A Additional Explanation for the Methods

A.1 Derivation of Eq. 10, 11, 12

Eq. 10: $\forall F \in \mathbb{R}^{b\times D}$, $X(Y;K) = YK^+K + F(I - KK^+)$ is a solution to $Y = XK$.

Proof. From the property of Moore-Penrose pseudo-inverse [Moo] we know that

$$KK^+K = K. \tag{1}$$

Additionally, for a orthonormal K (with linearly independent columns), we have

$$K^+K = I. \tag{2}$$

Then, right-multiply $X(Y;K)$ by K, we get

$$\begin{aligned} X(Y;K)K &= YK^+K + F(I - KK^+)K \\ &= Y + FK - FK = Y. \end{aligned} \tag{3}$$

□

Eq. 11 can be derived similarly.
Eq. 12 can be derived similarly knowing that $K^+ = VS^{-1}U^T$.

B More Experimental Results

B.1 More Experiment Setup

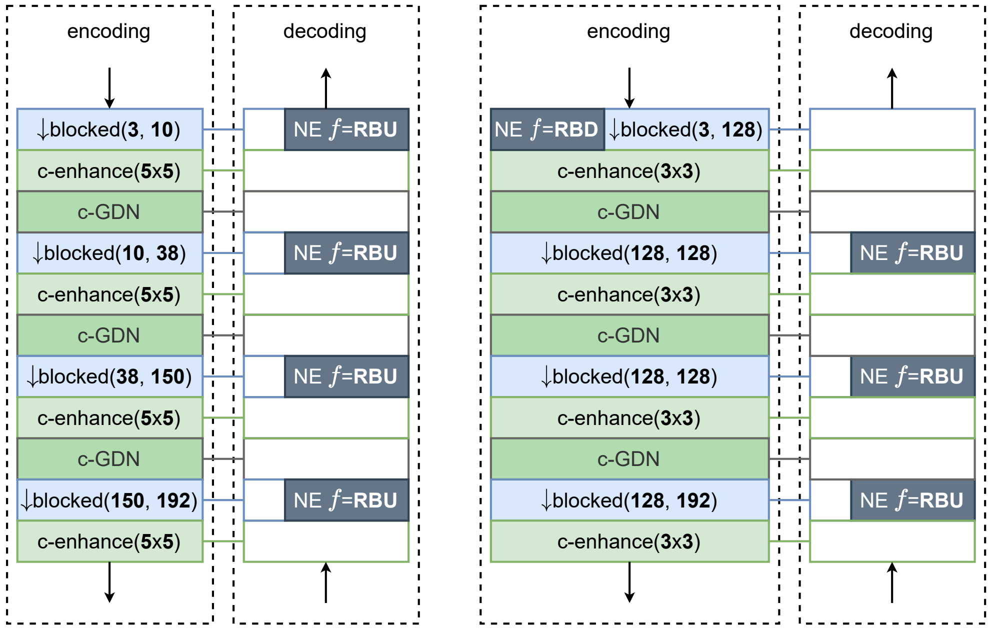
The detailed encoding and decoding transform is illustrated in Fig. 1. To extend the idempotent framework to near-idempotent framework, we change the first blocked convolution in the encoding transform to non-surjective by increasing its output channel number from 10 to 128. Since this blocked convolution is no longer surjective, it is no longer right-invertible. However, its corresponding layer in the decoding transform can be make surjective and right-invertible. Thus, we make its corresponding layer in the decoding transform surjective, and use the null-space enhancement on the encoding side.

Coupling structure [Dinh et al., 2016] used in coupling enhancement (c-enhance) and coupling GDN (c-GDN) is illustrated in Fig. 2(a). For c-enhance, the scale $s(\cdot)$ and translation $t(\cdot)$ are convolution. For c-GDN, the scale $s(\cdot)$ and translation $t(\cdot)$ are GDN [Ballé et al., 2015]. Following the usage guideline in [Dinh et al., 2016], we concatenate two coupling structure with the opposite way of splitting in one c-enhance/c-GDN.

In the ablation study, we test what will happen if all the blocked convolution are changed from right-invertible to invertible. To do this, we increase the output dimension of blocked convolution to its input dimension. Specifically, the output channel number is set to 4 times the input channel number, since the blocked convolution has a stride of 2. The changed setting for the blocked convolutions are $(3, 12)$, $(12, 48)$, $(48, 192)$, $(192, 768)$, respectively.

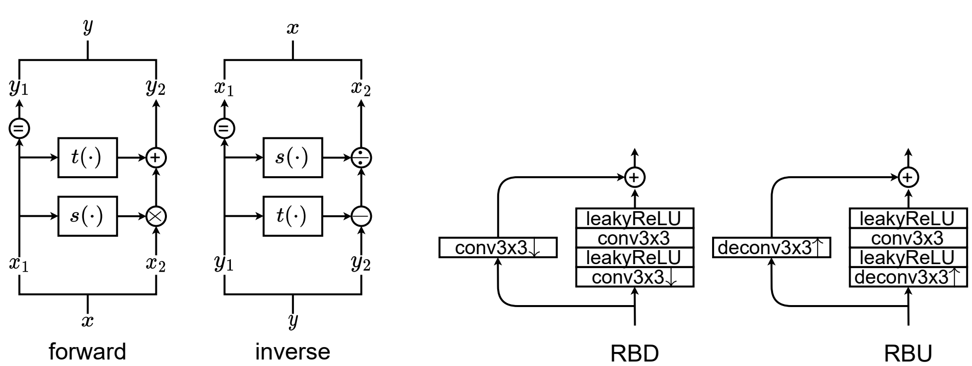
We include the implement of other codecs as follows. For codecs that have open-source implementations, we use that implementation. For codecs that do not have open-source implementations, we either use the data provided in the paper, or re-implement by ourselves if the detailed architecture is provided.

- Implementations from CompressAI [Bégaint et al., 2020]: Balle2017[Ballé et al., 2017], Balle2018[Ballé et al., 2018], Cheng2020[Cheng et al., 2020], JPEG2000Taubman et al. [2002], BPG444Sullivan et al. [2012], VTM444[Bross et al., 2021]
- Data from the original papers: Helminger2021[Helminger et al., 2021], Cai2022[Cai et al., 2022]
- Our re-implementation: Kim2020[Kim et al., 2020]. Specifically, we re-implement the FI loss proposed in this work on Balle2018 [Ballé et al., 2018].



(a) Details of the proposed imdempotent framework (b) Details of the proposed near-imdempotent framework

Figure 1: Detailed encoding and decoding transform for the proposed idempotent and near-idempotent framework. **blocked** refers to the proposed block-rearranged convolution. The (input, output) channels are annotated in the brackets, and stride is annotated using down-arrow. **NE** refers to the proposed null-space enhancement, and f is learned parametric function. Here RBU/RBD are the residual block used for upsampling/downsampling (Fig. 2(b)), respectively. **c-enhance** refers to the proposed coupling enhancement using coupling structure (Fig. 2(a)). The kernel size of the convolution is annotated in the brackets. **c-GDN** refers to the proposed right-invertible normalization using coupling structure (Fig. 2(a)). Blanked rectangular refer to the right-inverse/inverse of the corresponding layer, and has no additional weights.



(a) Couping structure used in c-enhence and c-GDN (b) Residual block used in null-space enhancement

Figure 2: Submodules used in the proposed framework: (a) coupling structure used in c-enhance and c-GDN; (b) residual block downsampling (RBD) and residual block upsampling (RBU) used in the $f(\cdot)$ of null-space enhancement.

B.2 More Quantitative & Qualitative Results

See Fig. 3-7 for more quantitative results.
See Fig. 8-11 for more qualitative results.

C More Discussion

C.1 Limitation

In this work, the surjective encoding transform is constructed using function composition of simple surjections. This construction strategy limits the latent dimension to be non-increasing throughout the encoding transform. This limitation contradicts the mainstream design logic of neural network, and is harmful to expressiveness.

A function composition of surjections is always a surjection, but a surjection needs not to be a function composition of surjections [Mac Lane, 2013]. Thus this restriction could be lifted by more advanced construction strategy of surjection.

C.2 Broader Impact

Improve the rate-distortion of re-compression has positive social impact. Re-compression constantly happens in the transmission and redistribution of image data. Reducing the bitrate can save the resources, energy and the carbon emission during these processes.

C.3 Reproducibility Statement

All theoretical results are proven in Appendix. A. For experimental results, all the datasets used are publicly available, and the implementation details are provided in Appendix. B. Furthermore, the source code for reproducing experimental results are provided in supplementary materials.

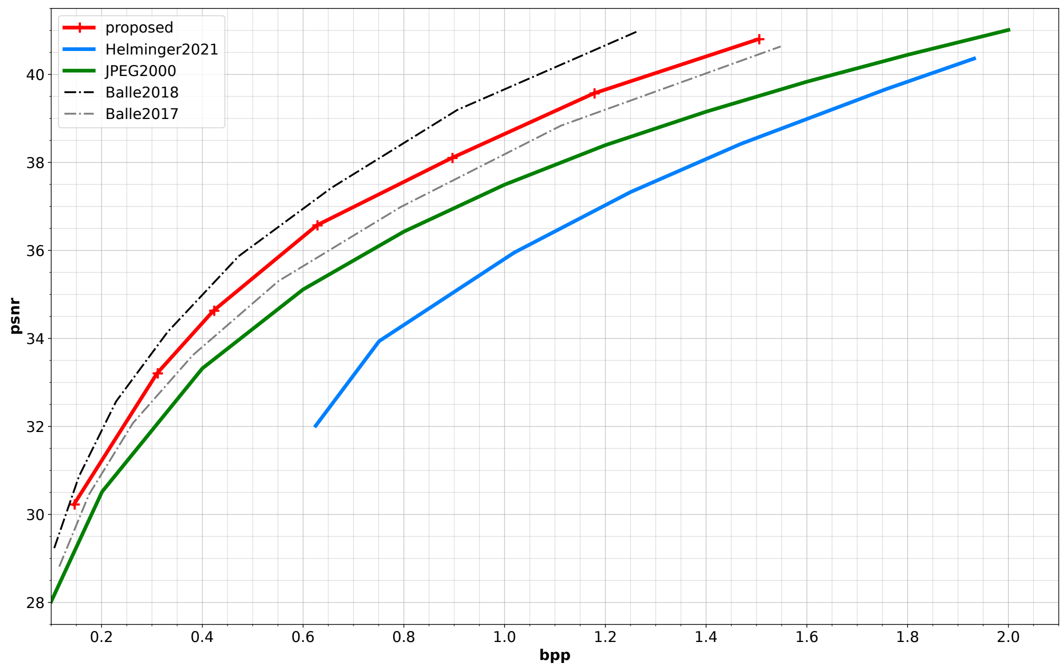


Figure 3: **PSNR-BPP** curve on **Tecnick**. Note that `Balle17` [Ballé et al., 2017] and `Balle18` [Ballé et al., 2018] are NOT idempotent and plotted here only for reference.

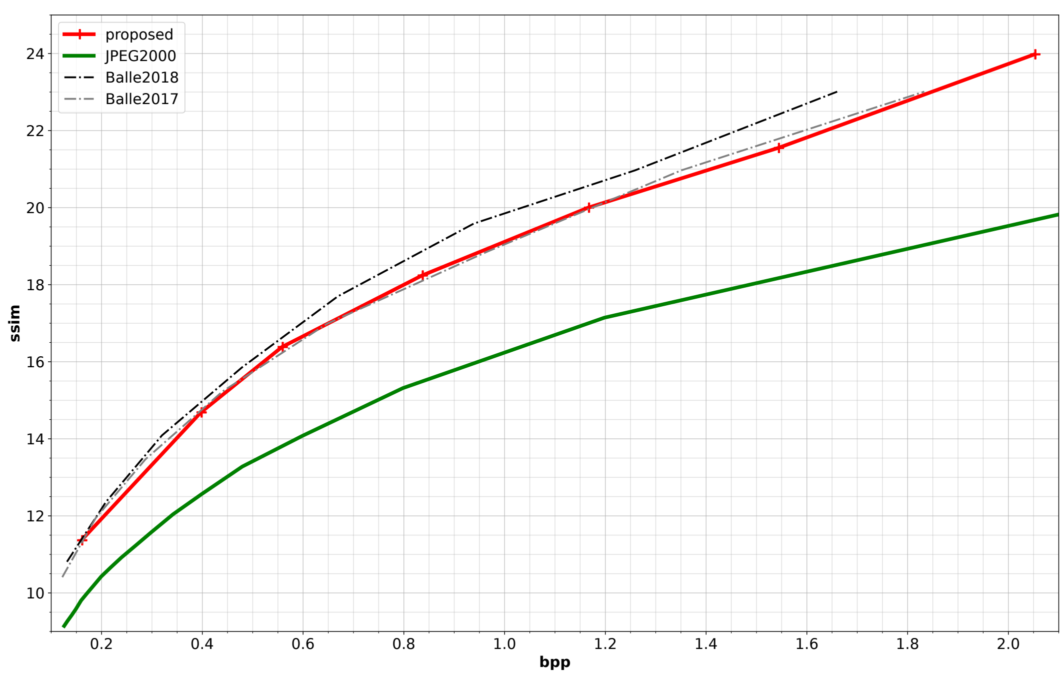


Figure 4: **MSSSIM-BPP** curve on **Kodak**. All models are optimized for minimizing MSE. Note that `Balle17` [Ballé et al., 2017] and `Balle18` [Ballé et al., 2018] are NOT idempotent and plotted here only for reference.

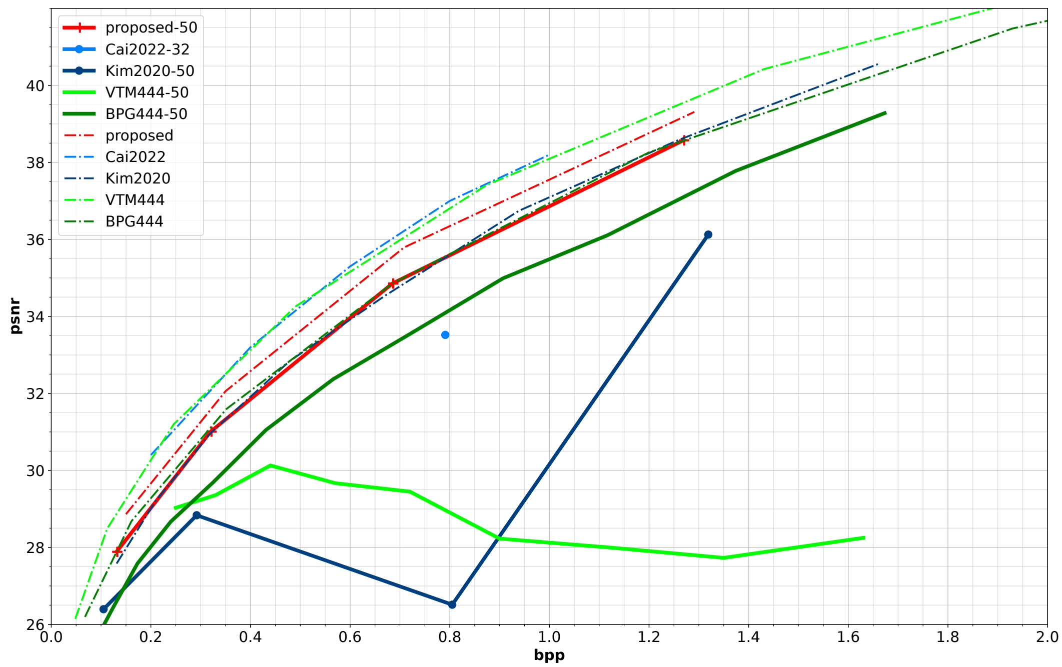


Figure 5: **PSNR-BPP** curve on **Kodak**. First-time compression performance is plotted in dotted line, and re-compression performance (upto 50 times) is plotted in solid line.

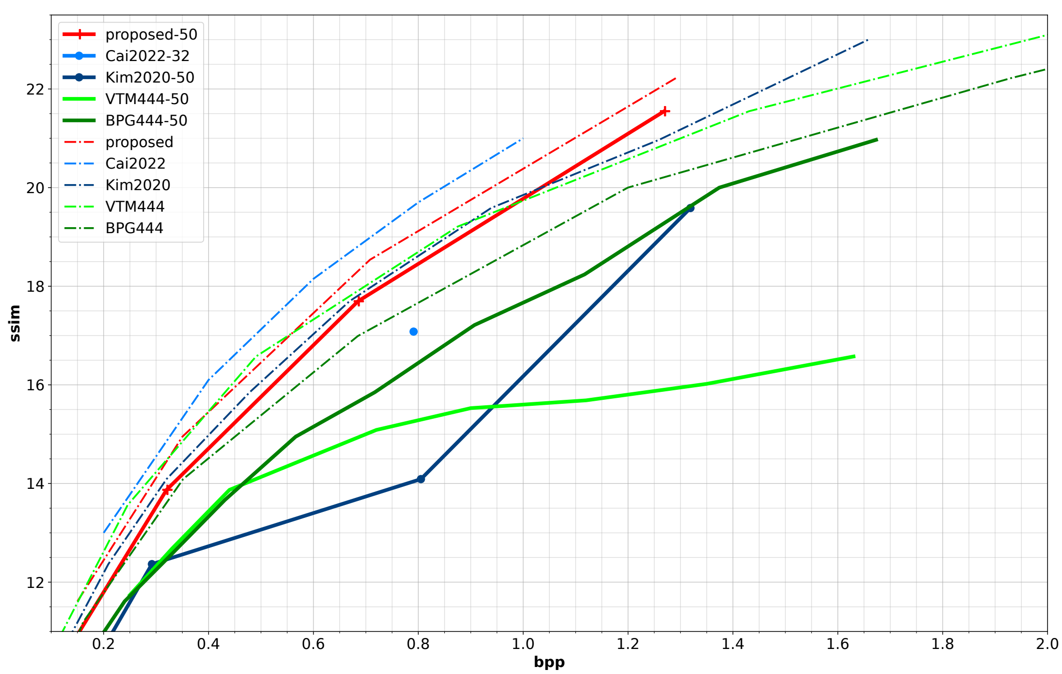


Figure 6: **MSSSIM-BPP** curve on **Kodak**. All models are optimized for minimizing MSE. First-time compression performance is plotted in dotted line, and re-compression performance (upto 50 times) is plotted in solid line.

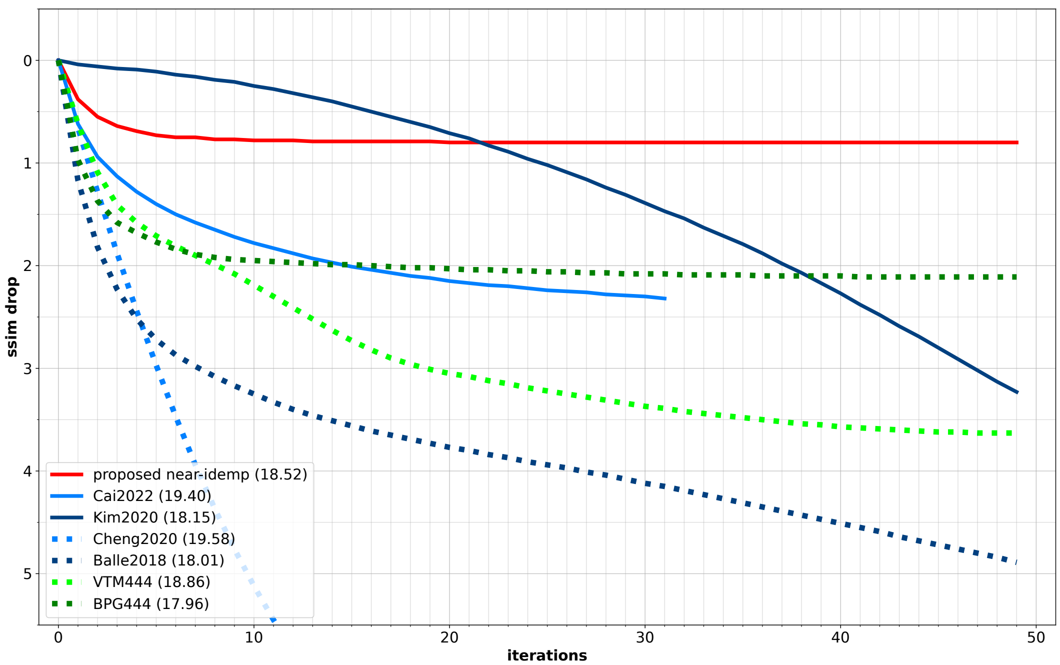


Figure 7: **MS-SSIM** drop upto 50 re-compression of near-idempotent codecs on **Kodak**. First-time MS-SSIM is annotated in $(\cdot)$. All models are optimized for minimizing MSE.

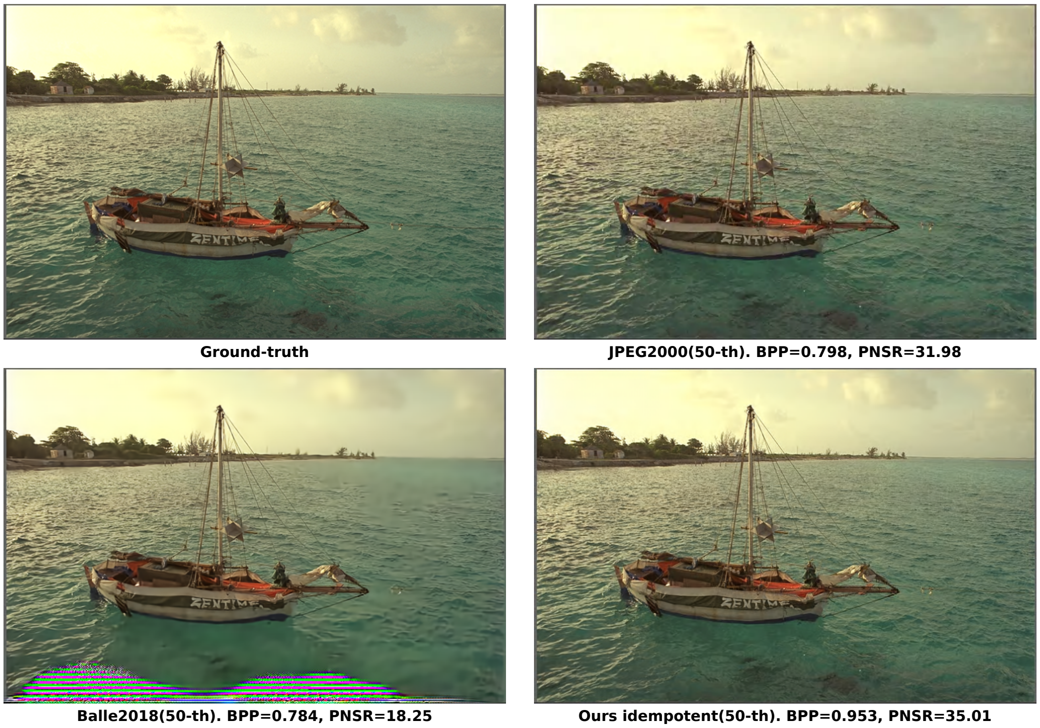


Figure 8: Qualitative comparison on reconstructed `kodim06` image after 50 times re-compression.

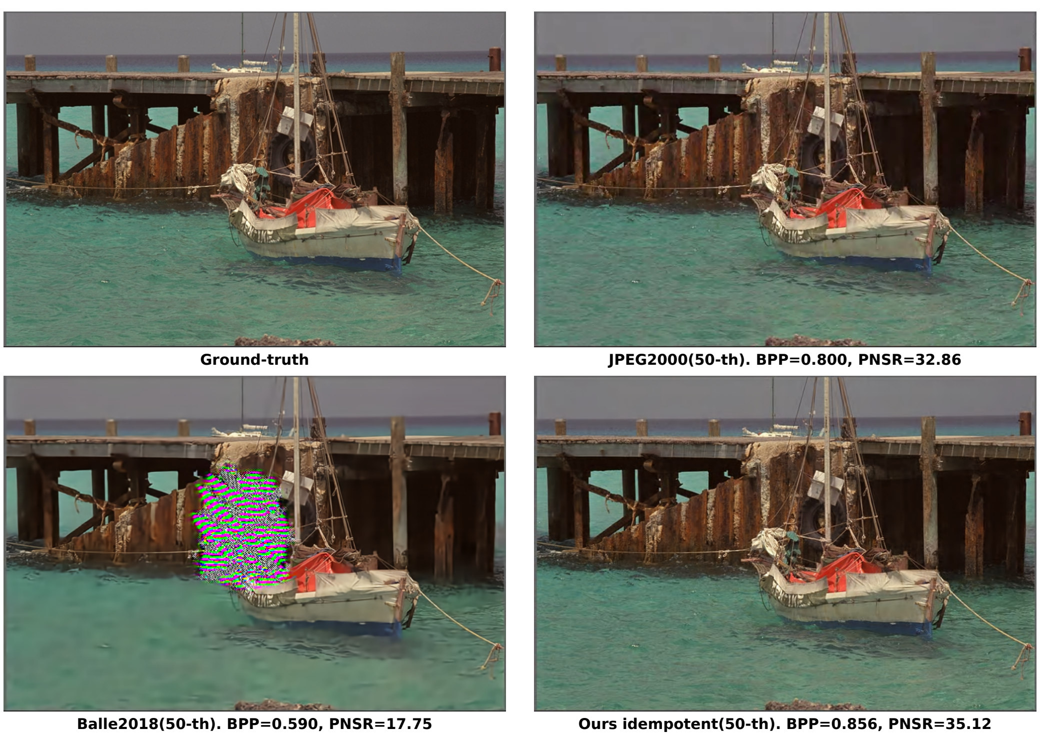


Figure 9: Qualitative comparison on reconstructed `kodim11` image after 50 times re-compression.

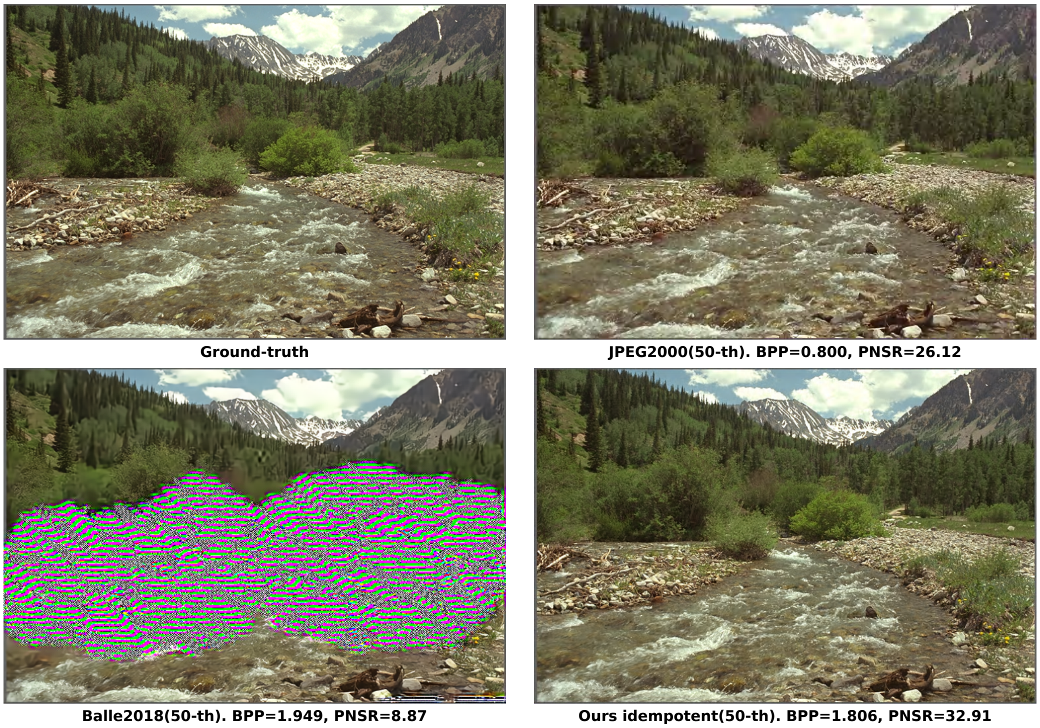


Figure 10: Qualitative comparison on reconstructed `kodim13` image after 50 times re-compression.

Figure 11: Qualitative comparison on reconstructed `kodim24` image after 50 times re-compression.

References

Moore-penrose inverse. URL `https://en.wikipedia.org/wiki/Moore-Penrose_inverse`.

J. Ballé, V. Laparra, and E. P. Simoncelli. Density modeling of images using a generalized normalization transformation. *arXiv preprint arXiv:1511.06281*, 2015.

J. Ballé, V. Laparra, and E. P. Simoncelli. End-to-end optimized image compression. In *5th International Conference on Learning Representations, ICLR 2017*, 2017.

J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston. Variational image compression with a scale hyperprior. In *International Conference on Learning Representations*, 2018.

J. Bégaint, F. Racapé, S. Feltman, and A. Pushparaja. Compressai: a pytorch library and evaluation platform for end-to-end compression research. *arXiv preprint arXiv:2011.03029*, 2020.

B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J.-R. Ohm. Overview of the versatile video coding (vvc) standard and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(10):3736–3764, 2021.

S. Cai, Z. Zhang, L. Chen, L. Yan, S. Zhong, and X. Zou. High-fidelity variable-rate image compression via invertible activation transformation. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 2021–2031, 2022.

Z. Cheng, H. Sun, M. Takeuchi, and J. Katto. Learned image compression with discretized gaussian mixture likelihoods and attention modules. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7939–7948, 2020.

L. Dinh, J. Sohl-Dickstein, and S. Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

L. Helminger, A. Djelouah, M. Gross, and C. Schroers. Lossy image compression with normalizing flows. In *Neural Compression: From Information Theory to Applications–Workshop@ ICLR 2021*, 2021.

J.-H. Kim, S. Jang, J.-H. Choi, and J.-S. Lee. Instability of successive deep image compression. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 247–255, 2020.

S. Mac Lane. *Categories for the working mathematician*, volume 5. Springer Science & Business Media, 2013.

G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand. Overview of the high efficiency video coding (hevc) standard. *IEEE Transactions on circuits and systems for video technology*, 22(12): 1649–1668, 2012.

D. S. Taubman, M. W. Marcellin, and M. Rabbani. Jpeg2000: Image compression fundamentals, standards and practice. *Journal of Electronic Imaging*, 11(2):286–287, 2002.