The Devil Is in the Details: Window-based Attention for Image Compression Supplementary Material

The supplementary document provides more experimental details of proposed methods. Firstly, we show the visual quality comparison of our method and classical methods in Figure 1. Secondly, in Figure 2, we provide some examples of the proposed method with MSE and MS-SSIM metrics, respectively. Finally, we show more rate-distortion results and compare with other methods in Figure 3 and 4.

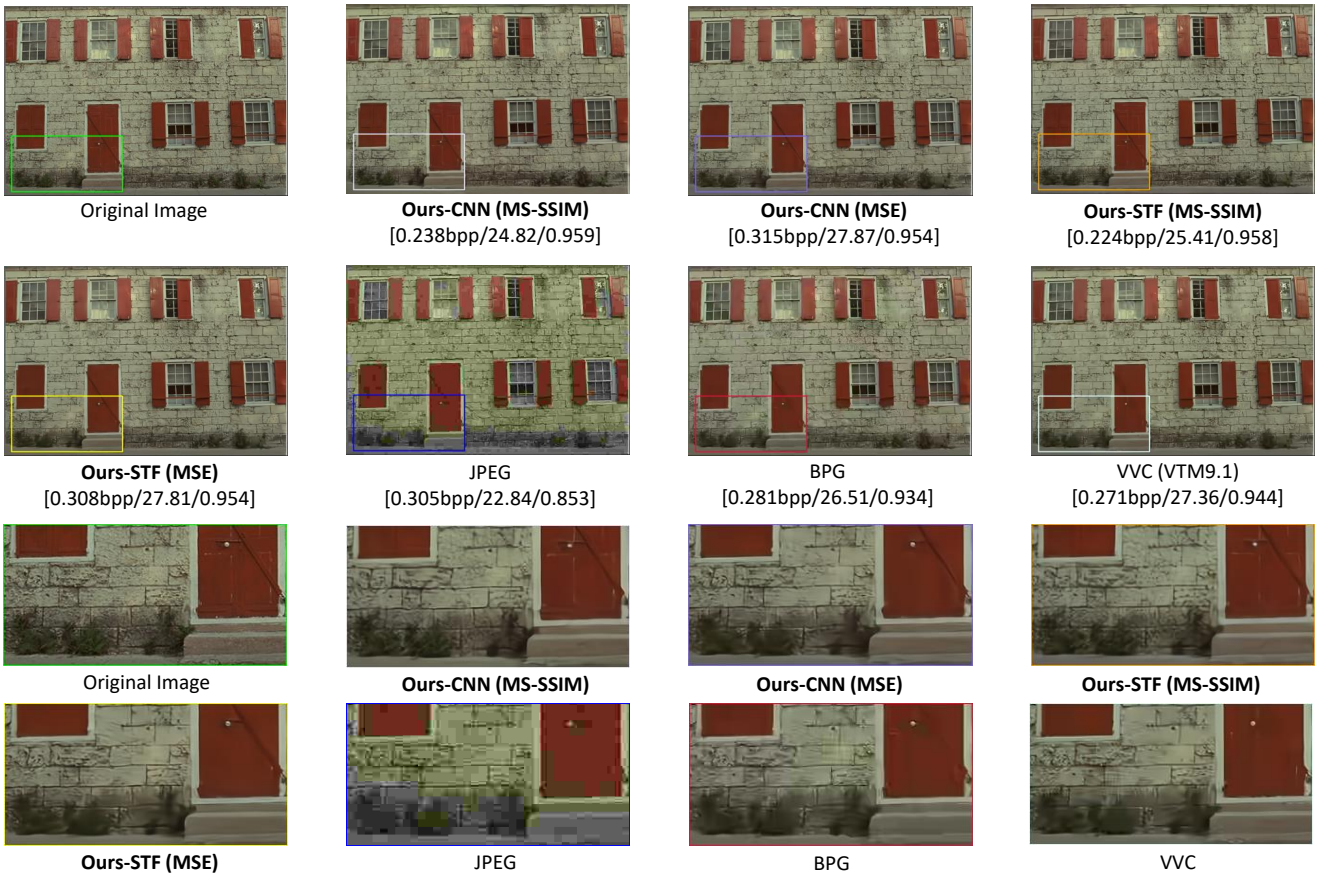

Figure 1. At similar bit rates, our methods provide the highest visual quality on *Kodim01.png*.

Ours-CNN (MSE)

Ours-CNN (MS-SSIM)

Ours-STF (MSE)

Ours-STF (MS-SSIM)

Ours-CNN (MSE)

Ours-CNN (MS-SSIM)

Ours-STF (MSE)

Ours-STF (MS-SSIM)

Figure 2. Reconstructed images *Kodim20.png* and *Kodim04.png* at similar bit rates. The models trained for MS-SSIM assign more detail to low contrast regions (such as grass and hair), while the models trained for MSE assign more detail to high contrast regions (such as text). Note that our CNN-based model provides higher visual quality on low contrast regions. Because in our STF model, the space redundancy within patches may lead to the loss of detail.

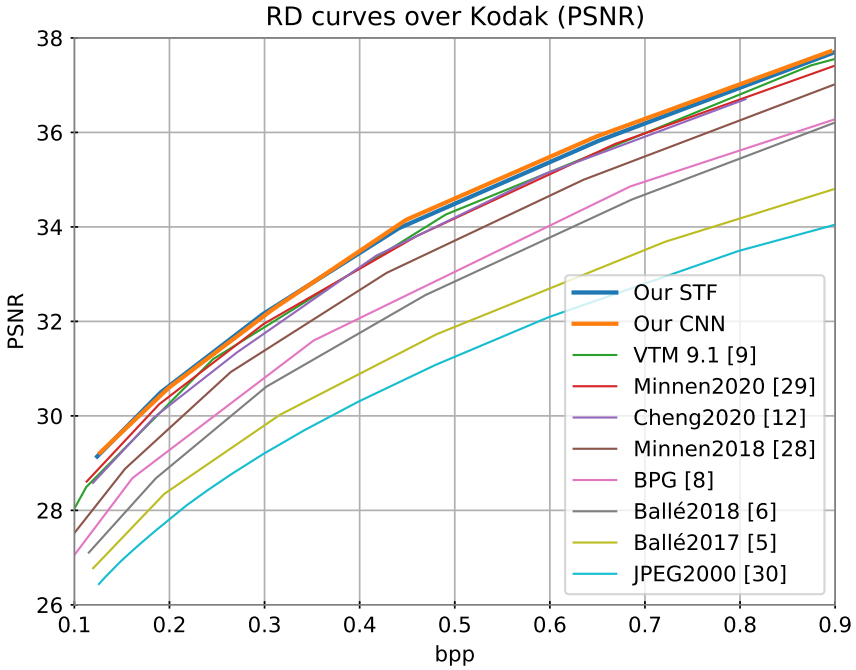

Figure 3. Rate-distortion curves for PSNR covering a wide range of conventional and ANN-based compression methods on Kodak dataset. All ANN-based compression methods are optimized for MSE. The RD points are obtained from their paper or CompressAI platform.

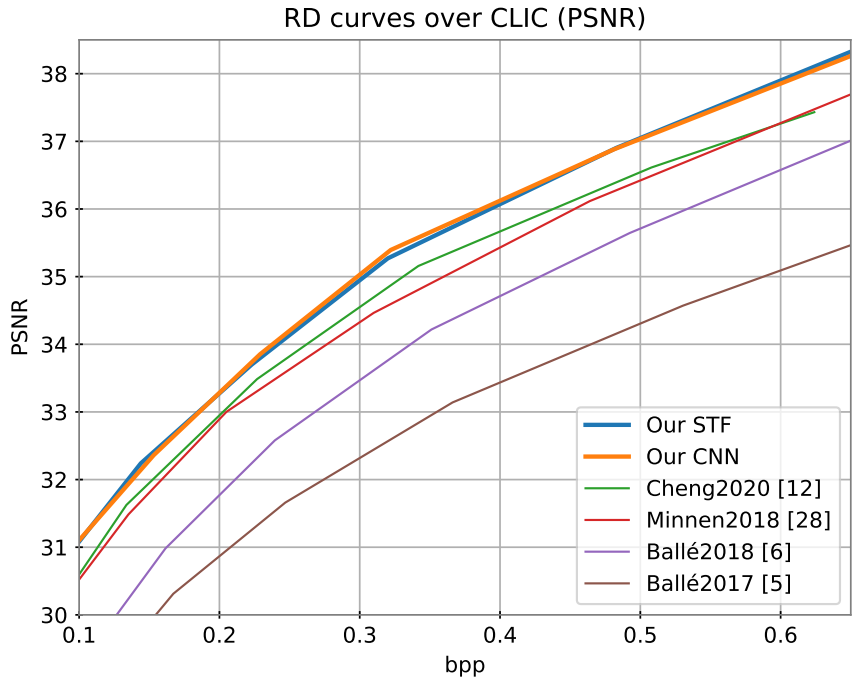

Figure 4. Rate-distortion curves for PSNR covering a wide range of ANN-based compression methods on CLIC professional validation dataset. All ANN-based compression methods are optimized for MSE. The RD points of Minnen2018, Ballé2018, and Ballé2017 are evaluated on CompressAI platform.