

Super-Resolución en Imágenes

Tesis de Maestría - PEDECIBA Informática

Mercedes Marzoa Tanco ¹

Tutores:

Andrés Almansa ^{1,2}

Matias Di Martino ^{1,3}

Gonzalo Tejera¹

Agosto 2019

[1] Facultad de Ingeniería - Universidad de la República

[2] Université Paris Descartes

[3] Duke University

Agenda

1. Introducción
2. Marco Teórico
3. Trabajos Relacionados
4. Solución Propuesta y Resultados
5. Conclusiones

Introducción

- Proyecto CNES - MAP5
- Estudiar diferentes algoritmos de super-resolución, en particular para su utilización en imágenes satelitales



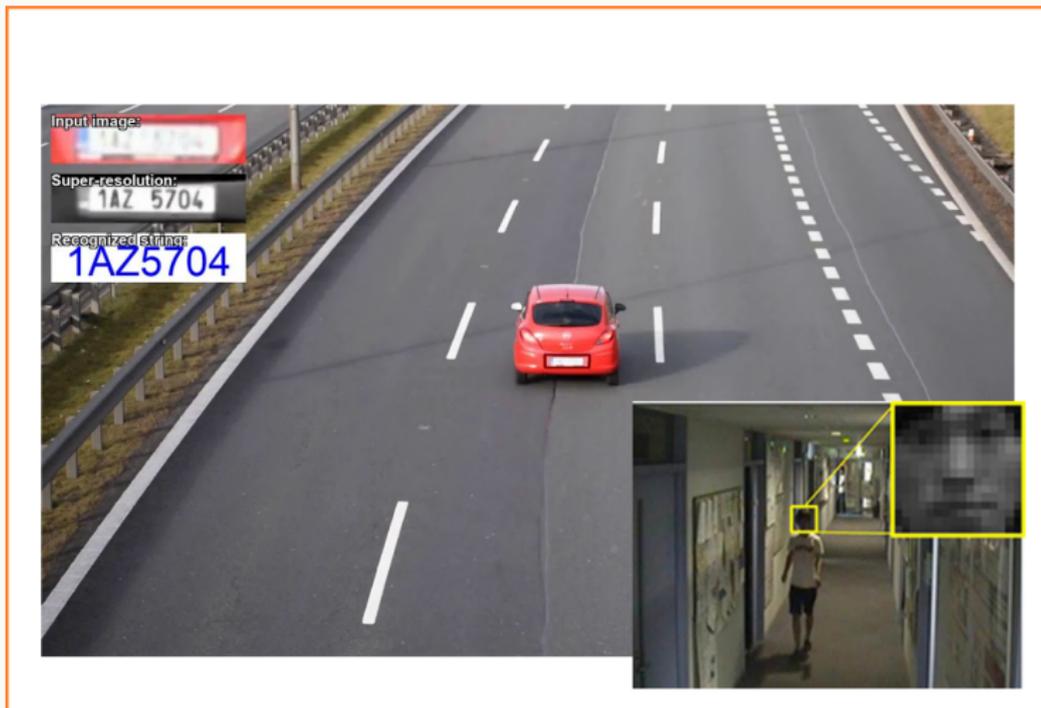
Introducción

- Super-resolución: proceso de mejorar la resolución espacial de una imagen estimando la alta resolución a partir de una o varias imágenes degradadas de baja resolución.



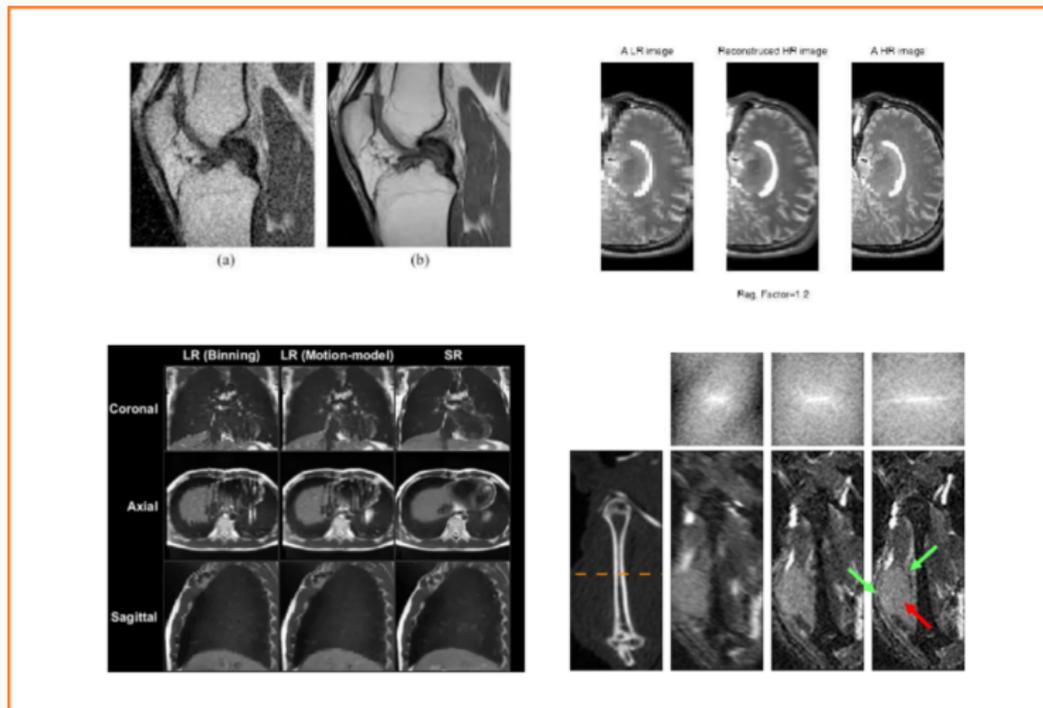
Introducción - Motivación

- Video-vigilancia



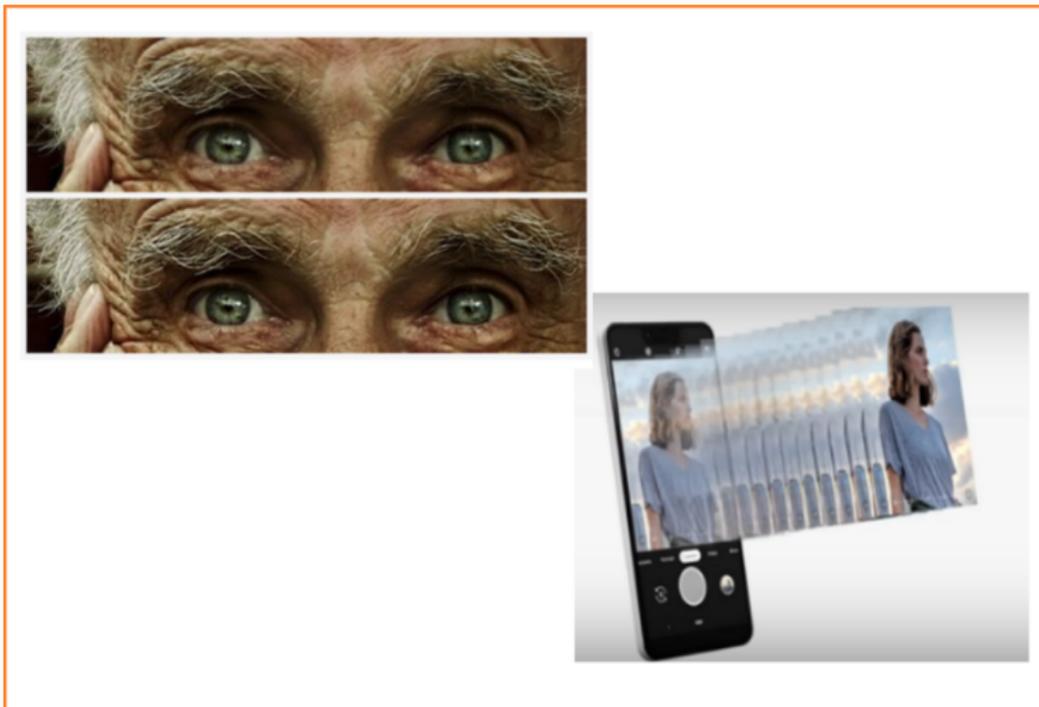
Introducción - Motivación

- Aplicaciones medicas



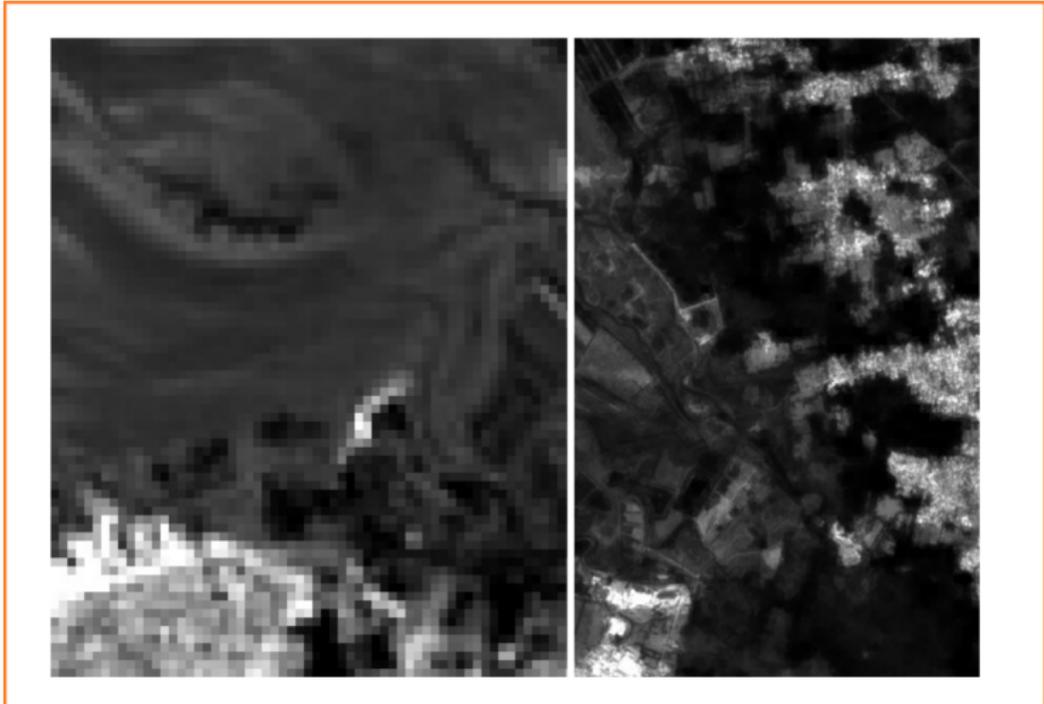
Introducción - Motivación

- Dispositivos móviles



Introducción - Motivación

- Imágenes Satelitales



En este proyecto:

- Estudiar la factibilidad de utilizar redes neuronales para mejorar la resolución de las imágenes satelitales

En este proyecto:

- Estudiar la factibilidad de utilizar redes neuronales para mejorar la resolución de las imágenes satelitales
- Responder las siguientes preguntas:
 - ¿Son mejores las redes neuronales de SR multi-imagen que los algoritmos clásicos de SR?
 - Dada una serie de imágenes de baja resolución con un desplazamiento desconocido entre ellas, ¿tiene sentido hacer SR mono-imagen, registrarlas y luego utilizar un algoritmo variacional para realizar la SR?
 - ¿Cómo impacta la elección de los métodos de interpolación utilizados en la generación de la base de datos de entrenamiento?

Marco Teórico

Imágen Digital

Una imagen digital es una representación discreta de un campo de luz continuo:

$$u(i, j) = S_{i,j}(k * u(x, y)) + n, (x, y) \in \Omega \subset \mathbb{R}^2. \quad (1)$$



Escena

$$u(x, y)$$



Desenfoque óptico

$$k * u(x, y)$$



Muestreo

$$S_{i,j}(k * u(x, y))$$



Ruido

$$S_{i,j}(k * u(x, y)) + n$$

Imágen Digital

Una imagen digital es una representación discreta de un campo de luz continuo:

$$u(i, j) = S_{i,j}(k * u(x, y)) + n, (x, y) \in \Omega \subset \mathbb{R}^2.$$



Escena

$u(x, y)$



Desenfoque óptico

$k * u(x, y)$



Muestreo

$S_{i,j}(k * u(x, y))$

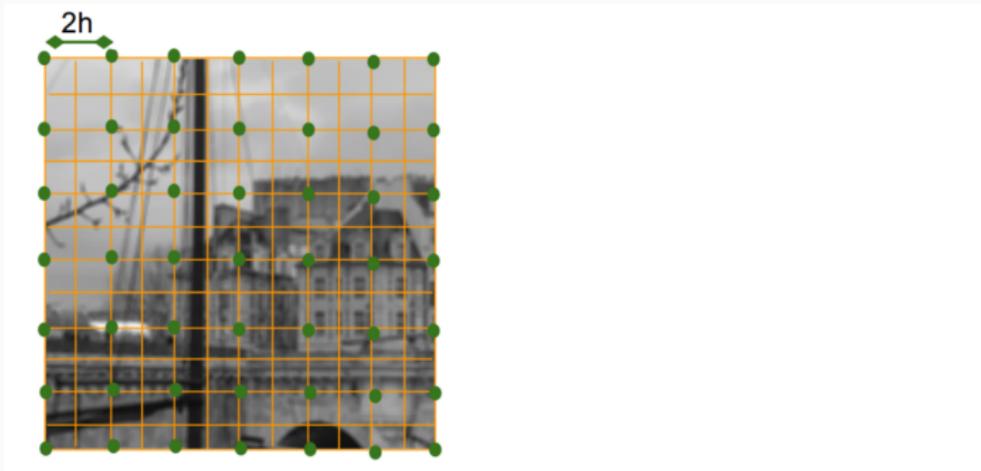


Ruido

$S_{i,j}(k * u(x, y)) + n$

Súper-Resolución

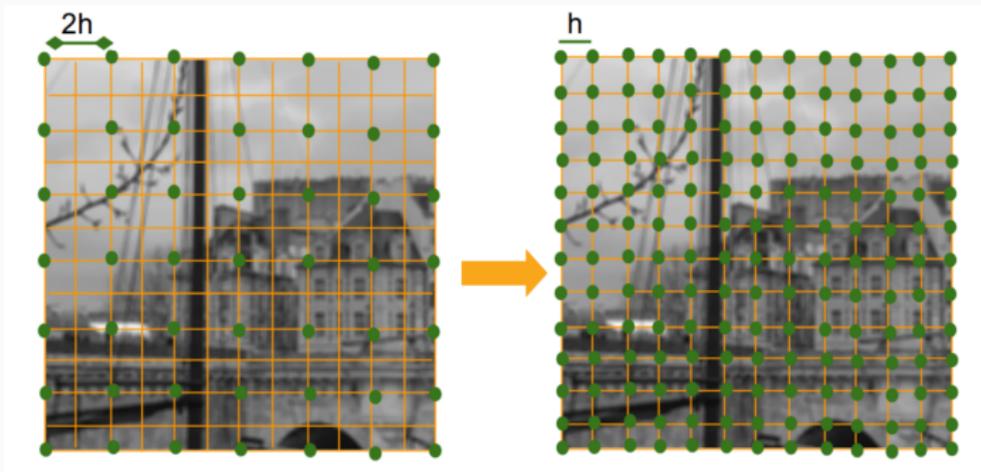
Estimación a partir de una muestra de baja resolución u^{rh} de su
contraparte de alta resolución u^h , siendo r el coeficiente de resolución.



$$r=2$$

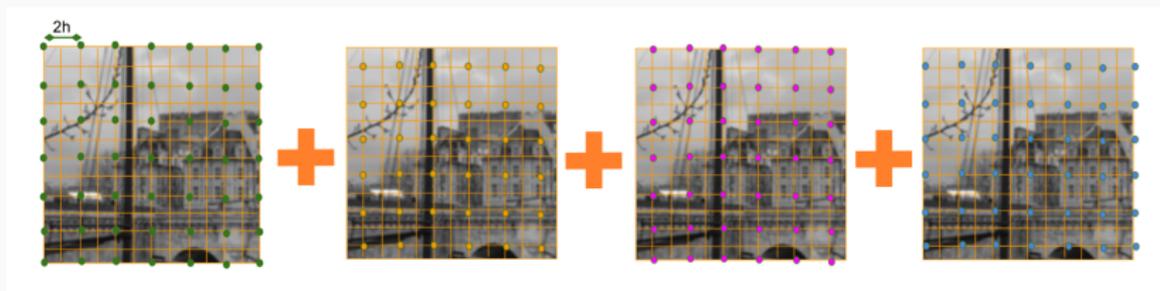
Súper-Resolución

Estimación a partir de una muestra de baja resolución u^{rh} de su
contraparte de alta resolución u^h , siendo r el coeficiente de resolución.



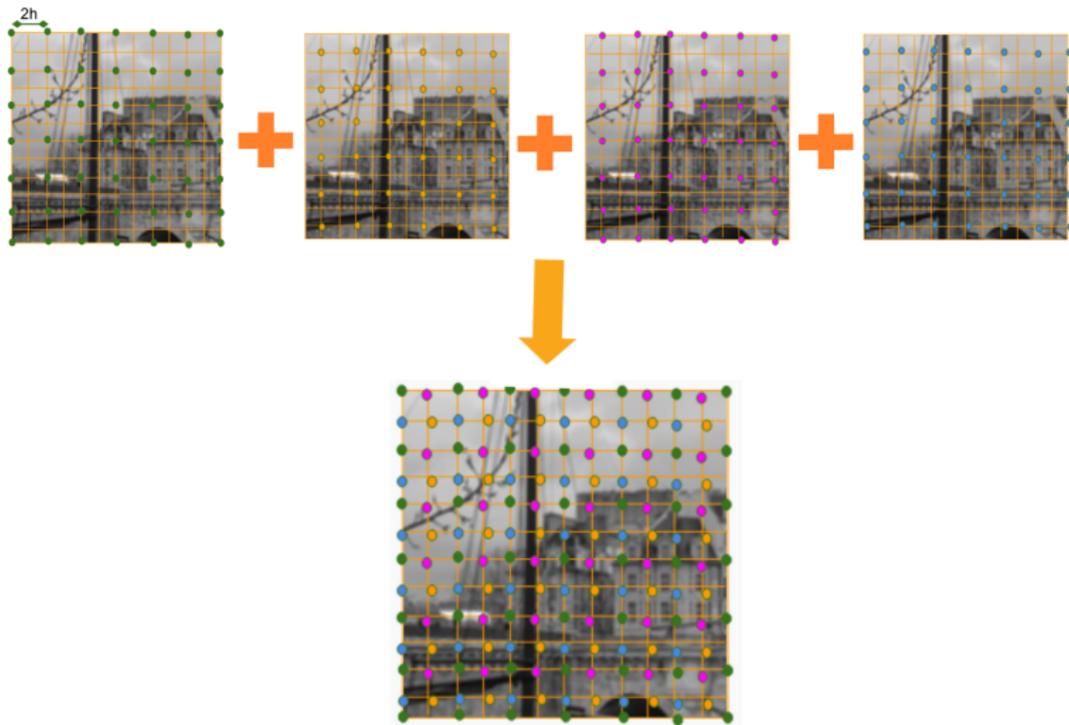
$$r=2$$

Súper-Resolución multi-imagen



Imágenes de baja resolución desplazadas

Súper-Resolución multi-imagen



Súper-Resolución multi-imagen

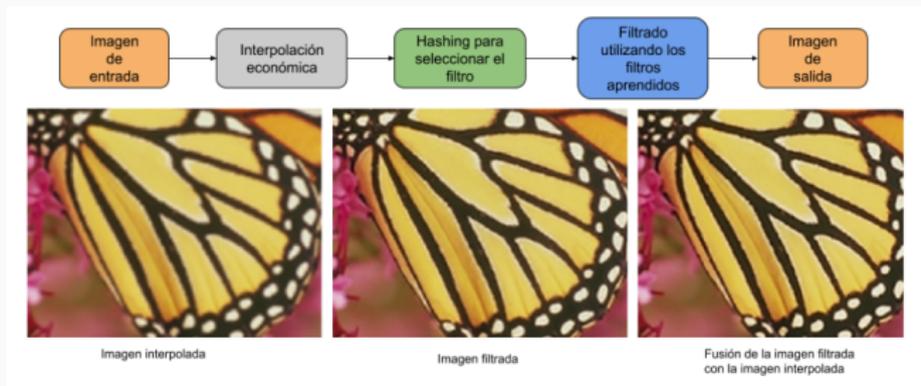


$$u(i, j) = S_{i, j}(k * \phi_p(u(x, y))), \quad p = 1, \dots, n, \quad (x, y) \in \Omega \subset \mathbb{R}^2$$

Trabajos Relacionados

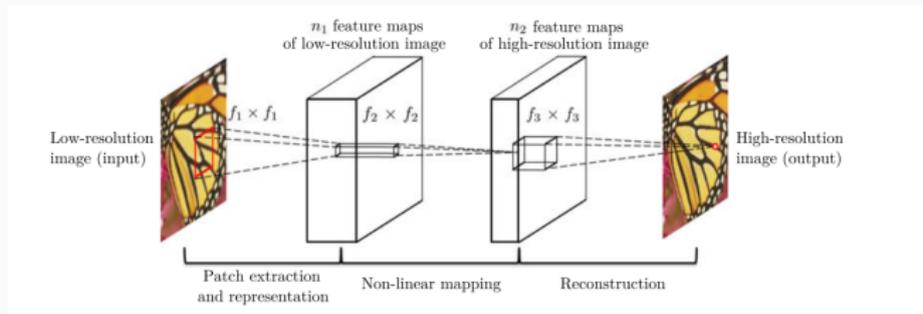
Super resolución Mono-Imagen

- RAISR - Romano et al.



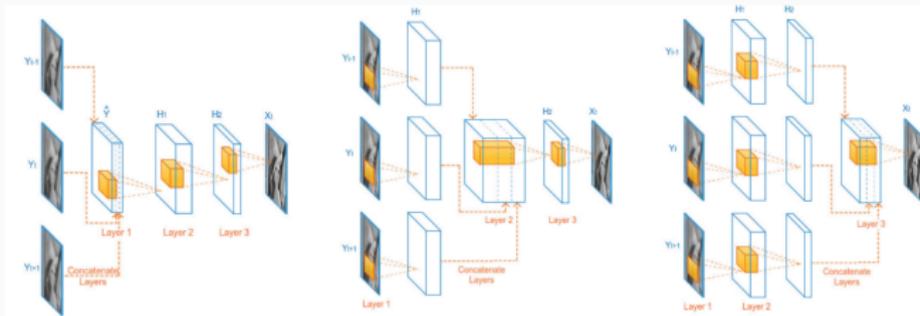
Super resolución Mono-Imagen

- Image super-resolution using deep convolutional network - Dong et al.



Super resolución Multi-Imagen

- Video super-resolution with convolutional neural networks - Kappeler et al.



Solución Propuesta y Resultados

Diferentes familias de soluciones:

- Métodos variacional tradicionales
- Redes neuronales:
 - Mono-Imagen
 - Multi-Imagen

Método variacional Shannon Total Variation (STV)

$$u_{HR} = \underset{u: \Omega \rightarrow \mathbb{R}}{\operatorname{argmin}} \|Au - u_0\|_2^2 + \lambda \operatorname{STV}_2(u) \quad (2)$$

donde,

- u_0 es la imagen observada
- A el operador lineal que modela la convolución con la respuesta al impulso del dispositivo de adquisición.
- λ el parámetro de regularización
-

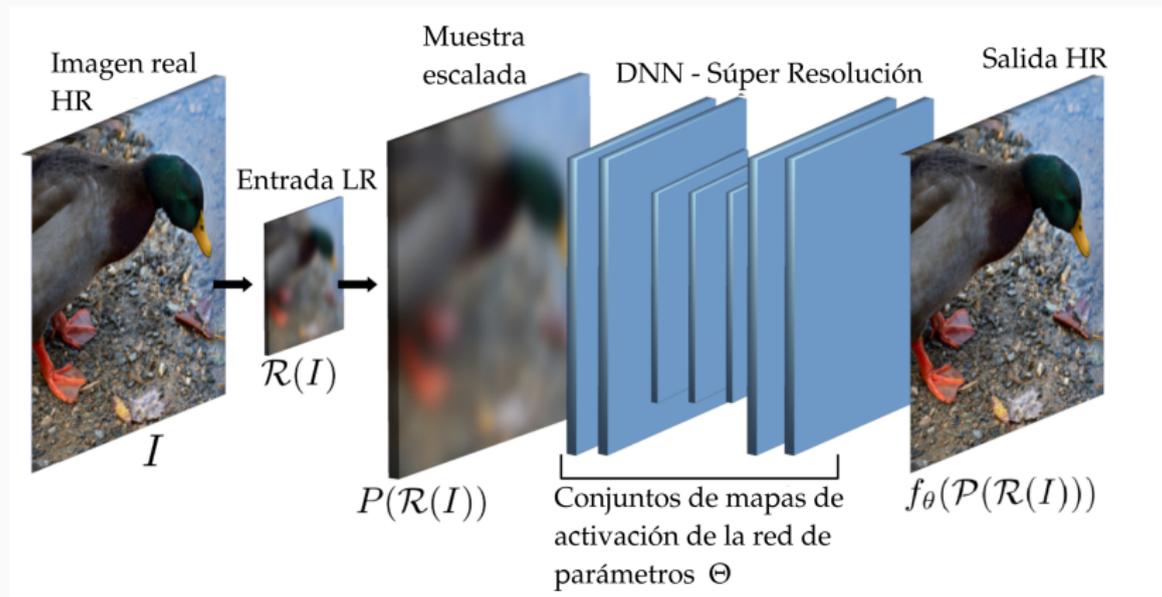
$$\operatorname{STV}_2(u) = \frac{1}{4} \sum_{(k,l) \in \Omega_2} \|\nabla_2 u(k,l)\|_2, \quad (3)$$

donde $\nabla_2 u(k,l) = \nabla U(\frac{k}{2}, \frac{l}{2})$ es el gradiente de la interpolación de Shannon de u en el punto $(k/2, l/2)$.

Se plantea el problema de forma dual y se utiliza para resolverlo la implementación *stvsuperres* de Abergel et al. la cuál se basa en el algoritmo Chambolle-Pock.

$$u_{HR} = \operatorname{argmin}_{u: \Omega \rightarrow \mathbb{R}} \max_{p: \Omega_2 \rightarrow \mathbb{R}, q \in \mathbb{R}^\omega} \left\langle \frac{\lambda}{4} \nabla_2 u, p \right\rangle + \langle Au, q \rangle - \delta_{\beta^*}(p) - \left\| \frac{q}{2} + u_0 \right\|_2^2. \quad (4)$$

Súper Resolución basada en redes neuronales profundas



DIV2K - Timofte et al.

- 1000 imágenes alta resolución
- Contenido variado: imágenes de personas, flora, fauna, entornos naturales, ciudades y objetos



MatDB - Matias Di Martino

- 14 imágenes de alta resolución
- Variedad de texturas



SatDB - Imágenes satelitales cortesía de la agencia espacial francesa (CNES)



Resolución sobre el terreno: 25cm
Simulación imagen trasladada.

PSNR

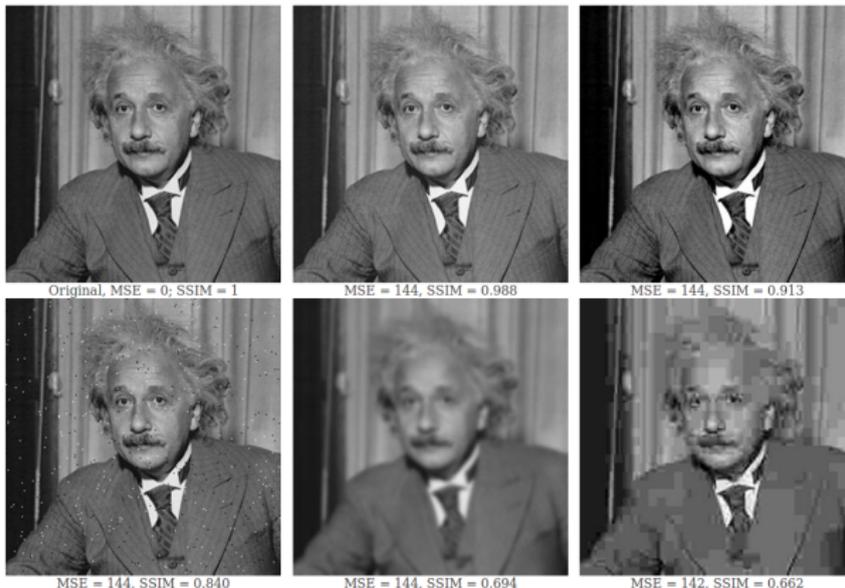
$$PSNR = 10 \log_{10} \left(\frac{I_{max}^2}{\frac{1}{n} \sum_x (I(x) - \hat{I}(x))^2} \right), \quad (5)$$

donde

- $I(x)$ imagen original
- $\hat{I}(x)$ imagen a comparar
- I_{max} máximo valor posible de la señal (ej: 255 para imágenes de 8 bits).

Mayor de PSNR significa mayor similitud de las imágenes que se están comparando.

SSIM



$$SSIM(X, Y) = l(X, Y)c(X, Y)s(X, Y)$$

Modela cualquier distorsión como una combinación de luminiscencia, contraste y estructura.

Súper Resolución multi-imagen: DDVS

Detail-revealing Deep Video Super-resolution

Xin Tao et al.



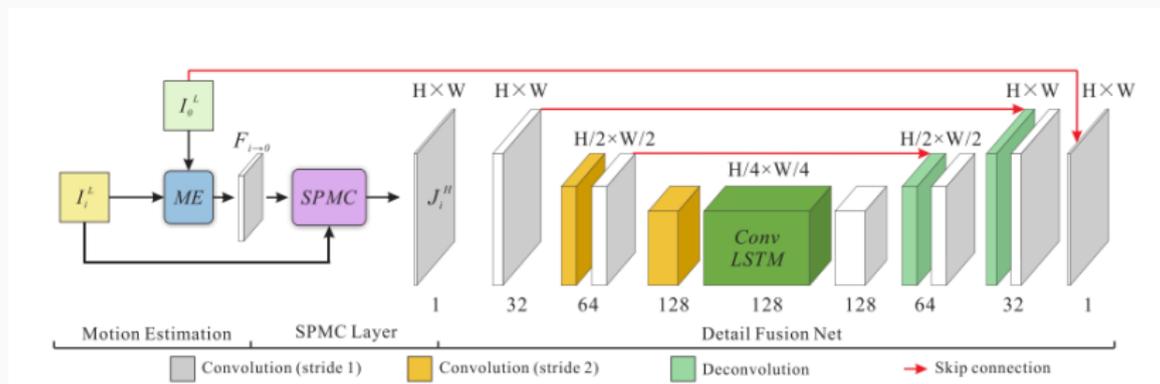
Bicubic $\times 3$



Using 3 frames #3

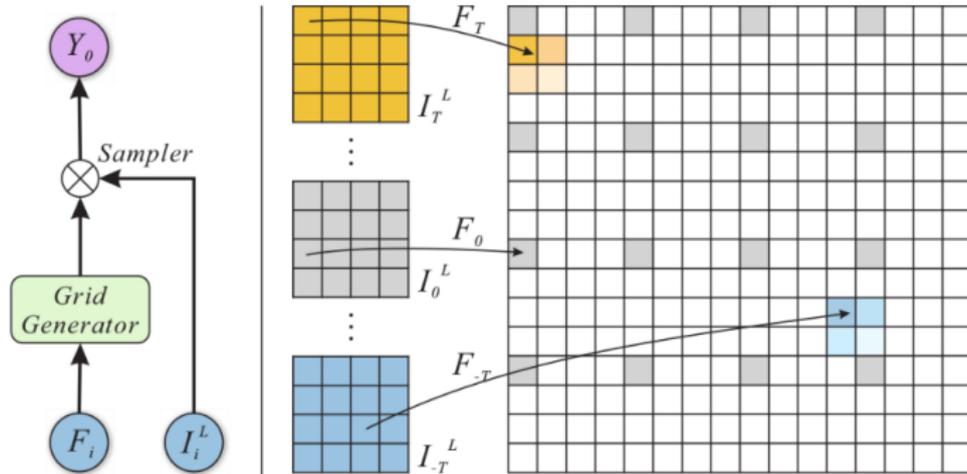
Súper Resolución multi-imagen: DDVS

- Se realiza la estimación de movimiento y la fusión de imágenes utilizando una red neuronal end-to-end.



Súper Resolución multi-imagen: DDVS

Capa de compensación de movimiento SMPC

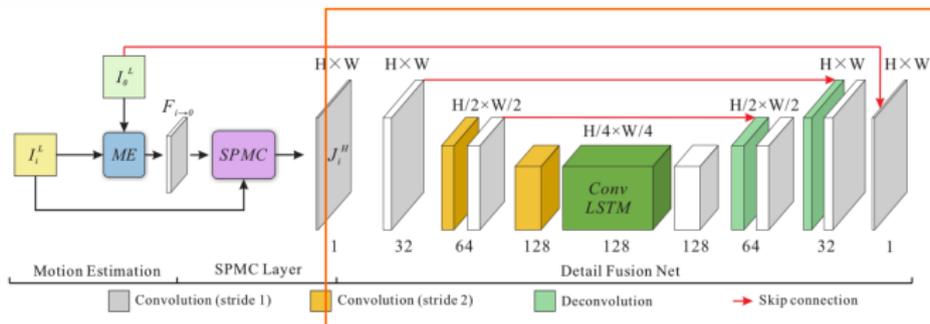


$$\begin{pmatrix} x_p^s \\ y_p^s \end{pmatrix} = W_{F,\alpha} \begin{pmatrix} x_p \\ y_p \end{pmatrix} = \alpha \begin{pmatrix} x_p + u_p \\ y_p + v_p \end{pmatrix}$$

$$J_q^H = \sum_{p=1} J_p^L M(x_p^s - x_q) M(y_p^s - y_q)$$

$$M(x) = \max(0, 1 - |x|).$$

Capa de fusión de detalles



$$\mathbf{f}_i = \text{Net}_E(J_i^H; \theta_E)$$

$$\mathbf{g}_i, \mathbf{s}_i = \text{ConvLSTM}(\mathbf{f}_i, \mathbf{s}_{i-1}; \theta_{LSTM})$$

$$I_0^{(i)} = \text{Net}_D(\mathbf{g}_i, \mathbf{S}_i^E; \theta_D) + I_0^{L\uparrow}$$

Función de costo:

- Primera etapa:

$$L_{ME} = \sum_{i=-T}^T \|I_i^L - \tilde{I}_{0 \rightarrow i}^L\|_1 + \lambda_1 \|\nabla F_{i \rightarrow 0}\|_1$$

- Segunda etapa:

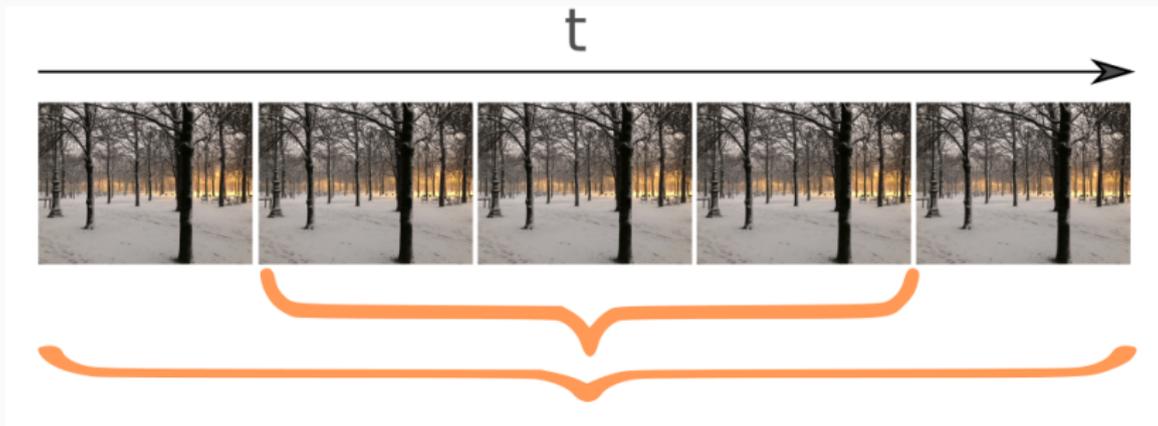
$$L_{SR} = \sum_{i=-T}^T k_i \|I_0^H - I_0^{(i)}\|_2^2$$

- Tercera etapa:

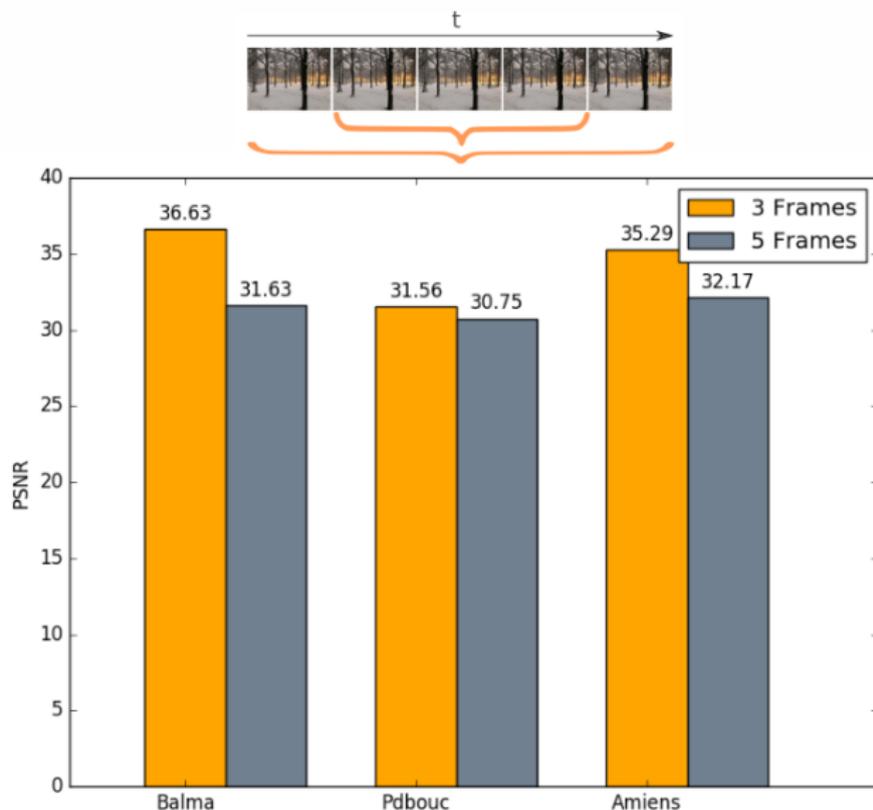
$$L = L_{SR} + \lambda_2 L_{ME}$$

- Se dispone del modelo ya entrenado
 - Imágenes de alta resolución sin ruido
 - Video submuestreado utilizando la interpolación bicúbica
- Se compara el resultado de la red con el algoritmo STV

Prueba 1: Número óptimo de imágenes



Prueba 1: Número óptimo de imágenes



Prueba 2: Imágenes con ruido

Resultado NN

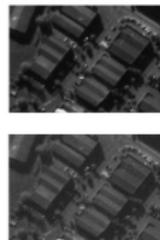
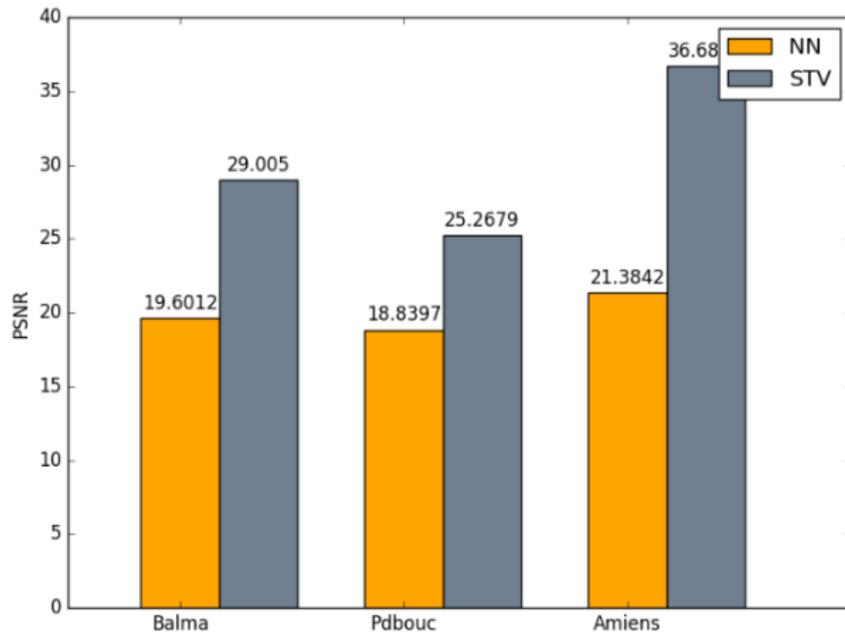


Prueba 2: Imágenes con ruido

Resultado STV

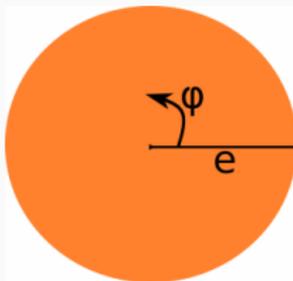


Prueba 2: Imágenes con ruido



Prueba 3: Ruido en las traslaciones

$$\varepsilon^j = (e \cos \varphi^j, e \sin \varphi^j), \varphi^j \sim \mathbb{U}[0, 2\pi]$$

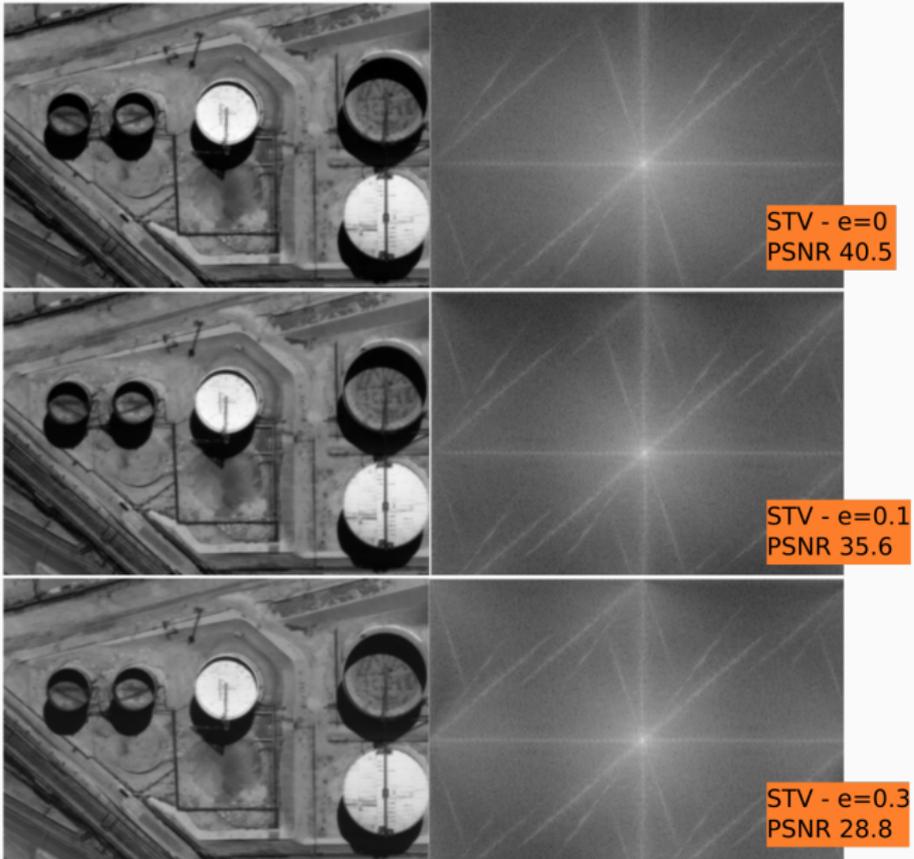


Se realiza la operación varias veces para obtener estadísticas del rendimiento del método STV para varios sorteos del error.

Prueba 3: Ruido en las traslaciones



Resultados - DDVS

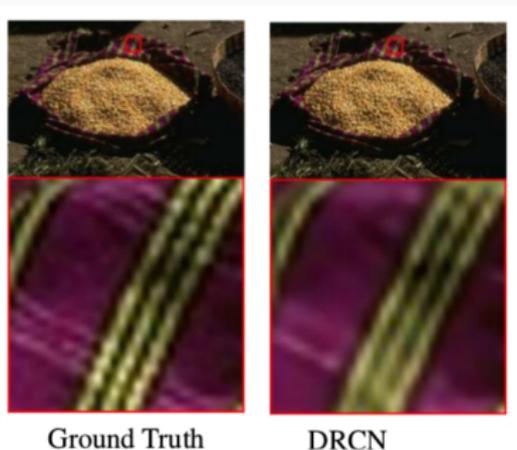


- Dado que las redes multi-imagen disponibles actualmente no presentan mejores resultados que los algoritmos variacionales, se estudian las redes mono-imagen con el fin de determinar si permiten obtener un mejor registrado, para luego aplicar algoritmos de SR estándar.

Súper Resolución mono-imagen: DRCN

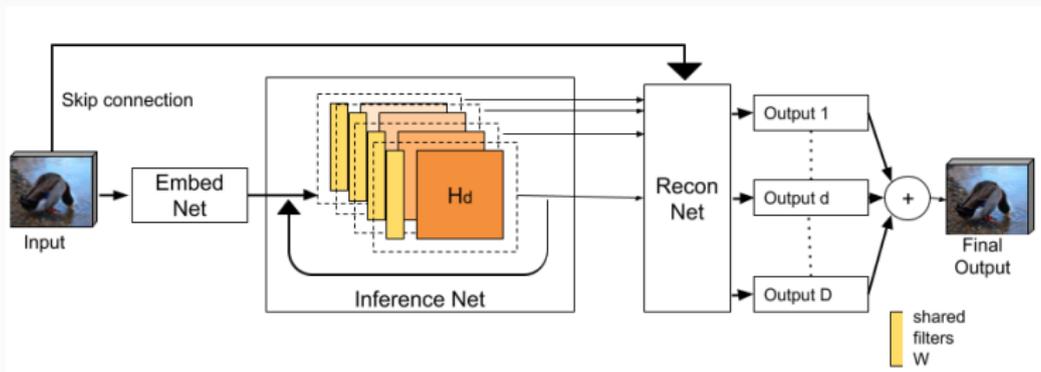
Deeply-Recursive Convolutional Network for Image Super-Resolution

Kim et al.



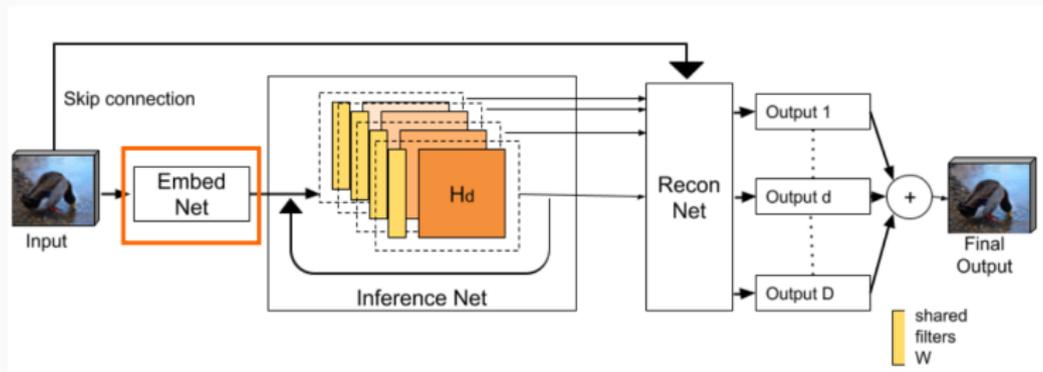
Súper Resolución mono-imagen: DRCN

- Red neuronal convolucional profunda con una capa recursiva de hasta 16 recurrencias



- Entrada: imagen x interpolada
- Salida: $\hat{y} = f(x)$
- $f(x) = f_3(f_2(f_1))$

Súper Resolución mono-imagen: DRCN

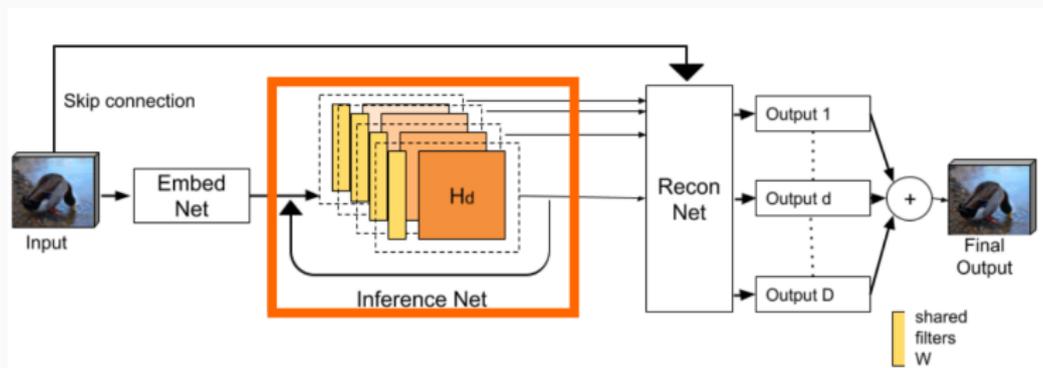


$$H_{-1} = \max(0, W_{-1} * x + b_{-1})$$

$$H_0 = \max(0, W_0 * H_{-1} + b_0)$$

$$f_1(x) = H_0$$

Súper Resolución mono-imagen: DRCN

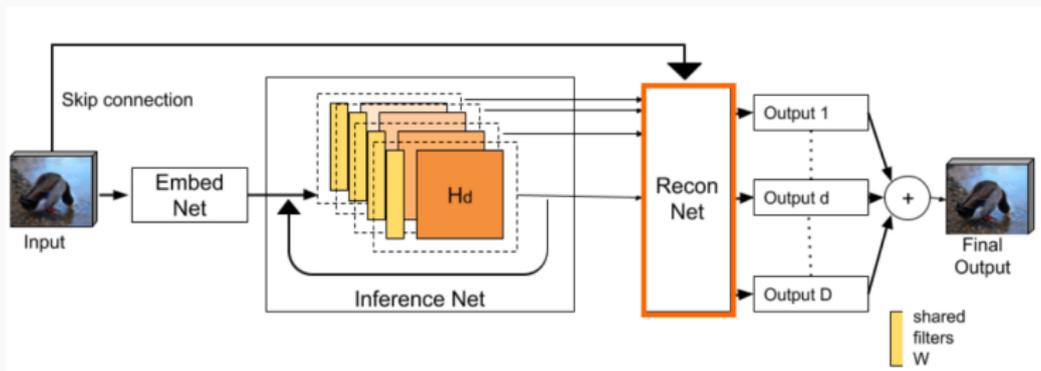


$$g(H) = \max(0, W * H + b)$$

$$H_d = g(H_{d-1}) = \max(0, W * H_{d-1} + b) \text{ para } d = 1, \dots, D$$

$$f_2(H) = (g \circ g \circ \dots \circ)g(H) = g^D(H)$$

Súper Resolución mono-imagen: DRCN

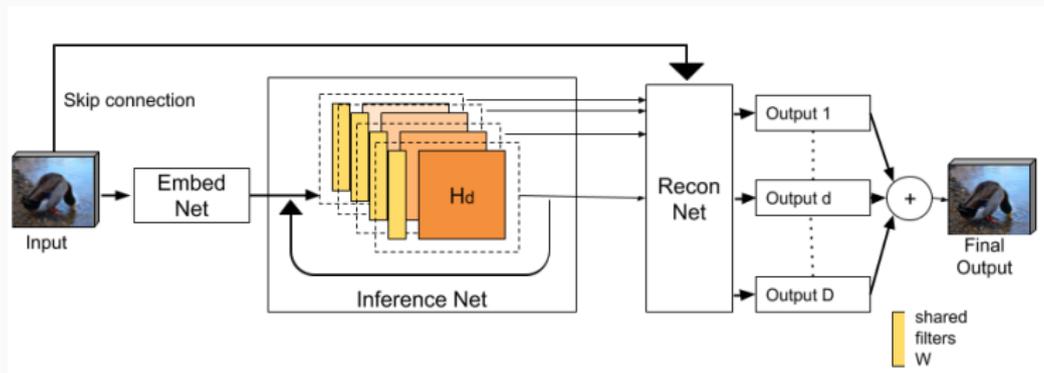


$$H_{D+1} = \max(0, W_{D+1} * H_D + b_{D+1})$$

$$\hat{y} = \max(0, W_{D+2} * H_{D+1} + b_{D+2})$$

$$f_3(H) = \hat{y}$$

Súper Resolución mono-imagen: DRCN



La salida final es el promedio de pesos de todas las predicciones intermedias:

$$\hat{y} = \sum_{d=1}^D w_d \cdot \hat{y}_d$$

$$\hat{y}_d = f_3(x, g^{(d)}(f_1(x))),$$

Función de costo

- Salidas intermedias:

$$l_1(\theta) = \sum_{d=1}^D \sum_{i=1}^N \frac{1}{2DN} \|y^{(i)} - \hat{y}_d^{(i)}\|^2$$

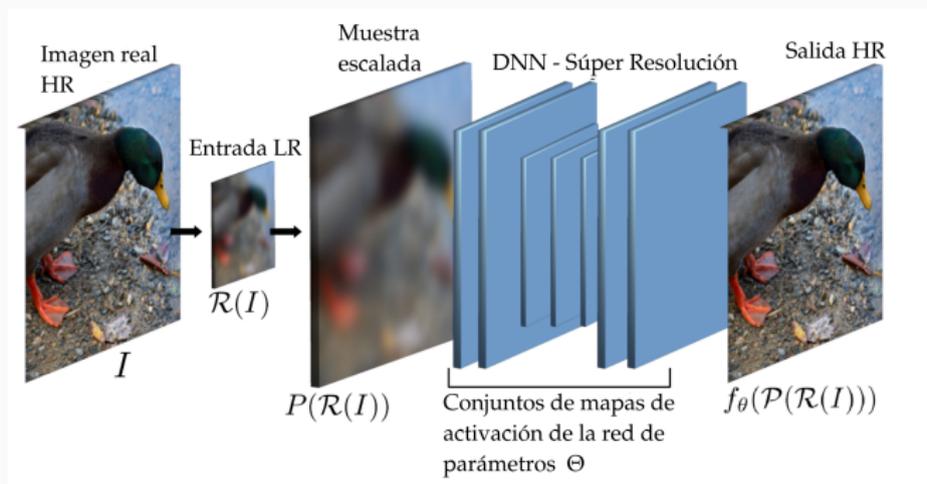
- Salida final:

$$l_2(\theta) = \sum_{i=1}^N \frac{1}{2N} \|y^{(i)} - \sum_{d=1}^D w_d * \hat{y}_d^{(i)}\|^2$$

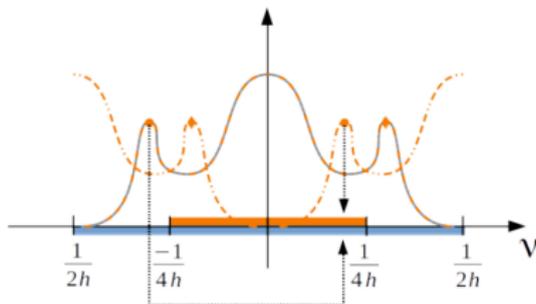
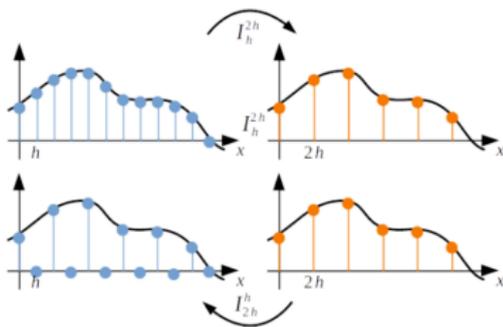
- Función de costo final:

$$L(\theta) = \alpha l_1(\theta) + (1 - \alpha) l_2(\theta) + \beta \|\theta\|^2$$

¿Cómo impacta la elección de los métodos de interpolación utilizados?



Restricción y Prolongación



$$R : I_{\Omega^h} \rightarrow I_{\Omega^{2h}}$$

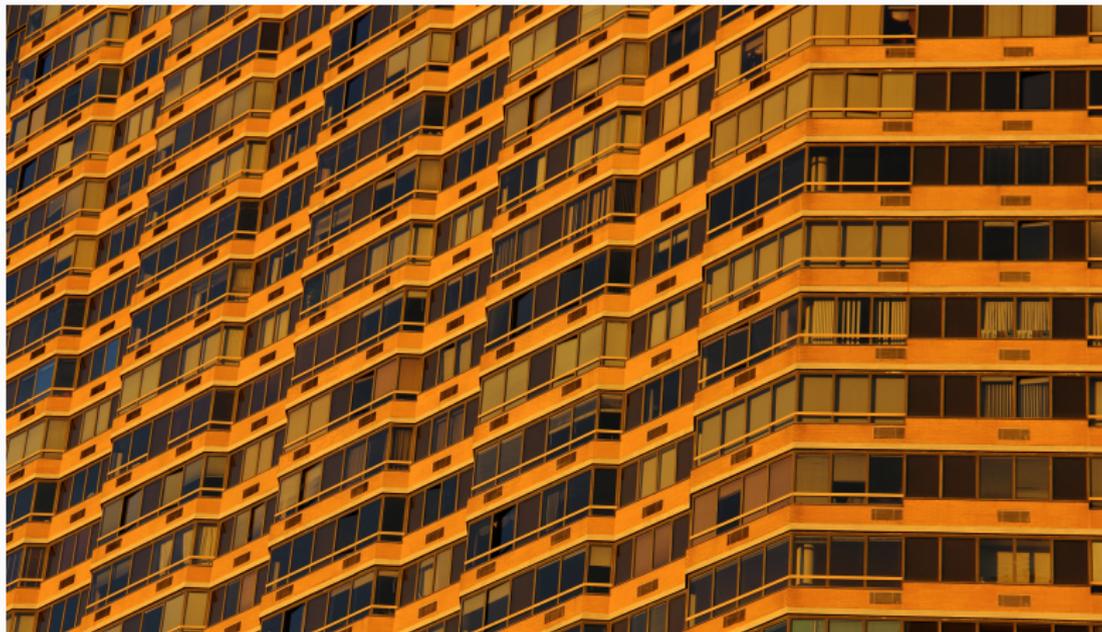
$$R(u^h) = R_0(K_h * u^h)$$

$$R_0(u^h)(i, j) = u^h(2i, 2j)$$

$$P : I_{\Omega^{2h}} \rightarrow I_{\Omega^h}$$

$$P(u^{2h}) = K^h * P_0(u^{2h}),$$

$$P_0(u^{2h})(i, j) = \begin{cases} u^{2h}(i/2, j/2) & \text{Si } i, j = \dot{2} \\ 0 & \text{en otro caso} \end{cases}$$

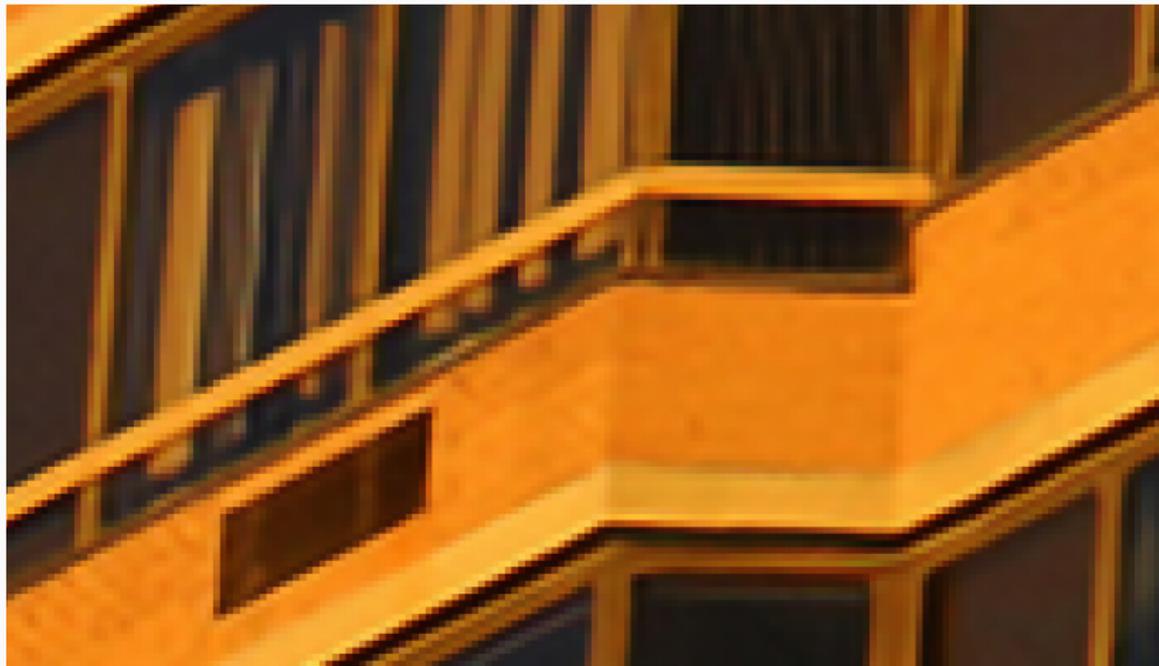


LR - P: Bicúbica



Resultados - DRCN

HR - P-train: Bicúbica - P-test: Bicúbica

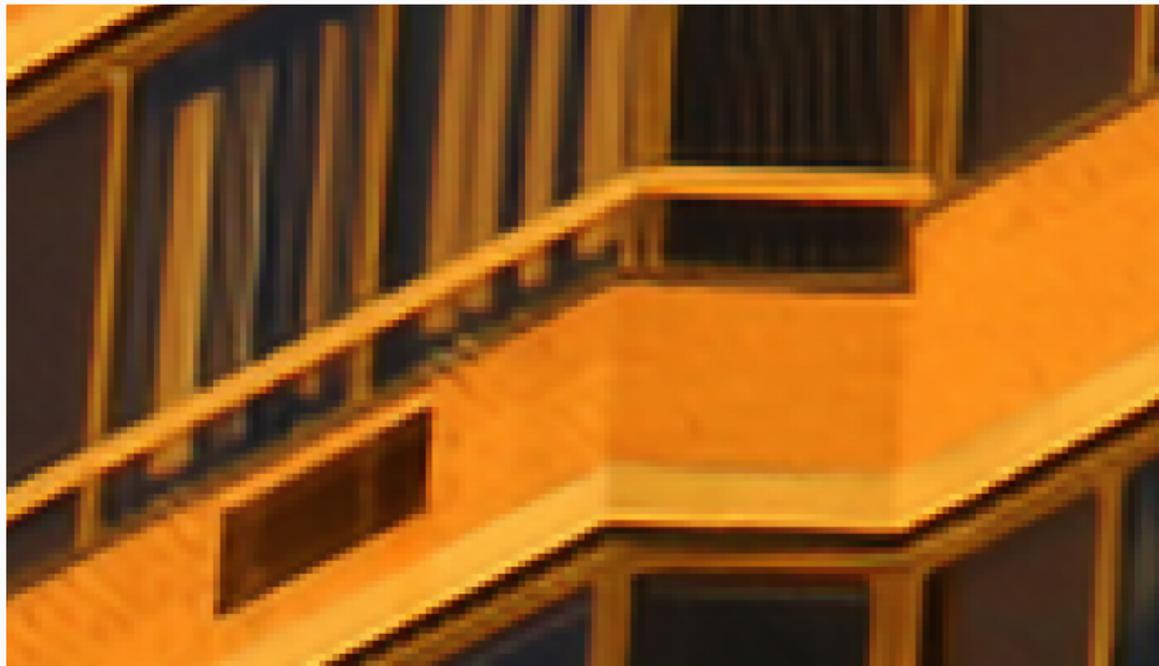


LR - P: Bicúbica



Resultados - DRCN

HR - P-train: Bicúbica - P-test: Seven





Interpolación Bicubica



HR - Bicúbica



Resultados - DRCN

LR - HR - Bicúbica



Resultados - DRCN

PSNR/SSIM según la interpolación utilizada

Base de Datos	Interpolación (Test)	Interpolación (Train)			
		Bilineal	Bicubic	Seven	Shannon
DIV2K	Bilineal	32.67/0.68	31.40/0.67	31.60/0.67	29.18/0.59
	Bicubic	31.47/0.70	32.73/0.70	30.09/0.69	30.36/0.63
	Seven	31.92/0.68	30.75/0.67	32.75/0.69	29.05/0.59
	Shannon	26.80/0.62	26.96/0.61	26.92/0.62	31.37/0.65
MatDB	Bilineal	37.59/0.83	35.70/0.82	37.89/0.83	35.05/0.78
	Bicubic	37.50/0.84	38.21/0.84	36.26/0.83	36.14/0.81
	Seven	37.19/0.82	35.44/0.81	38.31/0.83	34.89/0.78
	Shannon	34.79/0.79	35.14/0.79	34.36/0.78	36.70/0.82
SatDB	Bilineal	37.59/0.90	34.26/0.87	37.30/0.89	33.78/0.83
	Bicubic	36.97/0.89	38.21/0.91	34.98/0.87	34.88/0.86
	Seven	36.70/0.89	33.93/0.87	38.53/0.90	33.68/0.83
	Shannon	35.01/0.87	35.93/0.87	34.22/0.86	37.41/0.89
Promedio	Bilineal	35.95/0.80	33.79/0.79	35.60/0.80	32.67/0.73
	Bicubic	35.31/0.81	36.38/0.82	33.78/0.80	33.79/0.77
	Seven	35.27/0.80	33.37/0.78	36.53/0.81	32.54/0.73
	Shannon	32.20/0.76	32.68/0.76	31.83/0.75	35.16/0.79

PSNR/SSIM según la interpolación utilizada, utilizando imágenes con ruido

Base de Datos	Interpolación (Test)	Interpolación (Train)			
		Bilineal	Bicubic	Seven	Shannon
Promedio	Blineal	35.95/0.80	33.79/0.79	35.60/0.80	32.67/0.73
	Bicubic	35.31/0.81	36.38/0.82	33.78/0.80	33.79/0.77
	Seven	35.27/0.80	33.37/0.78	36.53/0.81	32.54/0.73
	Shannon	32.20/0.76	32.68/0.76	31.83/0.75	35.16/0.79
DIV2K noise	Blineal	32.83/0.62	31.16/0.60	32.42/0.62	29.88/0.56
	Bicubic	32.31/0.64	32.93/0.64	31.68/0.64	31.01/0.60
	Seven	32.19/0.62	31.07/0.59	32.82/0.63	29.83/0.56
	Shannon	28.05/0.60	29.35/0.61	28.85/0.61	31.48/0.63

Condiciones de prueba:

- Interpolación bicúbica
- Entrenamiento de la red (R y P) = interpolación (P) de la imagen de entrada
- Pilas de tres imágenes
- Tamaño 240×320

Comparación de las redes



Red	PSNR/SSIM
Mono-Imagen	27.37 / 0.94
Multi-Imagen	30.67 ± 0.27 / 0.96 ± 0.0



Red	PSNR / SSIM
Mono-Imagen	33.75 / 0.97
Multi-Imagen	40.55 ± 0.53 / 0.95 ± 0.0

Comparación de las redes



Comparación de las redes



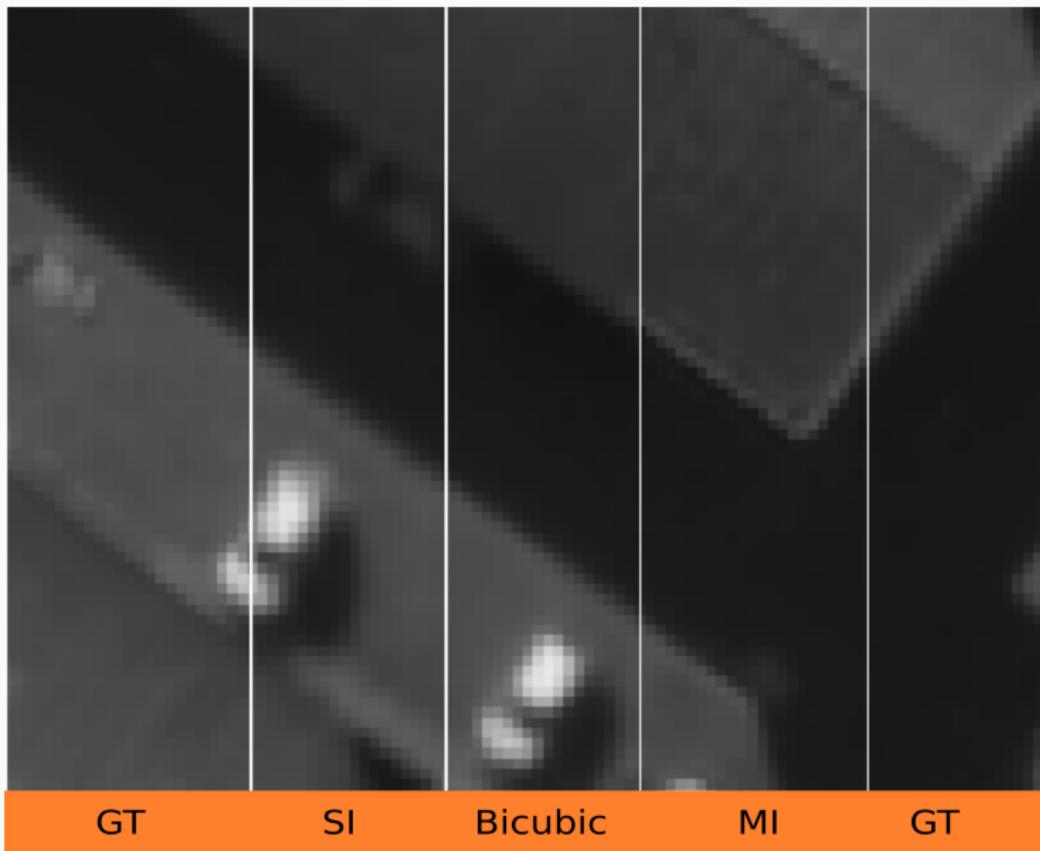
Comparación de las redes

Se evalúan los stacks de imágenes satelitales brindadas por el CNES, para tres imágenes en cada pila

Red	SI				MI
Train (R y P)	Bicubica	Bicubica	Shannon	Shannon	Bicubica
Test (P)	Bicubica	Shannon	Bicubica	Shannon	Bicubica
Amiens	33.08	34.05	34.43	35.25	36.54
Balma	37.87	40.9	39.66	42.14	39.37
Pdbouc	33.51	37.2	33.37	38.57	32.83
Promedio	34.82 ± 2.65	37.38 ± 3.42	35.82 ± 3.36	38.65 ± 3.44	36.24 ± 3.28

La elección adecuada de la interpolación tiene una ganancia en performance comparable con una red multi-imagen.

Comparación de las redes



Conclusiones

Estudio de factibilidad de utilizar redes neuronales para mejorar la resolución de las imágenes satelitales:

- El método de regularización STV presenta mejores resultados que las redes neuronales cuando se tiene un error en las traslaciones de menos de 0.3 pixeles.
- El tiempo de ejecución de las redes es significativamente inferior que para el algoritmo STV: 3 segundos contra aproximadamente 245.
- En caso de contar con tres imágenes con desplazamiento sub-píxel entre sí, el resultado de utilizar una red multi-imagen es de 5dB más que el obtenido al utilizar una red mono-imagen.

¿Cómo impacta la elección de los métodos de interpolación?

- El impacto en la performance al elegir la misma interpolación para crear la imagen sintética durante el entrenamiento que la elegida como interpolación para mostrar la imagen al comienzo de la red es de un promedio de 3dB más.
- Si se conoce el PSF de la cámara se debe elegir ese modelo para la restricción.
- En base a la experiencia empírica de este trabajo, elegir de núcleo para la prolongación el PSF de la cámara genera una ganancia promedio de 1dB, aunque no se puede asegurar que sea optima en todos los casos.

Preguntas
