

Super-Resolución de Imágenes: FSRCNN con Redes Residuales

Universidad de
San Andrés

Marcos Piotto (mpiotto@udea.edu.ar), Segundo Santos Torrado (ssantostorrado@udea.edu.ar)

Introducción

La super-resolución de imágenes individuales (*Single-Image Super-Resolution*) es un problema en visión por computadora que busca mejorar la resolución espacial de una imagen de baja resolución, generando una versión de alta resolución con más detalles y mayor nitidez.

Similaridad de Imágenes

Función de Costo

MSE (*Mean Squared Error*)

Métricas

PSNR (*Peak Signal-to-Noise Ratio*) mide la relación entre la potencia de la señal (la imagen original) y la potencia del ruido (las diferencias entre la imagen original y la reconstruida).

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right)$$

SSIM (*Structural Similarity Index Measure*) compara la luminancia, el contraste y la estructura de la imagen original y la degradada. El valor de SSIM varía de 0 a 1, donde 1 indica una coincidencia perfecta.



Figura 1: Se analiza que a MSE constante la percepción de calidad visual varía ampliamente.

Detalles de Implementación

- Datasets utilizados: *General100* y *T91* para entrenamiento, *BSD100* para validación, y *Set5* y *Set14* para métricas finales.
- Muestras por factor de escala: 1000 para $2\times$ y 10000 para $4\times$ para evitar *overfitting*.
- Tamaño de parche de 64 en todos los casos.

Conclusiones y Trabajo a Futuro

- La incorporación de conexiones residuales mejora el desempeño del modelo de super-resolución de imágenes FSRCNN.
- La versión compacta de FSRCNN ofrece la mejor relación cantidad de parámetros/*performance*.
- Exploración de soluciones del *state-of-the-art* de ISR como *ViT* o *GANs*.
- Búsqueda de hiperparámetros en todos los modelos.

Referencias

- [1] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network, 2016.

Metodología

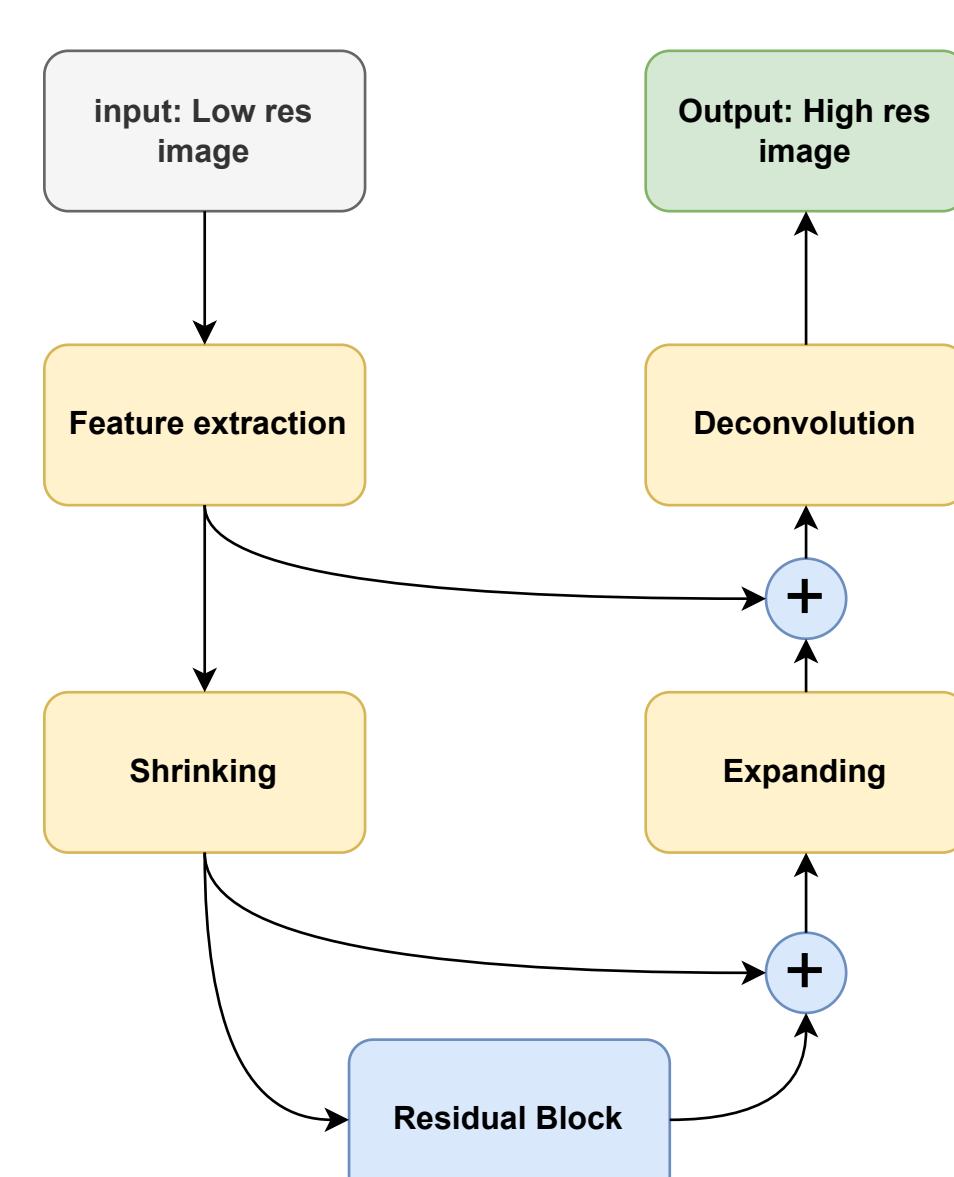


Figura 2: Arquitectura de RSRCNN (*Residual Super-Resolution Convolutional Neural Network*). Utiliza Conv2d, PreLU, BatchNorm2d, ConvTranspose2d

Factor de Escalado	Cantidad de Parámetros	
	$2\times$	$4\times$
FSRCNN (32, 5, 1)	2006	3542
FSRCNN (56, 12, 4)	9169	11857
RSRCNN (56, 12, 4)	14593	17281

Cuadro 1: Comparación de modelos implementados según su cantidad de parámetros entrenables.

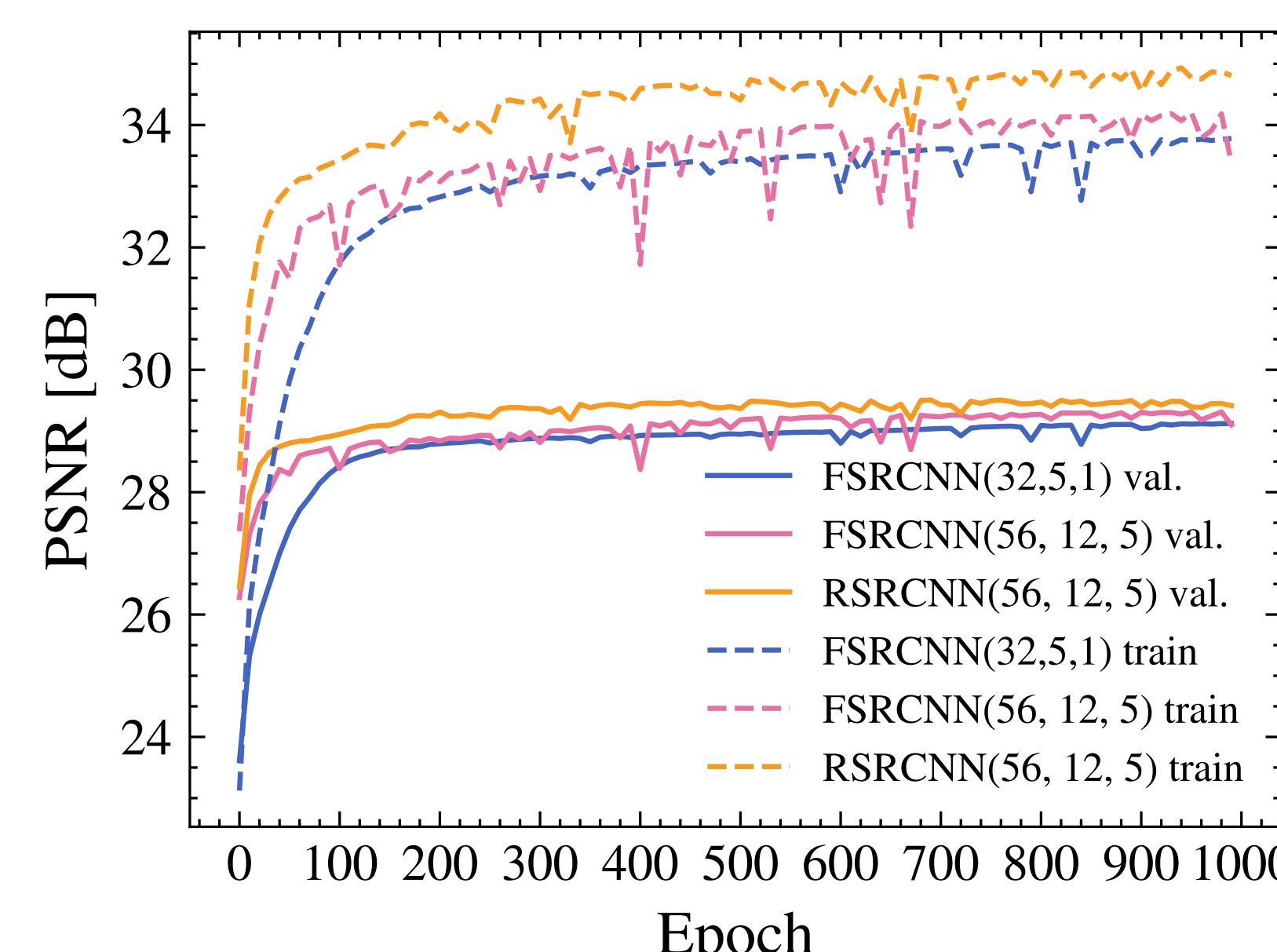


Figura 3: Curvas de PSNR de entrenamiento y validación para factor de escalado $2\times$ durante 1000 epochs. Se utilizó 1×10^{-3} de learning rate y batch size de 64.

Resultados

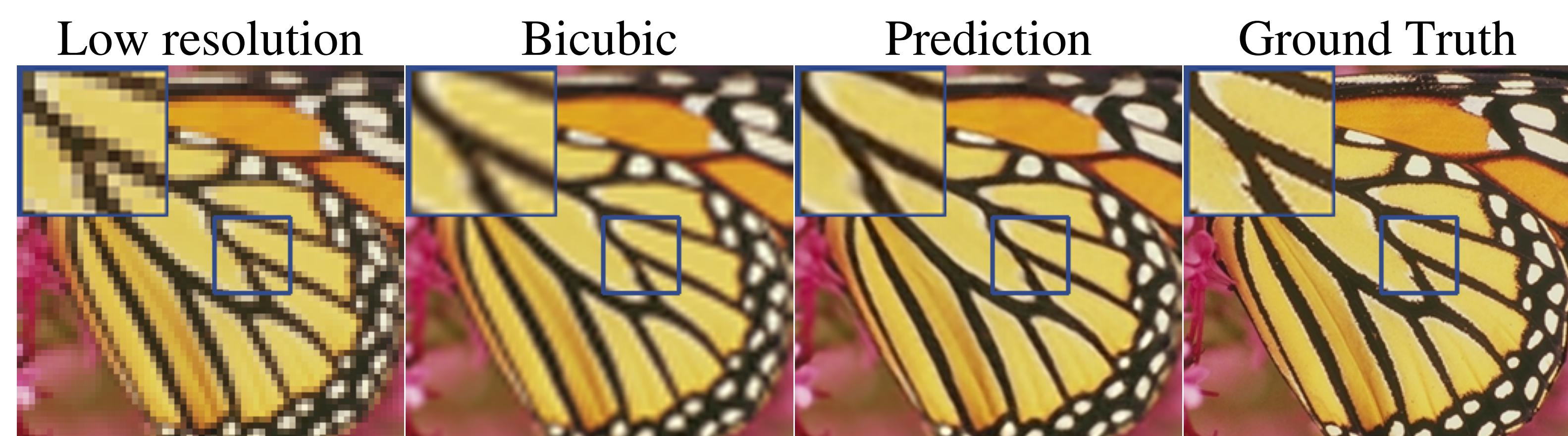


Figura 7: Resultados visuales de super-resolución con factor de escalado $4\times$. El canal Y (luma) es obtenido por el modelo RSRCNN, mientras que los canales Cb y Cr (croma) son ampliados empleando interpolación bicúbica.

Test Set	Factor de Escalado	FSRCNN (32, 5, 1)		FSRCNN (56, 12, 4)		RSRCNN (56, 12, 4)	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Set5	$2\times$	33.60	0.9419	34.26	0.9474	34.64	0.9488
	$4\times$	28.39	0.8365	26.85	0.7786	28.44	0.8399
Set14	$2\times$	29.75	0.8922	30.22	0.8959	30.40	0.8985
	$4\times$	25.44	0.7337	24.38	0.6838	25.59	0.7365

Cuadro 2: Evaluación de los modelos FSRCNN (en dos tamaños) y RSRCNN. Se analiza que RSRCNN minimiza las métricas de PSNR y SSIM en todos los apartados y FSRCNN tiene la mejor relación cantidad de parámetros/*performance*.

Data Augmentation

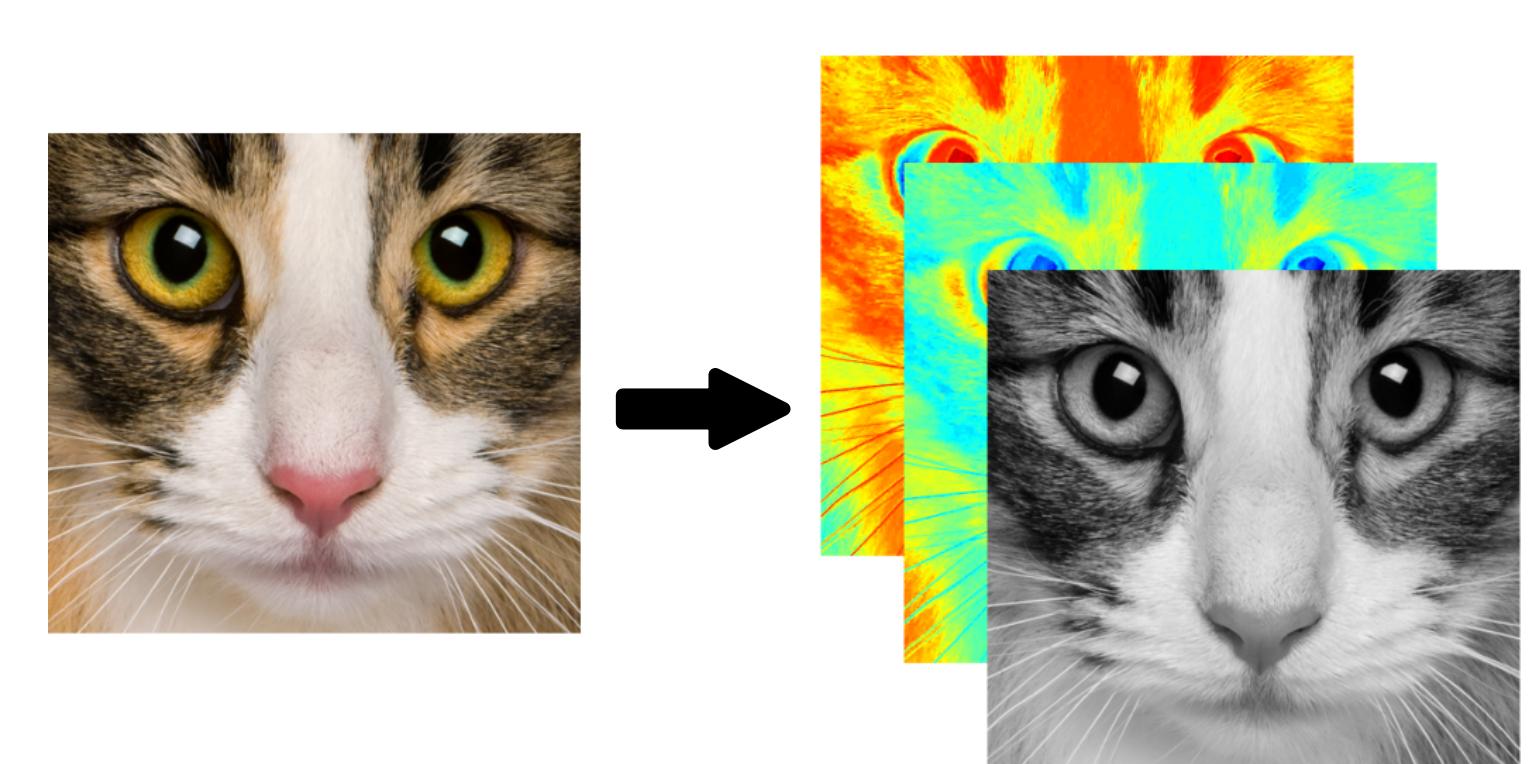


Figura 4: Transformación de espacio de color RGB a YCbCr (luma y croma). Se predice sobre el canal Y.



Figura 5: Rotación de 90° , 180° y 270° , y reescalado según factores de 0,9, 0,8, 0,7 y 0,6.

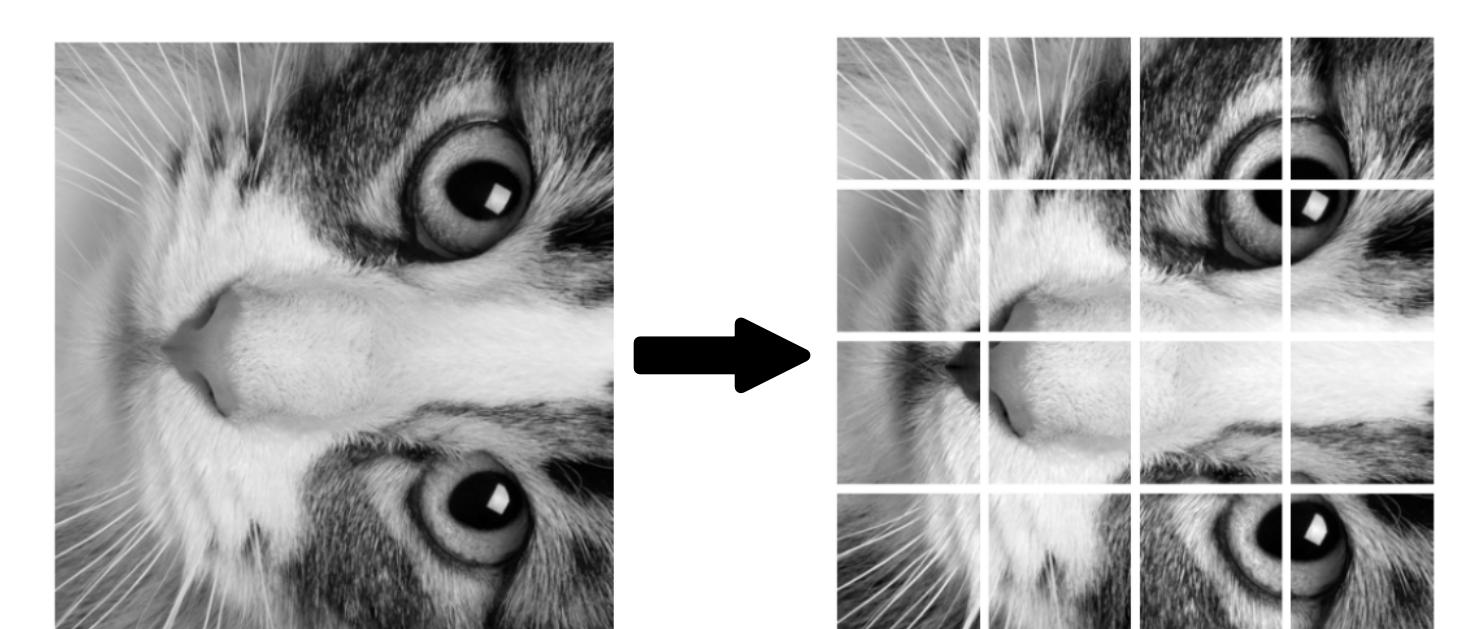


Figura 6: Fragmentación en parches uniformes para batches.