido a esto, es conveniente también utilizar la métrica *SSIM*.

6. Conclusiones

Se presentó un estudio comparativo de cuatro algoritmos de súper resolución basados en aprendizaje profundo. Se realizaron mediciones de *PSNR* y *SSIM* de las imágenes en las bases de datos, comúnmente utilizadas en súper resolución, para escalas de magnificación de 2x, 3x y 4x, con el fin de obtener los promedios en cada base de datos. Se muestran comparaciones cuantitativas y cualitativas de cada método. Los resultados demuestran que las arquitecturas más profundas probadas obtienen mejores resultados de *PSNR* y *SSIM*. En general, en los métodos muy profundos *VDSR* y *DRCN* siempre se obtuvieron los mejores resultados.

El uso del aprendizaje residual en las redes muy profundas *VDSR* y *DRCN* permite que el proceso de entrenamiento sea menos complejo, debido a que la cantidad de información que se procesa en la red se reduce y se logra minimizar los efectos del desvanecimiento del gradiente, el cual es un problema que surge en la medida que las redes se hacen más profundas. Los métodos poco profundos tienen más utilidad en aplicaciones que requieran menor complejidad. Si la calidad de las imágenes es lo que interesa, las redes muy profundas pueden valorarse como una opción.

En trabajos posteriores se explorará la influencia de la elección del diseño de la red neuronal, así como la función de pérdida en los rendimientos de estos algoritmos.

Agradecimientos. L. J. Rodríguez agradece a la UACJ por el apoyo brindado y al CONACYT por la beca otorgada para realizar sus estudios doctorales. H. Sossa agradece al Instituto Politécnico Nacional y a la SIP-IPN por el apoyo otorgado a través del proyecto SIP-IPN 20200630 para llevar a cabo esta investigación.

Referencias

1. Singh, A., Singh, J.: Survey on single image based super-resolution implementation challenges and solutions. *Multimedia Tools and Applications*, 79(3), pp. 1641–1672 (2019)
2. Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J., Liao, Q.: Deep learning for single image super-resolution: a brief review. *IEEE Trans. Multimedia*, 21(12), pp. 3106–3121 (2019)
3. Li, X., Wu, Y., Zhang, W., Wang, R., Hou, F.: Deep learning methods in real-time image super-resolution: A survey. *J. of Real-Time Image Processing*, pp. 1–22 (2019)
4. Li, D., Wang, Z.: Video superresolution via motion compensation and deep residual learning. *IEEE Trans. on Computational Imaging*, 3(4), pp. 749–762 (2017)
5. Huang, Y., Wang, W., Wang L.: Video super-resolution via bidirectional recurrent convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 40(4), pp. 1015–1028 (2018)
6. Lei, S., Shi, Z., Zou Z.: Super-resolution for remote sensing images via local–global combined network. *IEEE Geosci. Remote Sens. Lett.*, 14(8), pp. 1243–1247 (2017)
7. Shi, J., Li, Z., Ying, S., Wang, C., Liu, Q., Zhang, Q., Yan, P.: MR image super-resolution via wide residual networks with fixed skip connection. *IEEE J. Biomed. Health Inform.*, 23(3), pp. 1129–1140 (2019)
8. Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S. A., de Marvao, A., Dawes, T., O’Regan, D. P.: Anatomically constrained neural networks (ACNNS): Application to cardiac image enhancement and segmentation. *IEEE Trans. Med. Imag.*, 37(2), pp. 384–395 (2018)
9. Zhou, F., Yang, W., Liao, Q.: Interpolation-based image super-resolution using multisurface fitting. *IEEE Trans. Image Process.*, 21(7), pp. 3312–3318 (2012)
10. Mahmoudzadeh, Pasha, A., Kashou, N.: Interpolation-based super-resolution reconstruction: effects of slice thickness. *J. of Medical Imaging* 1(3), pp. 1–9 (2014)
11. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution as sparse representation of raw image patches. In: *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, AK, pp. 1–8 (2008)
12. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution via sparse representation. *IEEE Trans. Image Process.*, 19(11), pp. 2861–2873 (2010)
13. Timofte, R., De-Smet, V., Van-Gool, L.: A+: Adjusted anchored neighborhood regression for fast super-resolution. In: Cremers D., Reid I., Saito H., Yang MH. (eds) *Computer Vision – ACCV. Lecture Notes in Computer Science*, 9006, pp. 111–126 (2014)
14. Yang, C. Y., Yang, M. H.: Fast direct super-resolution by simple functions. In: *IEEE International Conf. on Computer Vision*, Sydney, NSW, pp. 561–568 (2013)
15. Schuler, S., Leistner, C., Bischof, H.: Fast and accurate image upscaling with super-resolution forests. In: *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, pp. 3791–3799 (2015)
16. Zhang, K., Gao, X., Tao, D., Li, X.: Single image super resolution with nonlocal means and steering kernel regression. *IEEE Trans. Image Process.*, 21(11), pp. 4544–4556 (2012)
17. Chang, H., Yeung, D. Y., Xiong, Y.: Super-resolution through neighbor embedding. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Washington, DC, USA, pp. 275–282 (2004)
18. Zhang, K., Wang, B., Zuo, W., Zhang, H., Zhang, L.: Joint learning of multiple regressors for single image superresolution. *IEEE Signal Process. Lett.*, 23(1), pp. 102–106 (2016)
19. Zhang, K., Tao, D., Gao, X., Li, X., Xiong, Z.: Learning multiple linear mappings for efficient single image super resolution. *IEEE Trans. Image Process.*, 24(3), pp. 846–861 (2015)
20. Dong, C., Loy, C. C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(2), pp. 295–307 (2015)

21. Wang, Z., Liu, D., Yang, J., Han, W., Huang, T.: Deep networks for image super-resolution with sparse prior. In: International Conf. on Computer Vision, pp. 370–378 (2015)
22. Dong, C., Loy, C. C., Tang, X.: Accelerating the superresolution convolutional neural network. In: European Conf. on Computer Vision (ECCV), pp. 1–22 (2016)
23. Kim, J., Lee, J. K., Lee, K. M.: Accurate image superresolution using very deep convolutional networks. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) Las Vegas, NV, pp. 1646–1654 (2016)
24. Kim, J., Lee, J. K., Lee, K. M.: Deeply-recursive convolutional network for image super-resolution. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1–8 (2016)
25. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep Laplacian pyramid networks for fast and accurate super-resolution. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 5835–5843 (2017)
26. Bishop, C. M.: Pattern recognition and machine learning. Springer, New York (2006)
27. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. MIT press (2016)
28. Bengio, Y., Simard, P., Frasconi, P.: Learning long-term dependencies with gradient descent is difficult. IEEE Trans. on Neural Networks 5(2), pp. 157–166 (1994)
29. Robots: Very deep convolutional networks for large-scale visual recognition. https://www.robots.ox.ac.uk/~vgg/research/very_deep/ (2020)
30. MATLAB: <https://www.mathworks.com/products/matlab.html>, last accessed 2020/03/20.
31. Pytorch: <https://pytorch.org/> (2020)
32. Bevilacqua, M., Roumy, A.; Guillemot, C., M., Alberi-Morel, L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: British Machine Vision Conference BMVC, pp. 1–10 (2012)
33. Zeyde, R., Elad, M., Protter, M.: On single image scaleup using sparse-representations. In: Boissonnat JD. et al. (eds) Curves and Surfaces. Curves and Surfaces 2010, Lecture Notes in Computer Science, 6920, pp. 710–730 (2010)
34. Huang, J. B., Singh, A., Ahuja, N.: Single image superresolution from transformed self-exemplars. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 5197–5206 (2015)
35. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: International Conf. Computer Vision, 2, pp. 416–423 (2001)
36. The Berkeley Segmentation Dataset and Benchmark. <https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/> (2020)
37. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions On Image Processing, 3(4), pp. 600–612 (2004)