Optimizing Image Compression: Perspectives on SVD-Based Restoration

Jiancong Chen[1], Peilin Chen[1], Shiqi Wang[1,2], and Zhu Li[3]
[1]Department of Computer Science, City University of Hong Kong
[2]Shenzhen Research Institute, City University of Hong Kong
[3]Department of Computer Science & Electrical Engineering (CSEE), University of Missouri
Email: Cc.chen@my.cityu.edu.hk, {plchen3, shiqwang}@cityu.edu.hk, lizhu@umkc.edu

***Abstract*—Recently, neural image codecs have demonstrated impressive performance in image compression. Current approaches primarily focus on designing sophisticated context mechanisms or network architectures for eliminating redundancies within image data, which leads to more and more computational resource requirements and increases the encoding and decoding time. In this paper, we analyze the degradation of compressed images from the perspective of Singular Value Decomposition (SVD) and propose a novel rate-distortion (R-D) optimization framework based on compressed image restoration. Specifically, our method introduces an SVD-based basis reconstruction error into the conventional R-D loss function, enabling enhanced detail restoration capabilities. The proposed optimization framework can be seamlessly applied to various codecs optimized by the R-D loss without introducing additional learnable parameters or inference overhead. Extensive experiments on public datasets show that our method achieves performance gains across several neural image codecs compared to the baseline, validating the effectiveness of the proposed optimization framework.**

***Index Terms*—Image compression, neural image codecs, rate-distortion optimization, singular value decomposition, image restoration**

I. Introduction

Image compression is a fundamental task in the field of image processing, with applications ranging from efficient image storage to transmission over bandwidth-constrained networks. The primary goal of image compression is to eliminate the redundancies in image data while retaining the essential information needed to reconstruct the images as closely as possible to the original. Achieving high compression ratios without significant quality degradation has been a long-standing challenge in this field. Over the past few decades, traditional image compression standards such as JPEG [1], JPEG2000 [2], HEVC [3], and VVC [4], have been widely adopted in practical applications due to their efficiency, benefiting from years of development and optimization. These codecs typically rely on hand-crafted modules, including fixed transformations, quantization, and loop filters, to reduce spatial redundancy and improve coding efficiency.

This work was supported in part by Shenzhen Science and Technology Program under Project JCYJ20220530140816037, in part by ITF GHP/044/21SZ and in part by Research Grants Council GRF-11200323. *(Corresponding author: Shiqi Wang)*

However, recent advances in deep learning have facilitated the development of neural image codecs, which have achieved performance comparable to, and even surpassing the traditional codecs in terms of reconstructed quality [5], [6]. Unlike traditional codecs that rely on individually hand-crafted modules, neural image codecs jointly optimize their components and utilize learned representations through various autoencoders, resulting in more efficient compression. These learned models adapt to the complex statistical properties of image data, enabling improved compression and reconstruction capabilities. Nevertheless, the focus on designing sophisticated structures, such as transformers [7] or a mixture of transformers and CNN [8], to exploit contextual information in image data has led to increased demands for computational resources.

In this work, we aim to enhance the performance of learned image compression by focusing on the restoration of compressed images. Singular Value Decomposition (SVD) is a powerful technique widely used in image restoration tasks such as deblurring [9], denoising [10], or providing a uniform solution for multiple degradations [11], [12]. We thoroughly examine the relationships between original and compressed images from the perspective of decomposition using SVD. Our extensive analysis reveals that restoring the singular vectors of compressed images can effectively recover a substantial portion of the information loss during compression. Such a finding is also observed in other degradations [12], claiming that the singular vectors contain contextual information and details. Building on these insights, we propose an SVD-based optimization framework for image compression that maximizes the preservation of critical information without increasing the bitrate. To achieve this, we incorporate compressed image restoration into the rate-distortion (R-D) optimization process by introducing an SVD-based basis reconstruction error term into the conventional R-D loss function.

The contributions of this work can be summarized as follows:

- We thoroughly analyze the degradation of compressed images from the perspective of decomposition using SVD, revealing that the restored singular vectors can effectively recover much of the information lost during compression.

- We propose an SVD-based optimization framework that integrates image restoration into the R-D optimization process by introducing a basis reconstruction error term to maximize information preservation without increasing the bitrate.
- Extensive experiments on public datasets show that our method improves compression performance across multiple neural image codecs, requiring no additional parameters or inference overhead.

II. Proposed Method

In this section, we first examine compressed images from the perspective of SVD to gain insights into their textural and structural characteristics (Sec. II-A). Building on these insights, we then present a novel framework for optimizing the neural image codecs (Sec. II-B).

A. Exploration on Compressed Images

SVD is a widely used technique for image restoration. To analyze the degradation of compressed images, we apply the SVD to both the compressed low-quality (LQ) data and their corresponding original high-quality (HQ) counterparts in both pixel space (image) and latent space (feature map). This process yields the singular vectors and values for both the LQ and HQ data, represented as follows:

$$\begin{aligned} X_{lq} &= U_{lq}\Sigma_{lq}V_{lq}^{\top}, \\ X_{hq} &= U_{hq}\Sigma_{hq}V_{hq}^{\top}, \end{aligned} \tag{1}$$

where X_{lq} and X_{hq} are the LQ and HQ data being factored, U_* and V_* represent the matrices of decomposed singular vectors, Σ_* represent the singular values in the form of a diagonal matrix. The data can be reconstructed through matrix multiplication in an inverse process. When reconstructing the LQ data, if we replace the left and right singular vectors with their HQ counterparts, i.e., $X_{rec} = U_{hq}\Sigma_{lq}V_{hq}^{\top}$, we can recover a substantial amount of the information lost during compression, thus improving the quality of the recomposed results. We refer to this process as “recomposition”, which we apply in both pixel and latent spaces of the compressed images, finding consistent improvements across various bitrates and neural image codecs.

Fig. 1 illustrates the quantitative analysis of the recomposed images for two examples using cheng2020-anchor [5] as the neural image codec, along with the visualization results at the lowest bitrate. The curves in the rightmost column indicate that recomposition in both spaces consistently enhance the PSNR, with recomposition in pixel space yielding a substantial improvement over the LQ image. The visualization results further demonstrate that pixel space recomposition restores significantly more details, while performance in latent space is constrained by the learned autoencoder of the neural image codecs. Therefore, considering both performance and the high-dimensionality of latent space, recomposition in pixel space emerges as the more effective choice.

As the reconstruction process can be expressed as a linear combination of a set of SVD basis components,

$$X = U\Sigma V^{\top} = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\top}, \tag{2}$$

where k is the rank of X; σ_i, $\mathbf{u}_i$, and $\mathbf{v}_i$ denote the i-th singular value, left singular vector, and right singular vector, respectively. In our recomposition process, we utilize the HQ singular vectors, which serve as high-quality bases containing richer contextual information and details, leading to improved reconstruction results.

B. SVD-based Optimization Framework

Neural image codecs are typically optimized using the R-D loss, defined as follows:

$$\mathcal{L} = \mathcal{R} + \lambda\mathcal{D}, \tag{3}$$

where $\mathcal{R}$ represents the estimated bitrate and $\mathcal{D}$ denotes the reconstruction distortion, the latter is often measured by pixel-wise mean square error (MSE). λ is a weighting factor that controls the trade-off between rate and distortion. The findings from our exploration in Sec. II-A suggest that recovering a portion of the corresponding HQ bases could significantly enhance the reconstruction quality. Although the HQ bases are not available during inference, we can impose some constraints during training to facilitate the encoder in learning more effective bases. Consequently, instead of solely optimizing for pixel distortion in terms of MSE, we also take the basis distortion into consideration. Specifically, the distortion is comprised of a pixel error and an SVD-based basis error, which can be formulated as,

$$\mathcal{D} = ||X_{hq} - X_{lq}||_2^2 + \mu||U_{hq}V_{hq}^{\top} - U_{lq}V_{lq}^{\top}||_1, \tag{4}$$

where X_{hq} and X_{lq} here represent the original and compressed images in pixel space, μ is a weight set to balance the pixel and basis errors, with a value of 10 in our experiments. The first term captures the pixel error measured by MSE, while the second term quantifies the difference between the compressed LQ and original HQ bases, collectively forming the distortion component of the R-D loss in Eqn. (3). The proposed SVD-based optimization framework can be applied to any neural image codecs optimized by the R-D loss function. It only incorporates SVD during the training phase to calculate the basis error, without introducing additional network parameters or computational overhead during inference.

III. Experimental Results

A. Implementation Details

We evaluate the proposed optimization framework on several neural image codecs optimized by the conventional R-D loss, including bmshj2018-hyperprior [13] (bmshj2018 for short), mbt2018 [14], cheng2020-anchor [5] (cheng2020 for short), and LIC-TCM [15]. For the first three codecs, we utilize the implementations available on the CompressAI[1] platform,

[1] https://interdigitalinc.github.io/CompressAI

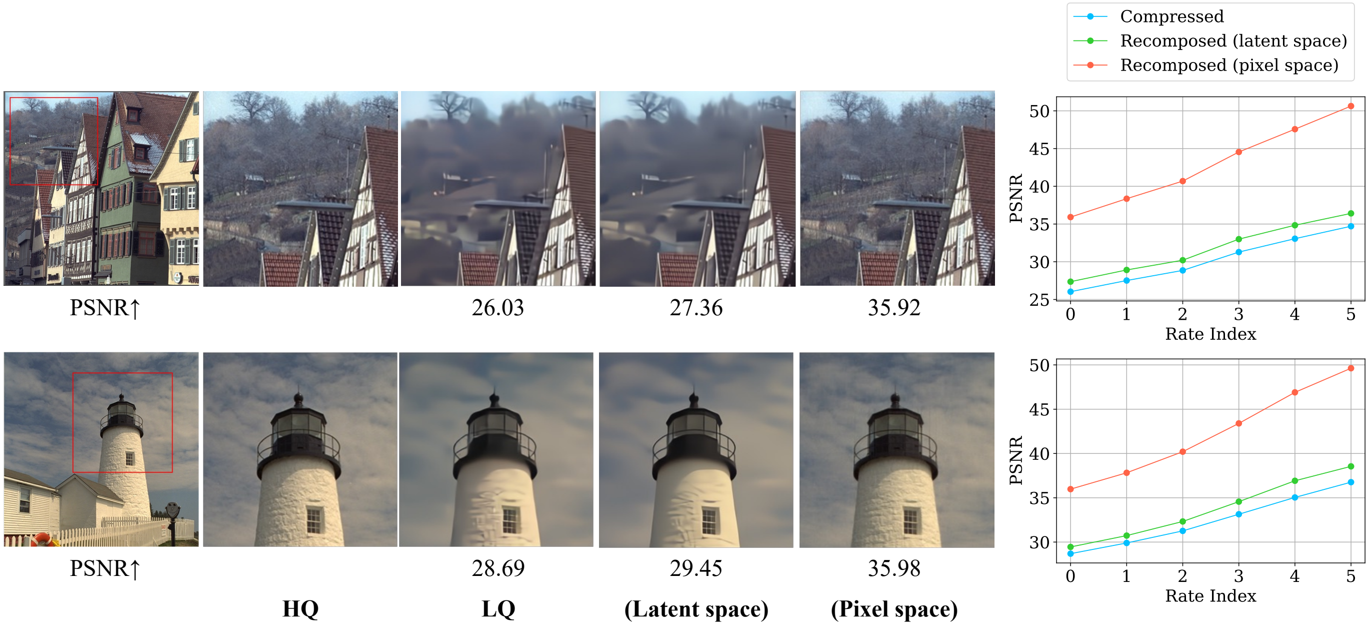


Fig. 1: Visualization of the compressed results (3rd column) and recomposed results in both latent space (4th column) and pixel space (5th column), the corresponding PSNR values are provided under the results. Zoom in for a better preview.

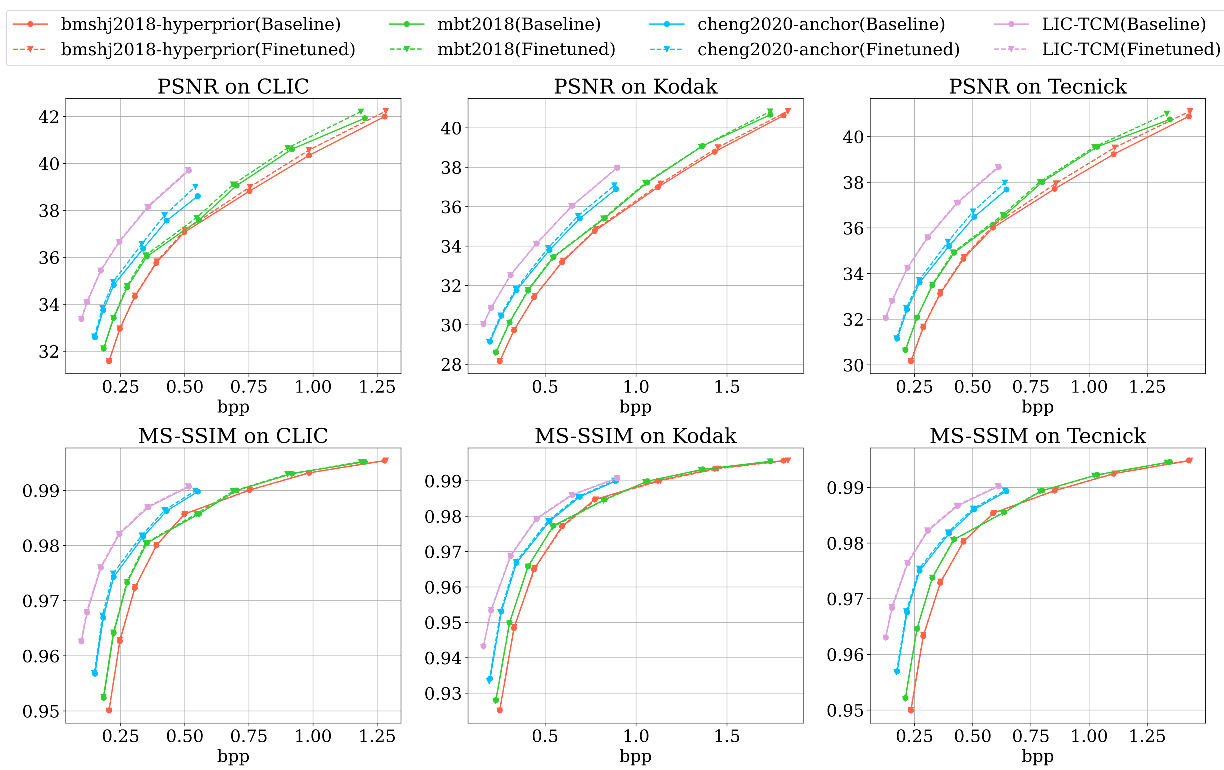
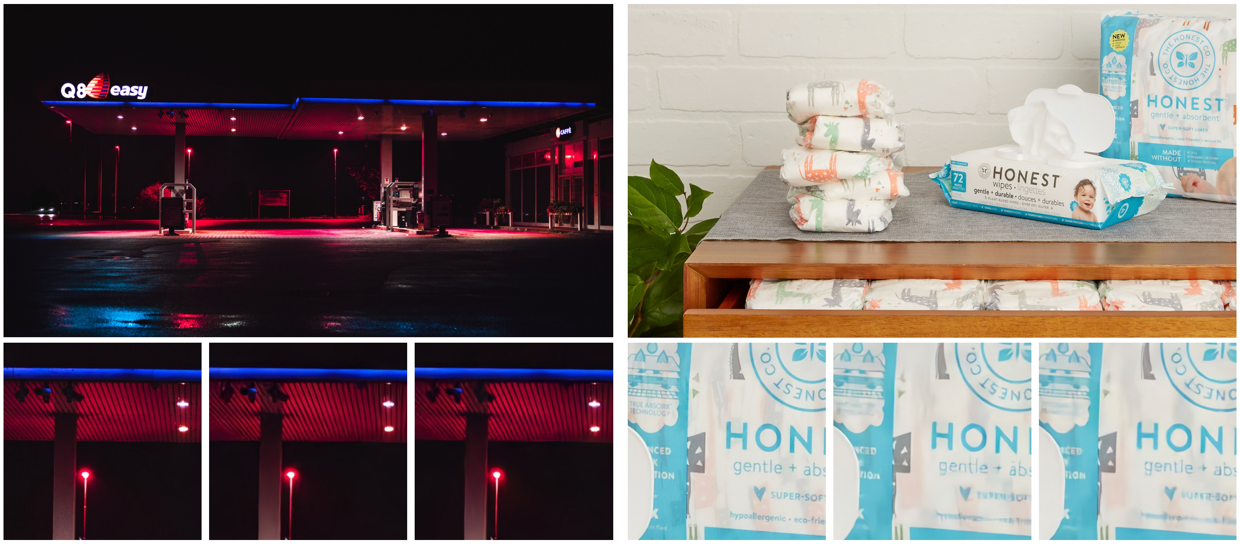


Fig. 2: Illustration of the proposed image compassion framework.

while LIC-TCM is evaluated using its official implementation. We fine-tune the pre-trained model weights provided by their original implementations rather than training from scratch, employing the loss function defined in Eqn. (3) and (4). For fine-tuning, we use randomly cropped image patches of size 256×256 from the DIV2K [16] training set, applying the



bpp / PSNR↑	0.1130 / 34.01	0.1104 / 34.35	bpp / PSNR↑	0.1425 / 31.45	0.1370 / 31.48
HQ	Baseline	**Ours**	HQ	Baseline	**Ours**

Fig. 3: Visualization of reconstructed images from CLIC professional dataset based on the original cheng2020-anchor codec (Baseline) and the one optimized with the proposed framework (Ours).

Model	Kodak	CLIC	Tecnick
bmshj2018	0.00	0.00	0.00
Ours(bmshj2018)	-1.74	-0.81	-1.96
mbt2018	0.00	0.00	0.00
Ours(mbt2018)	-2.49	-0.66	-1.27
cheng2020	0.00	0.00	0.00
Ours(cheng2020)	-5.00	-2.91	-3.92
LIC-TCM	0.00	0.00	0.00
Ours(LIC-TCM)	-1.14	-0.47	-0.84

TABLE I: The BD-rates (%) of our method on three test sets, which are calculated relative to the original codecs, with lower values indicating better compression performance.

Adam optimizer with an initial learning rate of $1e-5$ and a batch size of 32. For evaluation, we use the Kodak [17], CLIC [18] professional and the Tecnick [19] dataset. To evaluate the R-D performance, we measure the rate in bits per pixel (bpp) and assess the quality of the decompressed images using PSNR and MS-SSIM [20].

B. R-D Performance

To compare the R-D performance of the proposed optimization framework with the conventional R-D optimization, we present the R-D curves in terms of PSNR and MS-SSIM, as illustrated in Fig. 2. Our results indicate that the proposed method achieves comparable, or even superior, reconstruction quality while using slightly fewer bits in most cases compared to the baseline method utilizing pre-trained models. This improvement suggests that the SVD-based basis error effectively guides the reconstruction of better bases, thereby enhancing the quality of the final output. Moreover, the observed reduction in bitrates is primarily due to the introduction of the additional basis error term in the distortion calculation. As the distortion increases with a fixed λ, the bitrate tends to decrease when optimizing using the R-D loss depicted in Eqn. (3).

In Table I, we report the BD-rates [21] for quantitative comparisons between the traditional R-D optimization (upper) and our proposed optimization framework (lower). For each codec, the BD-rate is calculated relative to the pre-trained models, resulting in a BD-rate of 0 for the original codecs. Notably, the proposed method consistently achieves performance gains across different neural image codecs on various test sets. Furthermore, since we do not introduce additional network parameters or processes at inference time, the encoding and decoding time remain unchanged.

C. Qualitative Results

Fig. 3 presents visual comparisons of the baseline method and our approach, showcasing two examples from the CLIC professional dataset, with the original cheng2020-anchor used as the baseline. The corresponding bitrate and PSNR values are displayed beneath the results. Our method, which is optimized for both pixel distortion and basis distortion, achieves comparable visual quality to the baseline while utilizing less bitrate and even attaining higher quantitative PSNR metrics.

IV. Conclusions

In this paper, we present a novel rate-distortion optimization framework informed by our exploration of restoring compressed images through the perspective of decomposition and recomposition using SVD. Specifically, we introduce an SVD-based basis reconstruction error into the conventional R-D loss function, facilitating the learning of restoration-enhancing information. Our optimization framework can be seamlessly integrated into various codecs optimized by the R-D loss without introducing additional parameters or computational overhead during inference. Extensive experiments demonstrate that our method consistently achieves performance gains across multiple neural image codecs and datasets. These results underscore the effectiveness of leveraging SVD perspectives to enhance image compression efficiency.

References

[1] G. K. Wallace, "The JPEG still picture compression standard," *Communications of the ACM*, vol. 34, no. 4, pp. 30–44, 1991.

[2] D. Taubman, "JPEG 2000: Image compression fundamentals, standards and practice," 2002.

[3] 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*, vol. 22, no. 12, pp. 1649–1668, 2012.

[4] 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*, vol. 31, no. 10, pp. 3736–3764, 2021.

[5] 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 Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[6] D. He, Z. Yang, W. Peng, R. Ma, H. Qin, and Y. Wang, "ELIC: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5718–5727.

[7] A. B. Koyuncu, H. Gao, A. Boev, G. Gaikov, E. Alshina, and E. Steinbach, "Contextformer: A transformer with spatio-channel attention for context modeling in learned image compression," in *European Conference on Computer Vision*. Springer, 2022, pp. 447–463.

[8] J. Liu, H. Sun, and J. Katto, "Learned image compression with mixed transformer-cnn architectures," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 14 388–14 397.

[9] L. Huang, Y. Xia, and T. Ye, "Effective blind image deblurring using matrix-variable optimization," *IEEE Transactions on Image Processing*, vol. 30, pp. 4653–4666, 2021.

[10] Q. Guo, C. Zhang, Y. Zhang, and H. Liu, "An efficient svd-based method for image denoising," *IEEE transactions on Circuits and Systems for Video Technology*, vol. 26, no. 5, pp. 868–880, 2015.

[11] C. Bao, J.-F. Cai, and H. Ji, "Fast sparsity-based orthogonal dictionary learning for image restoration," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2013.

[12] J. Zhang, J. Huang, M. Zhou, C. Li, and F. Zhao, "Decomposition ascribed synergistic learning for unified image restoration," *arXiv preprint arXiv:2308.00759*, 2023.

[13] 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.

[14] D. Minnen, J. Ballé, and G. D. Toderici, "Joint autoregressive and hierarchical priors for learned image compression," vol. 31, 2018.

[15] J. Liu, H. Sun, and J. Katto, "Learned image compression with mixed transformer-cnn architectures," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 1–10.

[16] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 126–135.

[17] Eastman Kodak Company, "Kodak lossless true color image suite." http://r0k.us/graphics/kodak/

[18] G. Toderici, W. Shi, R. Timofte, L. Theis, J. Balle, E. Agustsson, N. Johnston, and F. Mentzer, "Workshop and challenge on learned image compression (CLIC2020)," in *CVPR*, 2020.

[19] N. Asuni, A. Giachetti, *et al.*, "Testimages: a large-scale archive for testing visual devices and basic image processing algorithms." in *STAG*, 2014, pp. 63–70.

[20] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2. Ieee, 2003, pp. 1398–1402.

[21] G. Bjontegaard, "Calculation of average psnr differences between rd-curves," *ITU-T SG16 Q*, vol. 6, 2001.