Lossy Image Compression with Compressive Autoencoders

L. Theis, W. Shi, A. Cunningham, F. Huszár

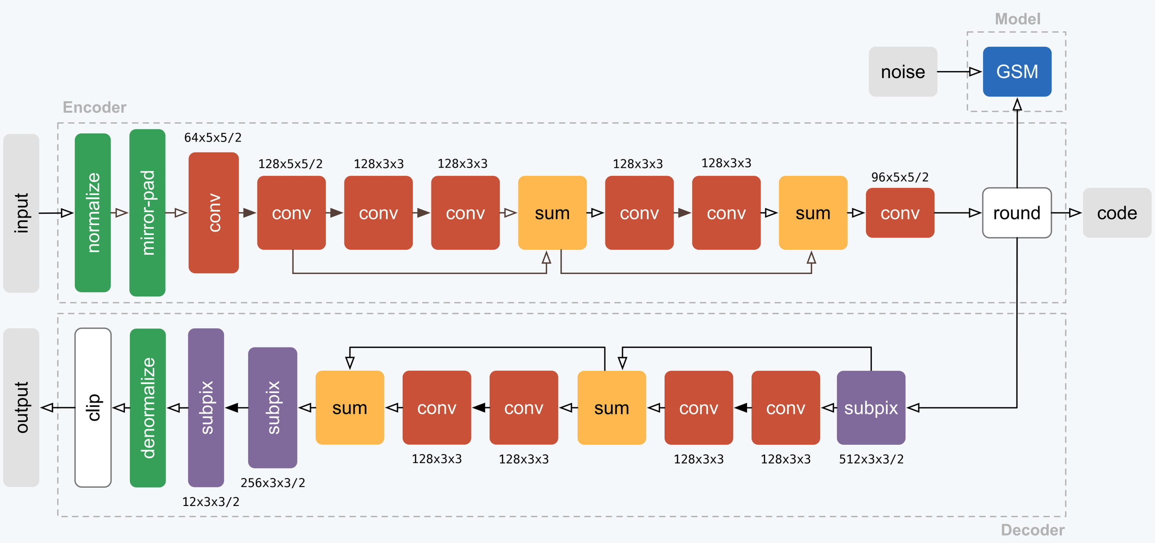
Twitter, London, UK

Introduction

Despite many advances in deep learning, the best **lossy compression** algorithms are still based on handcrafted algorithms.
This is largely due to the **non-differentiability of the rate-distortion tradeoff**.

The bit-rate as well as the quantization, [·], are non-differentiable. We explore a simple approach based on a differentiable approximation of the bit-rate and a redefinition of the derivative of quantization.

Compressive Autoencoder



A compressive autoencoder consists of an encoder f, a decoder g, and a probabilistic model Q:

$$f:\mathbb{R}^N \rightarrow \mathbb{R}^M, \quad g:\mathbb{R}^M \rightarrow \mathbb{R}^N, \quad Q:\mathbb{Z}^M \rightarrow [0,1]$$

The probabilistic model is used to assign a number of bits to coefficients produced by the encoder. Our decoder is based on sub-pixel convolutions [1].

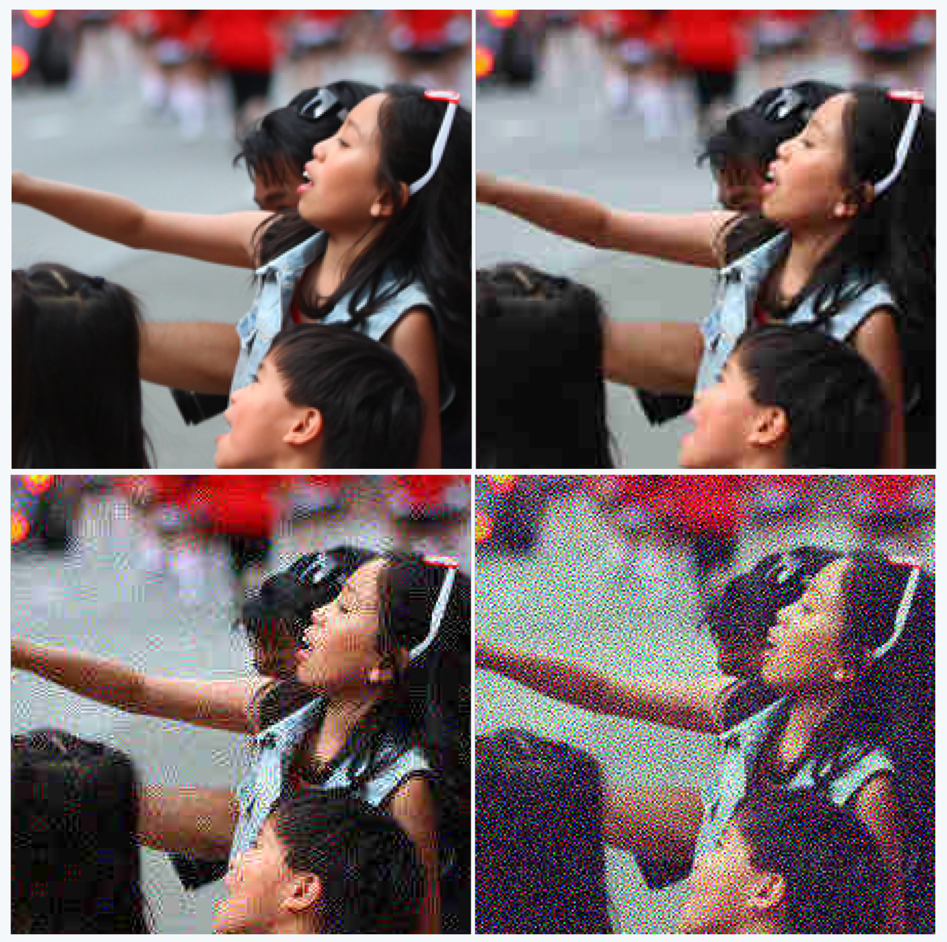
Quantization

One strategy previously explored is to replace quantization by additive noise [e.g., 1, 2].

$$[f(\mathbf{x})] \approx f(\mathbf{x}) + \boldsymbol{\varepsilon}$$

However, this introduces artefacts which can be perceptually very different from quantization artefacts, and therefore leads to biased distortion estimates.

Original | Rounding



Stochastic rounding | Uniform noise

Instead of replacing the quantization, we keep it in the forward pass and replace it's derivative in the backward pass:

$$\frac{d}{dy}[y] := \frac{d}{dy} r(y)$$

In practice, we found the identity to work well for r.

Bit-rate estimation

We express the discrete and non-differentiable Q in terms of a density q:

$$Q(\mathbf{z}) = \int_{[-.5,.5[^M} q(\mathbf{z}+\mathbf{u})\, d\mathbf{u}.$$

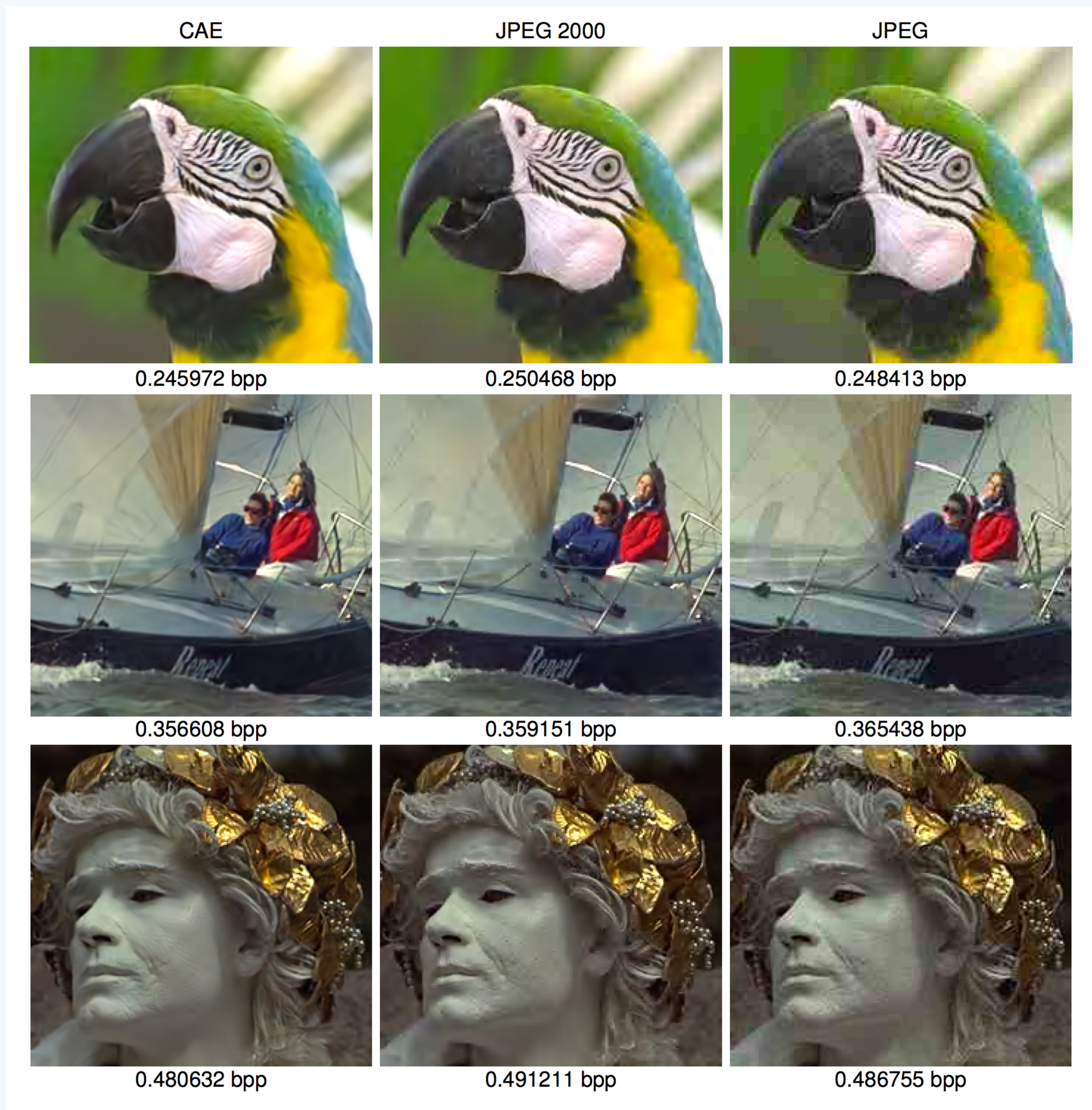
We optimize an upper bound on the bit-rate:

$$-\log_2 Q(\mathbf{z}) \leq \int_{[-.5,.5[^M} -\log_2 q(\mathbf{z}+\mathbf{u})\, d\mathbf{u},$$

Here, we use Gaussian scale mixtures to model the marginal distribution of coefficients:

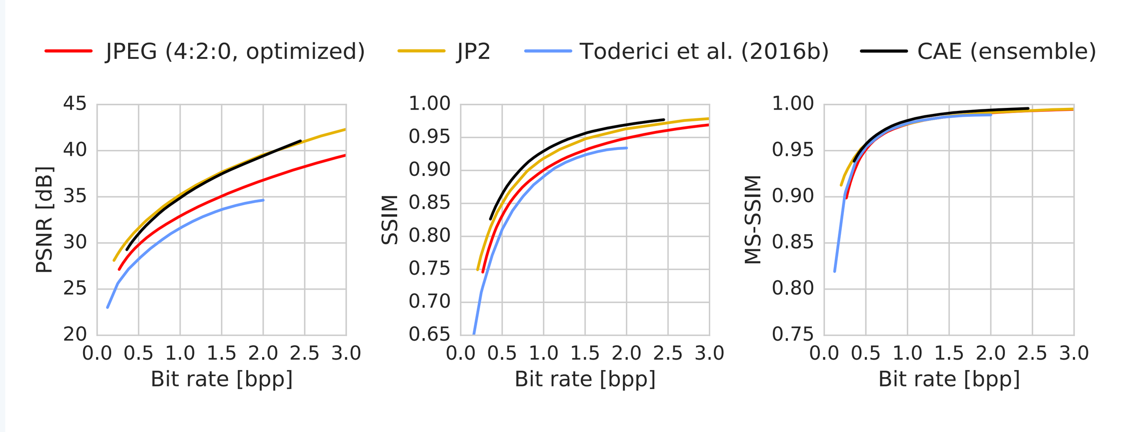
$$\log_2 q(\mathbf{z}+\mathbf{u}) = \sum_{i,j,k} \log_2 \sum_s \pi_{ks} \mathcal{N}(z_{kij}+u_{kij}; 0, \sigma_{ks}^2)$$

Qualitative results



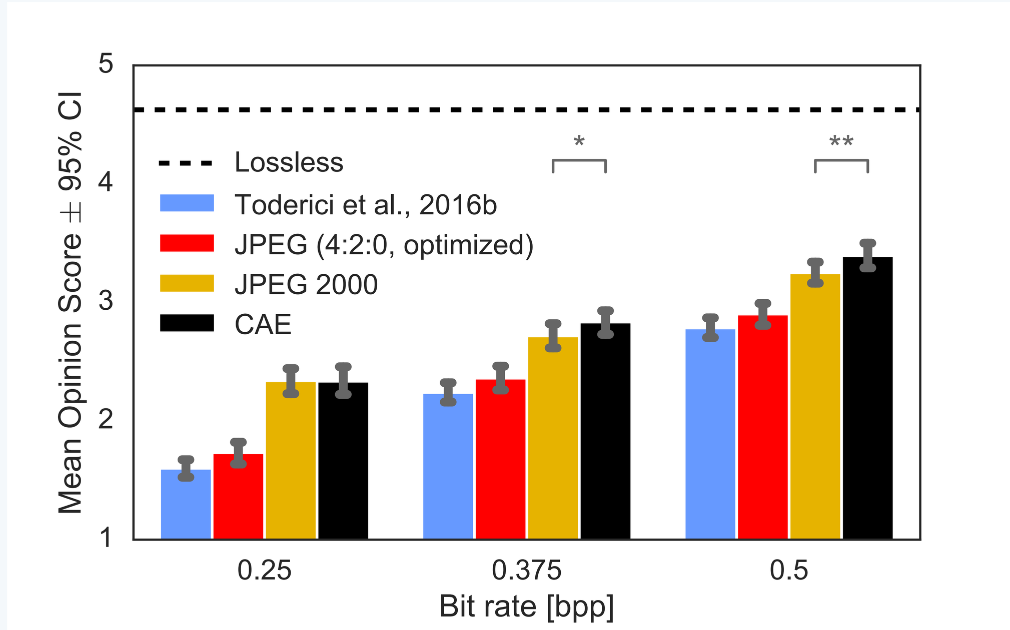
PSNR, SSIM, MS-SSIM

Results on the **Kodak dataset**:



Mean opinion scores

Following standard practices, we asked 24 naive subjects to rate images on a scale from 1 to 5:



References

[1] Shi et al., CVPR, 2016
[2] Balle et al., arXiv:1607.05006, 2016
[3] Toderici et al., arXiv:1608.05148, 2016

Resources

Compressed images and bit-rates:
http://theis.io/compressive_autoencoder/