NIF: A Fast Implicit Image Compression with Bottleneck Layers and Modulated Sinusoidal Activations

Lorenzo Catania, Dario Allegra

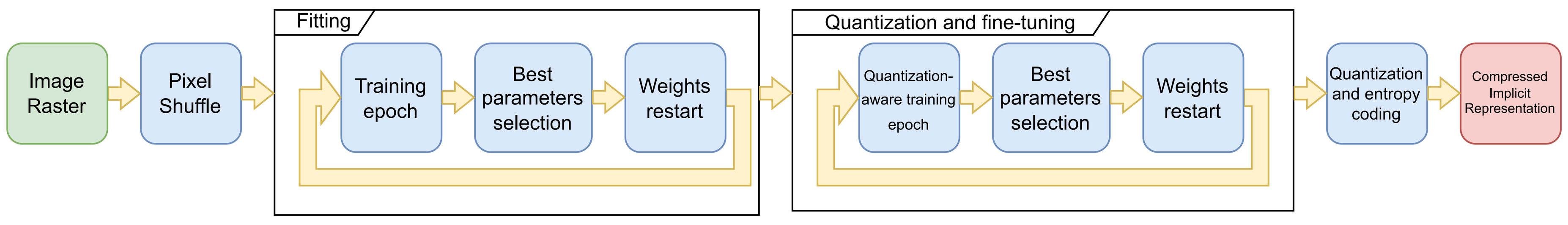
lorenzo.catania@phd.unict.it, dario.allegra@unict.it

Department of Mathematics and Computer Science, University of Catania

Overview

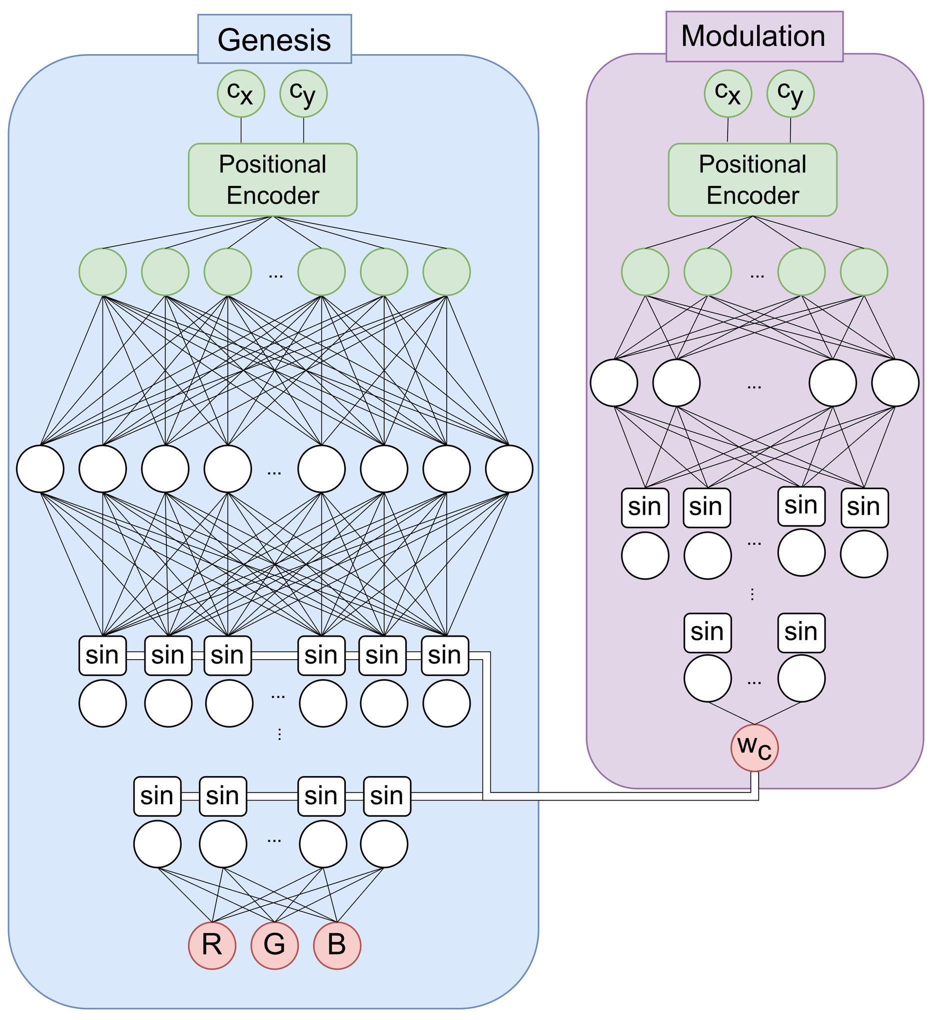
We propose a compression method based on Implicit Neural Representation, Neural Image Format (NIF), which relies on training a representational network that maps input coordinates to pixels. The network parameters are then quantized and compressed to minimize the size of encoded data. These methods are penalized by the computational cost to overfit a neural network on the signal and by the time required to obtain a compressed representation. Our proposal leverages various tweaks to aid the network during training and to reduce the time needed to obtain an optimal approximation of the target signal.

Training process



The training of a neural network typically consists of executing a number of epochs, each containing several iterations. However, many state-of-the-art implicit image compression approaches lack a clear subdivision of the overall training process. Our approach leverages a step-wise decomposition of both the fitting and quantization fine-tuning processes, as it enables one to perform some optimisations on the network at specific stages, such as weights restart.

Network architecture



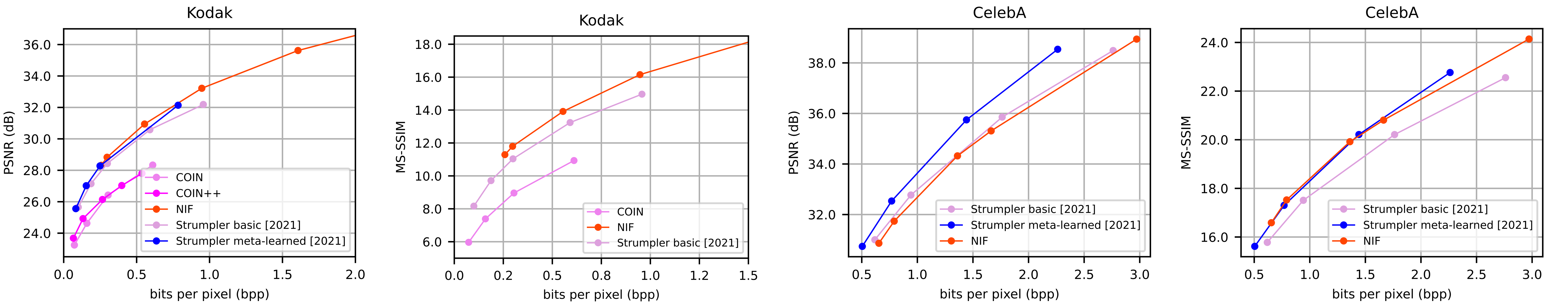
Our proposed architecture consists of two SIRENs[2]. A **Genesis** network responsible for calculating features of a pixel when fed with its coordinates. In contrast to traditional SIRENs, the period of each sinusoidal activation is altered based on the period variation provided by the **Modulation** network, a dedicated module that adjusts the hidden feature period, thereby adapting to variations in frequency across different regions of the image Therefore, the proposed activation function is:

$$\omega_c = \omega_0 + \sigma * f_m(c_x, c_y)$$

$$y_i = sin(\omega_c * (W_i * y_{i-1}^T + B_i))$$

Where σ is a hyper-parameter that scales period modulations, (c_x, c_y) are the coordinates given as input to the network from which the hidden features y_i are calculated, f_m is the function represented by the **Modulation network** that associates each coordinate to its period variation and ω_c is the resulting modulated period for coordinates (c_x, c_y).

Comparison to INR-based methods



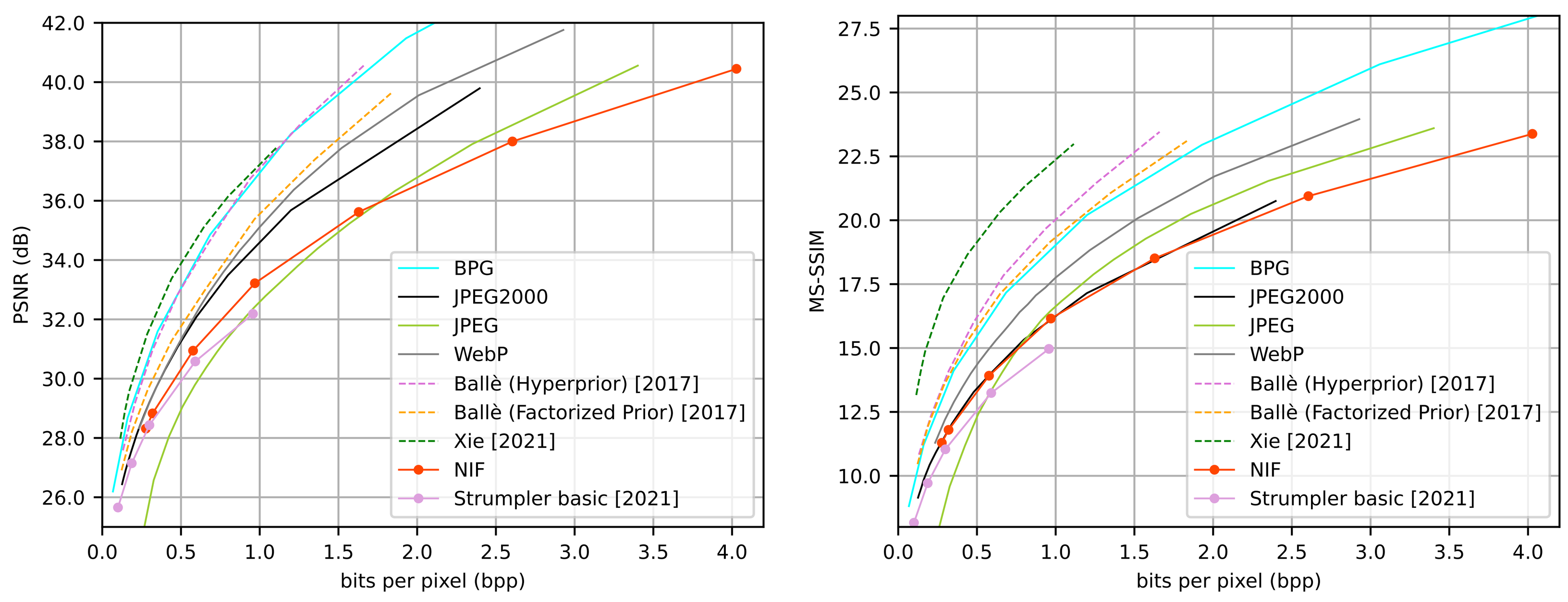
Our proposal NIF outperforms previous works on the field in terms of PSNR on Kodak while achieving comparable performance on CelebA. In terms of MS-SSIM, NIF consistently outperforms previous works on the field. The Strumpler et al. meta-learned approach is competitive on CelebA; this is not surprising since the meta-learning is performed on the same CelebA dataset, which is characterized by a limited images variability, leveraging consistent redundancies that are typically absent in real scenarios.

Encoding speed improvements

Downscale	×2					
Bits per pixel	~0.3bpp			~1.0bpp		
Method	Time	PSNR	SSIM	Time	PSNR	SSIM
COIN [1]	26'19"	28.84dB	0.777	55'45"	29.37dB	0.771
Strumpler [3]	70'15"	**35.00dB**	**0.920**	77'55"	**40.15dB**	0.970
NIF (Ours)	**2'42"**	32.33dB	0.891	**3'32"**	38.91dB	**0.972**
Downscale	×4					
Bits per pixel	~0.3bpp			~1.0bpp		
Method	Time	PSNR	SSIM	Time	PSNR	SSIM
COIN [1]	8'27"	33.88dB	0.917	16'4"	22.53dB	0.475
Strumpler [3]	22'17"	**42.39dB**	**0.984**	24'15"	46.73dB	0.993
NIF (Ours)	**2'24"**	39.09dB	0.977	**2'27"**	**47.35dB**	**0.997**

The advantage in terms of execution time with respect to previous methods is evident, especially at lower bitrates. For instance, at 0.3 bpp to encode x2 downsampled pictures NIF has about **x26** compression speed. At 1.0 bpp and x4 downsampled images, our proposal NIF achieves the best results in terms of metrics with a large time saving (**x10**).

Comparison to traditional codecs



Our proposal NIF clearly outperforms JPEG and performs comparably to modern codecs, such as BPG and WebP, in terms of PSNR at low bitrates. In terms of MS-SSIM, NIF performs similarly to JPEG2000. These results clearly reduce the gap between INR-based compression and classical methods with respect to the previous baseline given by the basic version of Strumpler[3], which is reported in the plot for reference.

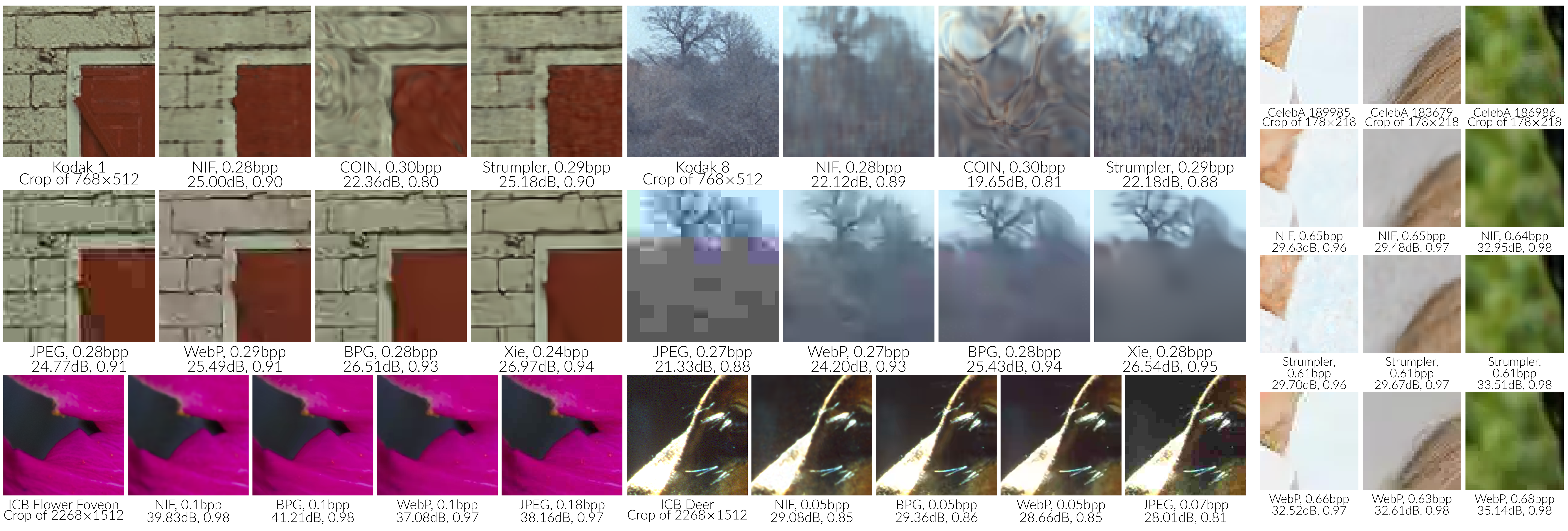
Hi-res image compression

bpp	Codec	PSNR	MS-SSIM
0.09bpp	JPEG	26.18dB	0.81
0.08bpp	WebP	28.84dB	0.89
0.08bpp	BPG	32.33dB	0.92
0.08bpp	NIF (Ours)	29.25dB	0.85
0.24bpp	JPEG	32.00dB	0.92
0.22bpp	WebP	32.66dB	0.93
0.21bpp	BPG	34.69dB	0.94
0.22bpp	NIF (Ours)	31.84dB	0.89

On the high-resolution images ICB dataset our proposal NIF outperforms JPEG and WebP in terms of PSNR at low bitrates, with a gain of +3.07dB against JPEG. Concerning higher bitrates, NIF is comparable to WebP and JPEG, but is still outperformed by BPG.

Visual comparisons

Values for bits-per-pixel, PSNR and MS-SSIM are reported below each image. These comparisons show that the decompressed images are less noisy with respect to previous INR-based methods and do not suffer from well-known distortions such as blocking artifacts, which are instead common in traditional approaches.



Kodak 1 Crop of 768×512 | NIF, 0.28bpp 25.00dB, 0.90 | COIN, 0.30bpp 22.36dB, 0.80 | Strumpler, 0.29bpp 25.18dB, 0.90 | Kodak 8 Crop of 768×512 | NIF, 0.28bpp 22.12dB, 0.89 | COIN, 0.30bpp 19.65dB, 0.81 | Strumpler, 0.29bpp 22.18dB, 0.88

JPEG, 0.28bpp 24.77dB, 0.91 | WebP, 0.29bpp 25.49dB, 0.91 | BPG, 0.28bpp 26.51dB, 0.93 | Xie, 0.24bpp 26.97dB, 0.94 | JPEG, 0.27bpp 21.33dB, 0.88 | WebP, 0.27bpp 24.20dB, 0.93 | BPG, 0.28bpp 25.43dB, 0.94 | Xie, 0.28bpp 26.54dB, 0.95

ICB Flower Foveon Crop of 2268×1512 | NIF, 0.1bpp 39.83dB, 0.98 | BPG, 0.1bpp 41.21dB, 0.98 | WebP, 0.1bpp 37.08dB, 0.97 | JPEG, 0.18bpp 38.16dB, 0.97 | ICB Deer Crop of 2268×1512 | NIF, 0.05bpp 29.08dB, 0.85 | BPG, 0.05bpp 29.36dB, 0.86 | WebP, 0.05bpp 28.66dB, 0.85 | JPEG, 0.07bpp 28.01dB, 0.81

CelebA 189985 Crop of 178×218 | CelebA 183679 Crop of 178×218 | CelebA 186986 Crop of 178×218

NIF, 0.65bpp 29.63dB, 0.96 | NIF, 0.65bpp 29.48dB, 0.97 | NIF, 0.64bpp 32.95dB, 0.98

Strumpler, 0.61bpp 29.70dB, 0.96 | Strumpler, 0.61bpp 29.67dB, 0.97 | Strumpler, 0.61bpp 33.51dB, 0.98

WebP, 0.66bpp 32.52dB, 0.97 | WebP, 0.63bpp 32.61dB, 0.98 | WebP, 0.68bpp 35.14dB, 0.98

References

[1] Emilien Dupont, Adam Golinski, Milad Alizadeh, Yee Whye Teh, and Arnaud Doucet. Coin: Compression with implicit neural representations. In *Neural Compression: From Information Theory to Applications–Workshop@ ICLR 2021*, 2021.

[2] Vincent Sitzmann, Julien N.P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Proc. NeurIPS*, 2020.

[3] Yannick Strümpler, Janis Postels, Ren Yang, Luc Van Gool, and Federico Tombari. Implicit neural representations for image compression. In *European Conference on Computer Vision*, pages 74–91. Springer, 2022.