

# Supplementary material: Slimmable Compressive Autoencoders for Practical Neural Image Compression

Fei Yang<sup>1,2</sup>, Luis Herranz<sup>2</sup>, Yongmei Cheng<sup>1</sup>, Mikhail G. Mozerov<sup>2</sup>

<sup>1</sup> School of Automation, Northwestern Polytechnical University, Xi'an, China

<sup>2</sup> Computer Vision Center, Universitat Autònoma de Barcelona, Barcelona, Spain

{fyang, lherranz, mozerov}@cvc.uab.es, chengym@nwpu.edu.cn

## A. Example of $\lambda$ -scheduling

Fig. A1 (top) illustrates the evolution of the RD curve from the initial naive training (in black) until the end of  $\lambda$ -scheduling (in red). Fig. A1 (bottom) illustrates the schedule for the  $\lambda$ s of the different subCAEs.

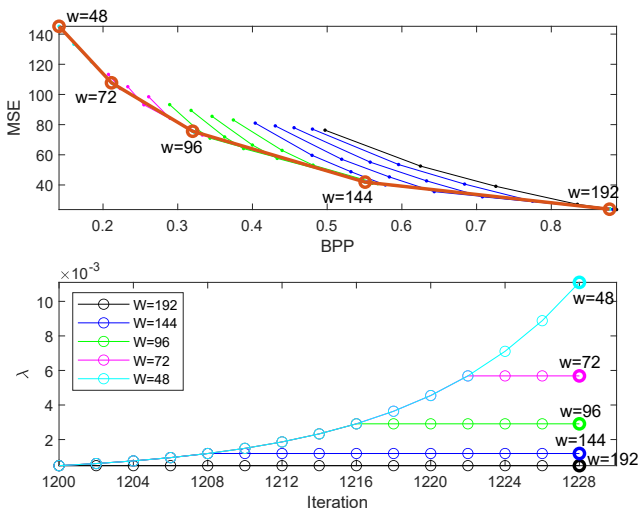


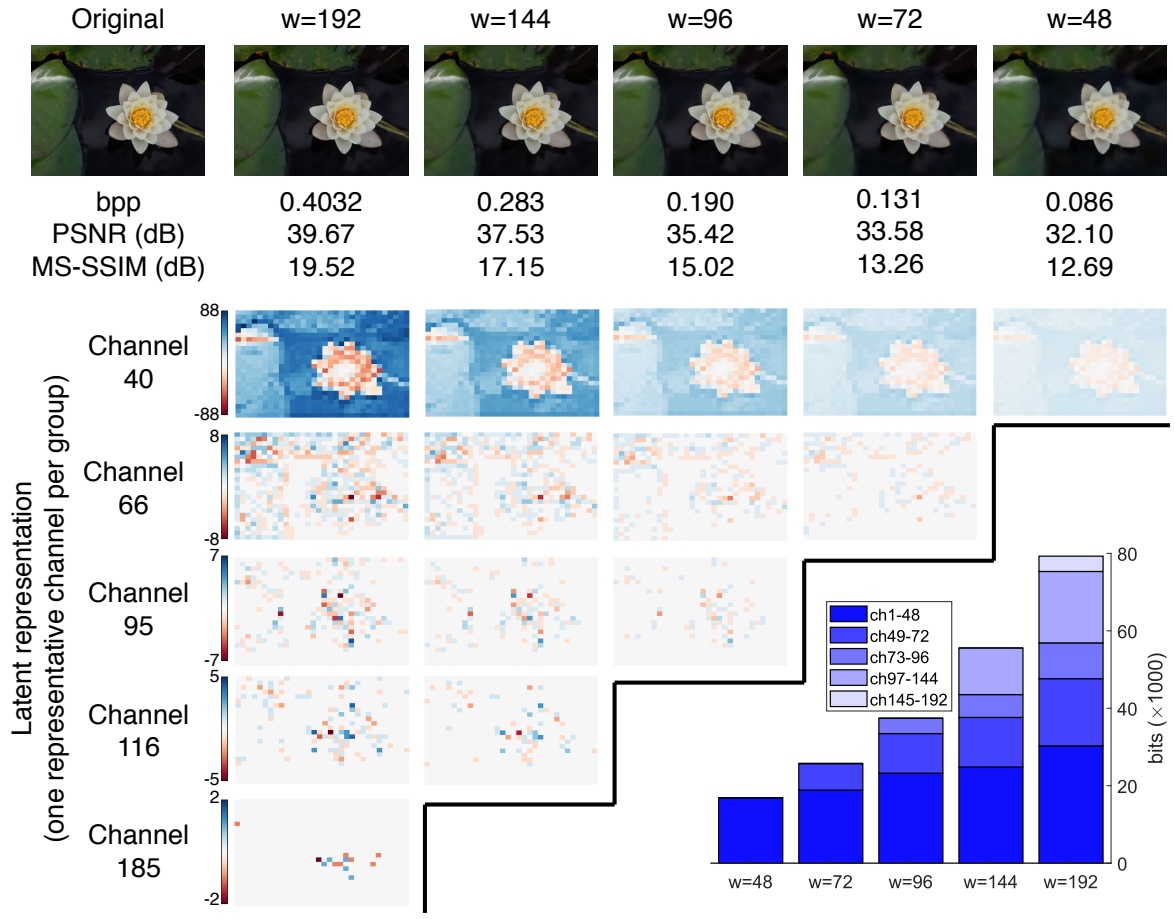
Figure A1: Evolution of the RD curve (top) and  $\lambda$  during the  $\lambda$ -scheduling phase. Naive training shown in black (top).

## B. Additional visualizations

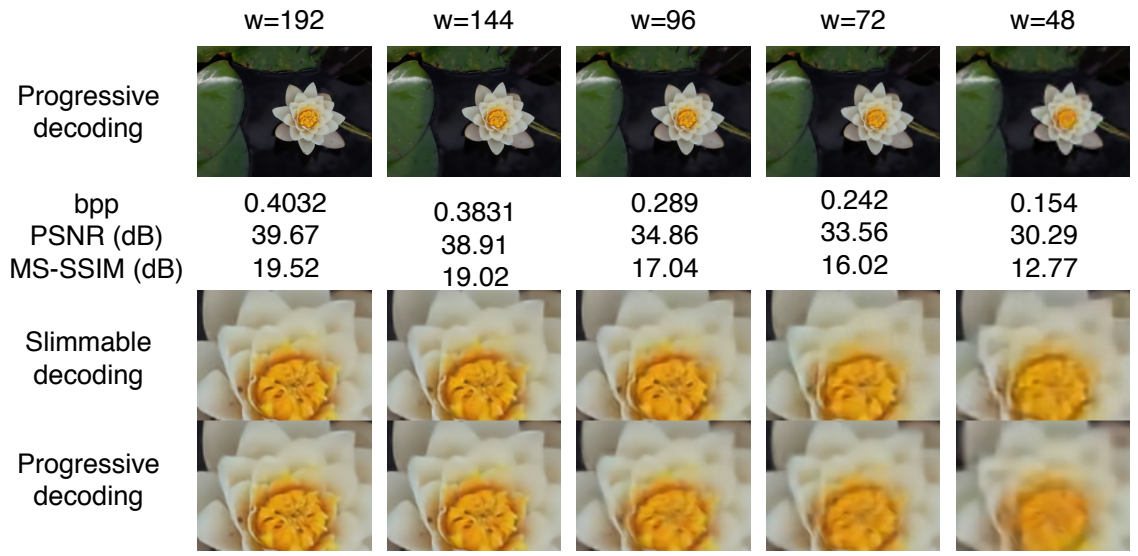
Fig. A2a shows an illustrative example of the reconstructed images and (quantized) latent representations obtained by the different subCAEs. The slimmable structure of the last layer of the encoder results in a latent representation also structured in five groups of channels (i.e. 1-48, 49-72, 73-96, 97-144 and 145-192). One representative channel per group is shown. The corresponding bit allocation is shown in Fig. A2b. Note that smaller subCAEs also allocate fewer bits in the first group of channels.

Fig. A2c illustrates progressive decoding when (quality)

scalability is enabled, compared with the default slimmable decoding (non scalable bitstreams). In this case each group of channels of  $w = 192$  in Fig. A2b correspond to the base stream (channels 1-48) and four enhancement streams, which can be progressively combined to improve quality. However, there is a noticeable increase in both rate (explained by the bit allocation shown in Fig. A2b) and distortion with the quality scalable bitstream.



(a) (b)



(c)

Figure A2: Illustrative example: (a, top) reconstructions using slimmable decoding, (a, bottom) selection of quantized latent maps (one per group of channels), (b) break down of bits spent in each group of channels, (c) scalable bitstream.