A. Learning Better Lossless Compression Using Lossy Compression – Supplementary

A.1. Q-Classifier Architecture

We show the architecture for the Q-Classifier in Table A1. *Residual* denotes a sequence of convolution, ReLU, convolution, with a skip connection adding the input to the output (as in [15], but without BatchNorm).

Layer	C_{in}	C_{out}	Filter	Stride
Conv + ReLU	3	64	5×5	2
Conv + ReLU	64	128	5×5	2
4× Residual	128	128	3×3	
Conv	128	256	5×5	2
4× Residual	256	256	3×3	
Channel-Avg.	256	256		
Linear	256	$\lvert\mathcal{Q}\rvert$		

Table A1. Q-Classifier architecture.

A.2. BPG Performance

Fig. A1 compares the performance of BPG on Kodak, in terms of PSNR, to the recent learned image compression approach from Minnen *et al*. [30]. The plot is digitized from Figure 2 in [30].

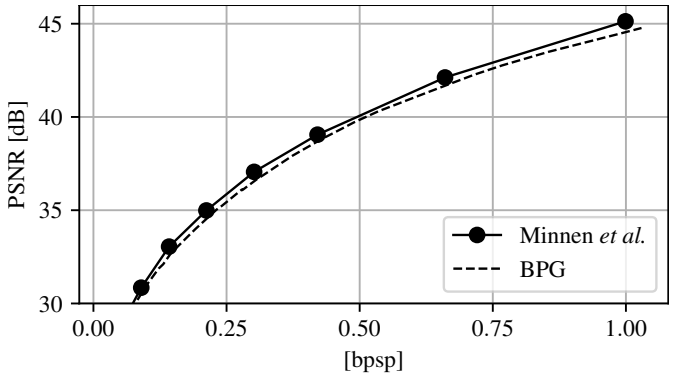

Figure A1. Comparing BPG to Minnen *et al*. [30]

A.3. Examples from the testing sets

We provide additional visual examples here:
https://data.vision.ee.ethz.ch/mentzerf/rc/rc_suppl_additional.pdf
Specifically, we show one image from each of our testing sets, alongside with the residual r and a sample from $p(r)$, which is expected to be visually similar to r. Please refer to Section 6.4 for details on sampling and the visualization.