QUALITY ASSESSMENT OF DEEP-LEARNING-BASED IMAGE COMPRESSION

Giuseppe Valenzise, Andrei Purica†, Vedad Hulusic‡, Marco Cagnazzo†,*

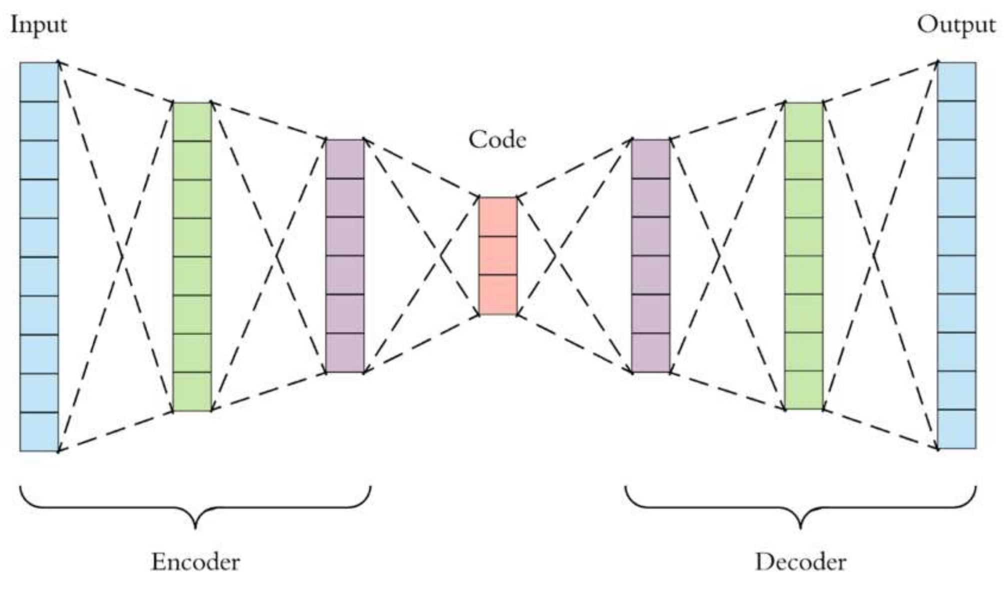
* L2S, UMR 8506, CNRS – CentraleSupélec – Université Paris-Sud
† LTCI, Télécom Paristech, Paris
‡Department of Creative Technology, Faculty of Science and Technology, Bournemouth University, UK

Abstract

Very recently, deep generative models have been used to optimize or replace some image coding, with very promising results. However, so far no systematic and independent study of the coding performance of these algorithms has been carried out. In this paper, we conduct a subjective evaluation of two recent deep-learning-based image compression algorithms, comparing them to JPEG 2000 and to the recent BPG image codec based on HEVC Intra. We found that compression approaches based on deep auto-encoders can achieve coding performance higher than JPEG 2000, and sometimes as good as BPG. We also show experimentally that the PSNR metric is to be avoided when evaluating the visual quality of deep-learning-based methods, as their artifacts have different characteristics from those of DCT or wavelet-based codecs. In particular, images compressed at low bitrate appear more natural than JPEG 2000 coded pictures, according to a no-reference naturalness measure.

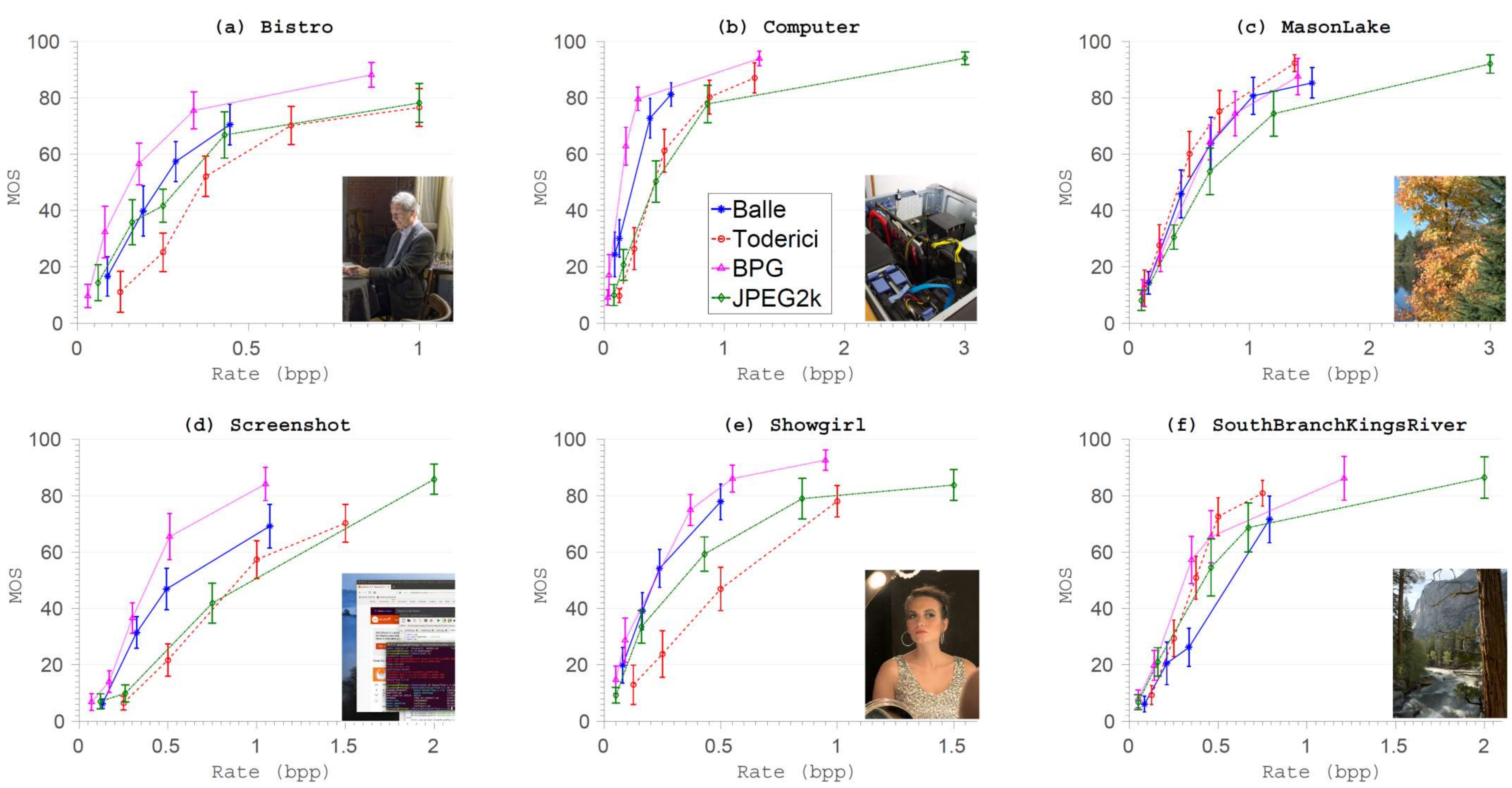
Context and motivation

- Traditional image coding methods mainly based on linear signal models (DCT, wavelet transform, linear prediction)
- Recently proposed image compression algorithms based on deep generative models such as (variational) auto-encoders or generative adversarial networks
- Goal: assess visual quality of images compressed with these methods



Subjective Experiment

- Compared methods:
  - **Ballé et al.** (2016) [1]: End-to-end image compression using a variational auto-encoder
  - **Toderici et al.** (2016) [2]: Variable bit rate image compression using recurrent auto-encoders
  - **JPEG 2000**
  - **BPG** (HEVC Intra)
- For Ballé et al., bitrate estimated by implementing a simple run-length + entropy encoding
- Test material: 6 uncompressed images of 736x960 pixels, total of 113 compressed stimuli
- 23 participants, no outliers detected
- Method: DSIS with continuous scale

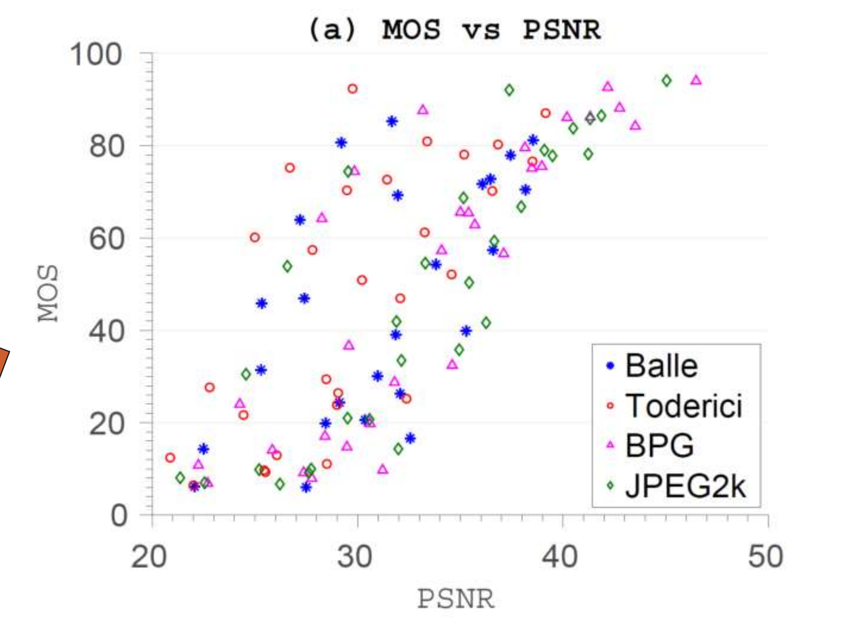


Objective Evaluation

Fidelity metrics

Metric	JPEG2K & BPG PCC	JPEG2K & BPG SROCC	JPEG2K & BPG RMSE	Balle & Toderici PCC	Balle & Toderici SROCC	Balle & Toderici RMSE	All methods PCC	All methods SROCC	All methods RMSE
PSNR	0.858	0.870	15.676	0.652	0.638	20.156	0.763	0.766	18.619
SSIM	0.902	0.908	13.154	0.829	0.830	14.864	0.866	0.871	14.392
MS-SSIM	**0.964**	**0.957**	**8.170**	0.917	0.907	10.6	0.941	**0.936**	9.77604
VSNR	0.888	0.896	14.011	0.740	0.731	17.881	0.815	0.815	16.677
VIF	0.962	0.953	8.348	**0.931**	**0.919**	**9.740**	**0.944**	**0.936**	**9.516**
UQI	0.812	0.802	17.813	0.815	0.821	15.423	0.813	0.807	16.751
IFC	0.925	0.917	11.615	0.922	0.907	10.282	0.922	0.910	11.163
NQM	0.897	0.899	13.499	0.803	0.794	15.842	0.852	0.848	15.078
WSNR	0.953	0.955	9.261	0.866	0.851	13.318	0.910	0.908	11.914

- For some metrics (especially PSNR), drop in accuracy with DL methods



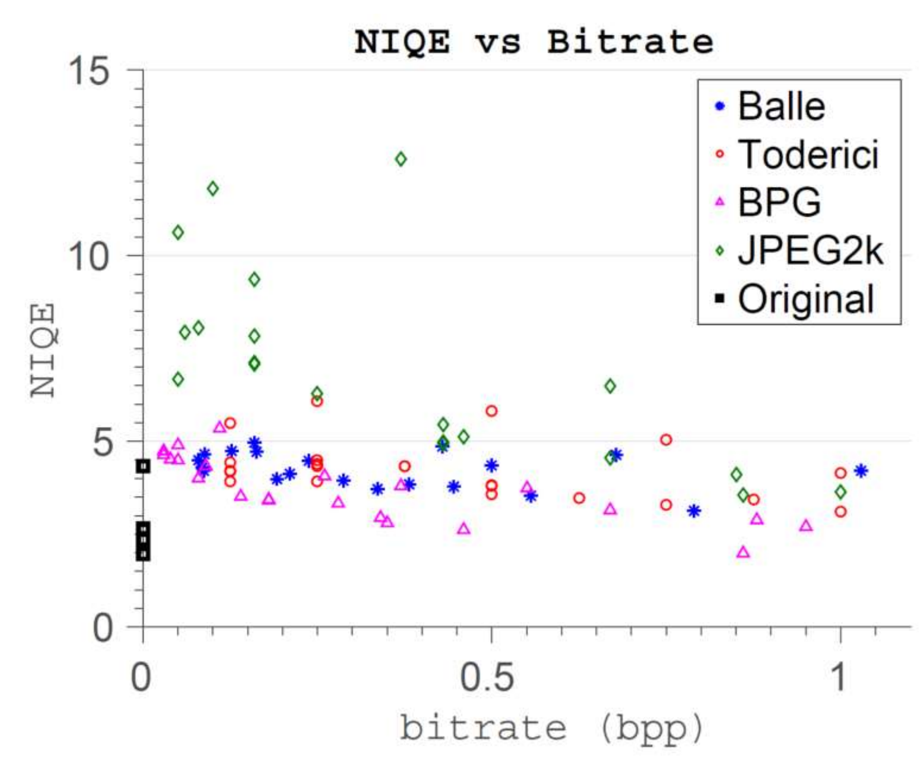
Qualitative results

MOS = 12.4
PSNR = 20.85 dB

MOS = 8.1
PSNR = 21.35 dB

(a) *(left) Ballé, 0.38 bpp; (right) JPEG 2K, 0.43 bpp* (b) *(left) Ballé, 0.09 bpp; (right) BPG, 0.08 bpp* (c)*(left)Toderici,0.125 bpp; (right)JPEG 2K, 0.1 bpp*

Naturalness



[1] J. Ballé, V. Laparra, and E. P. Simoncelli, “End-to-end optimized image compression,” in Int. Conf. on Learning Representations (ICLR), Toulon, France, Apr. 2017.
[2] G. Toderici, D. Vincent, N. Johnston, S. J. Hwang, D. Minnen, J. Shor, and M. Covell, “Full resolution image compression with recurrent neural networks,” in IEEE CVPR, Hawaii, USA, Jul. 2017, pp. 5435–5443.

MMSP 2018, Vancouver, Canada