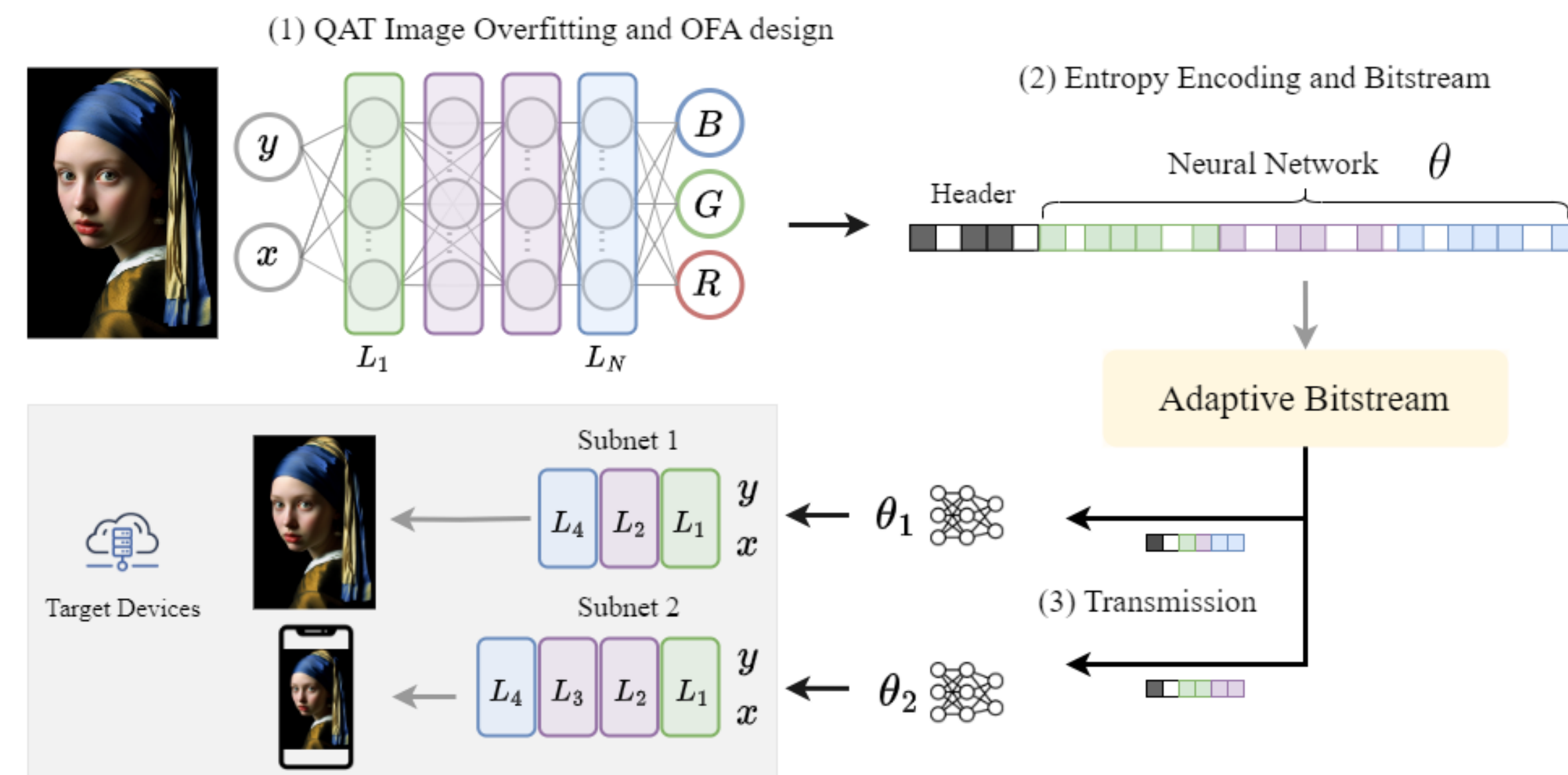


Introduction

Implicit Neural Images as novel image compression

No theoretical or practical method predicts the best INR model that fits a given signal



Contributions

Adaptive Neural Images SOTA neural representation for different inference or transmission requirements

4-bit quantization Reduced bit-per-pixel by 8 times, unharmed fidelity

Transversal framework Applicable to any INR architecture or modality, not limited to image

Method

SIREN^[1] and MFN^[2] as base architectures

Quantization-aware training with learnable scaling factors^[3] for near lossless compression

$$\bar{x} = \text{quantize} \left(\text{clamp} \left(\frac{x}{s}, -1, 1 \right) \right), \hat{x} = \bar{x} \times s$$

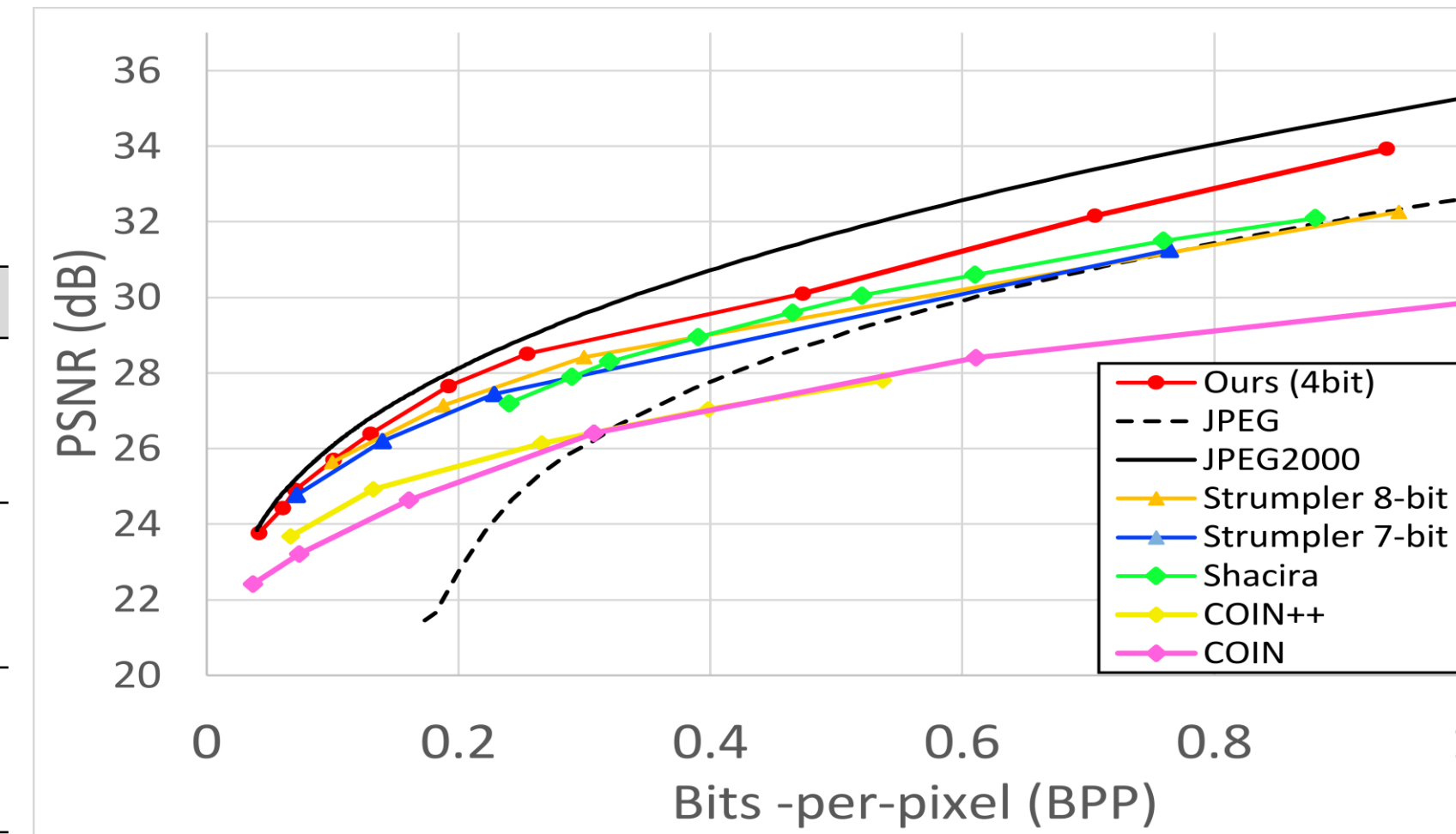
$$\bar{x} = \frac{\text{round}((x + 1) \times 2^{n-1})}{2^n}, n = \text{number of bits}$$

Once-for-All^[4] (OFA) SuperNet approach with 16 possible architectures options to serve multiple bpp/PSNR requirements

Experiments

4-bits model achieves SOTA PSNR/bpp tradeoff

BPP	Bitwidth	l×ch	PSNR↑
0.1	8-bit	3×64	25.36±2.64
	4-bit	2×128	26.39±2.62
0.5	8-bit	3×64	25.36±2.64
	4-bit	2×128	26.39±2.62
1.0	8-bit	3×64	25.36±2.64
	4-bit	2×128	26.39±2.62



SIREN Full Precision [5]

MFN Full Precision



PTQ 8-bit

PTQ 4-bit

PTQ 8-bit

PTQ 4-bit

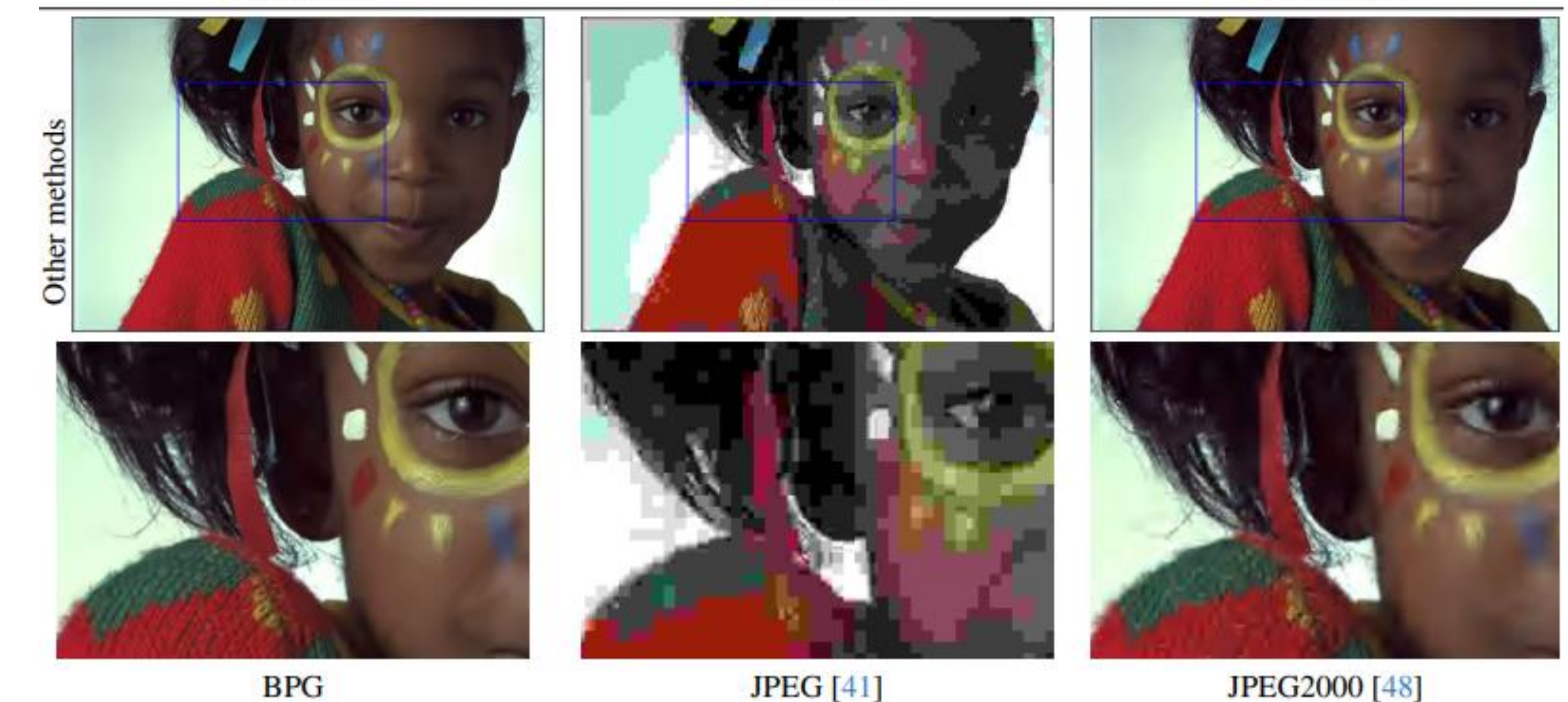
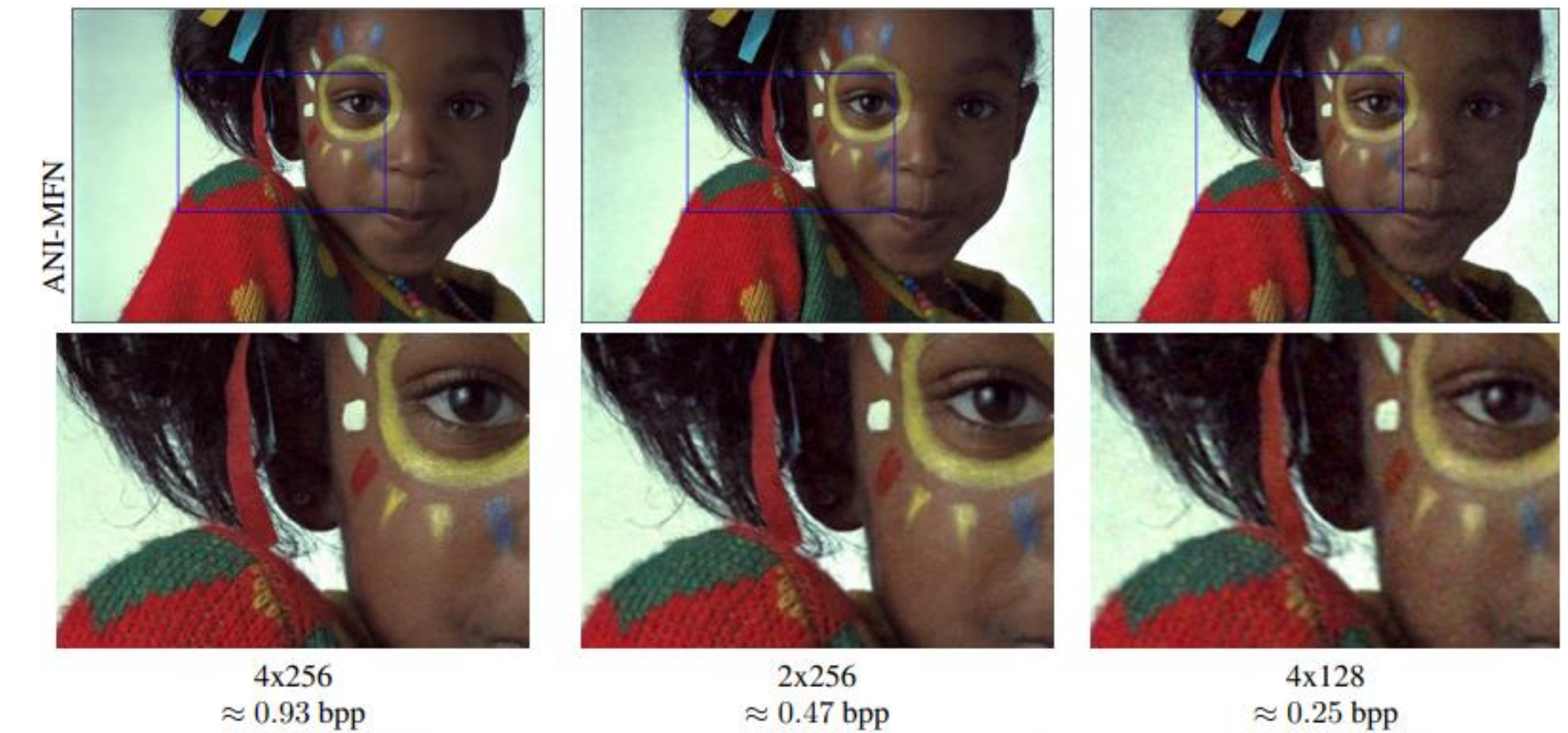


QAT 8-bit

QAT 4-bit

QAT 8-bit

QAT 4-bit



JPEG, JPEG2000 ≈ 0.23 bpp

References

- [1] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. Advances in neural information processing systems, 33:7462–7473, 2020.
- [2] Rizal Fathony, Anit Kumar Sahu, Devin Willmott, and J Zico Kolter. Multiplicative filter networks. In International Conference on Learning Representations, 2020.
- [3] Steven K Esser, Jeffrey L McKinstry, Deepika Bablani, Rathinakumar Appuswamy, and Dharmendra S Modha. Learned step size quantization. 2019.
- [4] Han Cai, Chuang Gan, Tianzhe Wang, Zhekai Zhang, and Song Han. Once-for-all: Train one network and specialize it for efficient deployment. 2019.
- [5] Emilian Dupont, Adam Golinski, Milad Alizadeh, Yee Whye Teh, and Arnaud Doucet. Coin: Compression with implicit neural representations. arXiv preprint arXiv:2103.03123, 2021.
- [6] Marcos V Conde, Andy Bigos, and Radu Timofte. Streaming neural images. In 2024 IEEE International Conference on Image Processing (ICIP), pages 2034–2040. IEEE, 2024.