



HAL
open science

INR-MDSQC: Implicit Neural Representation Multiple Description Scalar Quantization for robust image Coding

Trung Hieu Le, Xavier Pic, Marc Antonini

► **To cite this version:**

Trung Hieu Le, Xavier Pic, Marc Antonini. INR-MDSQC: Implicit Neural Representation Multiple Description Scalar Quantization for robust image Coding. 2023. ⟨hal-04148808⟩

HAL Id: hal-04148808

<https://hal.science/hal-04148808v1>

Preprint submitted on 26 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

INR-MDSQC: Implicit Neural Representation Multiple Description Scalar Quantization for robust image Coding

Trung Hieu Le, Xavier Pic, Marc Antonini, member IEEE

13S laboratory, Côte d'Azur University CNRS,UMR 7271, Sophia Antipolis, France

Abstract—Multiple Description Coding (MDC) is an error-resilient source coding method designed for transmission over noisy channels. We present a novel MDC scheme employing a neural network based on implicit neural representation. This involves overfitting the neural representation for images. Each description is transmitted along with model parameters and its respective latent spaces. Our method has advantages over traditional MDC that utilizes auto-encoders, such as eliminating the need for model training and offering high flexibility in redundancy adjustment. Experiments demonstrate that our solution is competitive with autoencoder-based MDC and classic MDC based on HEVC, delivering superior visual quality.

Index Terms—Multiple Description Coding(MDC), Autoregressive Model, Synthesis Model, MultiLayer Perceptron (MLP), Implicit Neural Representation (INR)

I. INTRODUCTION

Multiple Description Coding (MDC) has been studied for many years. In [1], the authors presented an efficient source coding solution able to manage packet errors, random bit errors and routing delays. MDC for image encoding involves encoding multiple representations of an image; if one is lost or corrupted during transmission, the remaining descriptions can still be used to reconstruct the original image with some quality degradation.

Classic MDC methods have typically dealt with some problems. The first application with a scalar quantizer was proposed in [2], where the index assignment refers to the process of mapping from the source to a set of output descriptions to achieve the best rate, redundancy, and distortion trade-off. This problem is complex. The wavelet transform MDC is based on [3], where authors confront issues of quantization and redundancy index assignment, and attempt to solve the problem of non-linearity during optimization by modeling each subband a Gaussian model. However, this model has limited accuracy at lower rates, and its complexity is very high. Standard-compliant MDC methods such as HEVC [4], [5], [6] can achieve high performance with low latency. However, rate distortion control is carried out with an empirical formula that is based on linear regression, which limits the quantization range and thereby constrains their performance.

Recent research has indicated the potential use of neural networks for image compression [7], [8], [9], but a few work have applied it to MDC. The most recent applications for MDC

are [10], [11], which employ Generative Networks and Compressive Autoencoders. However, these methods' drawbacks include the requirement of a high computing capacity for the training process. Furthermore, the training process must be performed with very large datasets to be efficient. This is even more challenging in the MDC context due to the redundancy adaptation mechanism, which requires retraining the model.

In recent image compression research using neural networks, the so-called *Implicit Neural Representation* (INR), the neural network learns to represent an image implicitly through its weights, a coordinate map, and possibly a latent space [12], [13]. More recently, the Coordinate-based Low Complexity Hierarchical Image Codec (COOL-CHIC) framework [14] has achieved performance close to the state of the art of the compressive autoencoder presented in [7], without the need for a training process. In this paper, we propose the Implicit Neural Representation Multiple Description Scalar Quantization Codec (INR-MDSQC) method based on the COOL-CHIC architecture. The advantages of the proposed solution are: no need for model training, high performance and flexible redundancy tuning.

In the rest of the paper, we first formulate in section II-A our MDC problem using this network. Then, we present in section II-B and II-C the detailed architectures of the synthesis model and the auto-regressive model, both of which are optimized during training. Following that, we discuss in section II-D the post-training quantization process designed for precision reduction. Lastly, in section II-E we outline the bitstream's organization and the decoding process. Finally, we show the experimental result in section III and conclude the paper in section IV.

II. PROPOSED METHOD

A. Multiple description problem statement

Inspired by the COOL-CHIC framework [14], we propose an overfitted INR-MDSQC network based on hierarchical latent scalar quantization. The architecture of the INR-MDSQC has three main components as illustrated in figure 1:

- 1) **Two sets of discrete hierarchical latent spaces:** \mathbf{y}_1 and \mathbf{y}_2 for descriptions 1 and 2, respectively. Then, \mathbf{y}_0 is constructed from the interlacing of \mathbf{y}_1 and \mathbf{y}_2 . For each set, $\hat{\mathbf{y}}_j, \forall j \in \{0, 1, 2\}$ represents their quantized versions.

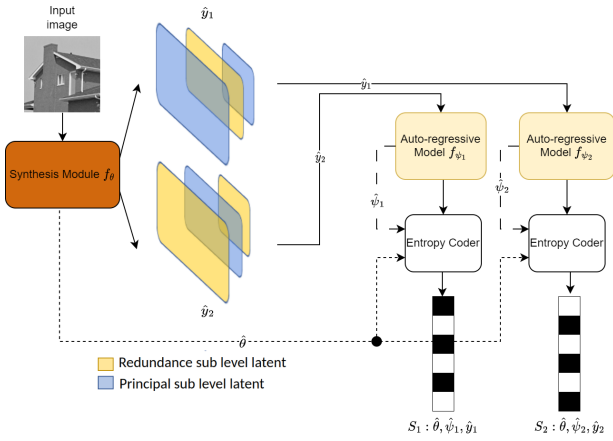


Fig. 1. **INR-MDSQC**: Given an image, the synthesis model (f_θ) divides it into two latent spaces, \hat{y}_1 and \hat{y}_2 . Each latent space is then compressed based on the probability estimates derived from the auto-regressive model. The entropy coder proceeds to compress both the model parameters (θ, ψ) and the pixels in the latent space to generate two descriptions S_1 and S_2

- 2) **Synthesis model (f_θ)**: A Multi-Layer Perceptron (MLP) that creates $\hat{y}_j, \forall j \in \{1, 2\}$ from the original image (θ represents its parameters).
- 3) **Auto-regressive model (f_{ψ_j})**: A MLP that estimates the distribution of subsequent pixels based on previously decoded pixels for latent spaces $\hat{y}_j, \forall j \in \{1, 2\}$ (ψ_j represents its parameters).

In the COOL-CHIC framework, image encoding is achieved by overfitting parameters $\{\theta, \psi, \hat{y}\}$ to the image characteristics, and transmission is carried out by transmitting these parameters. INR-MDSQC takes inspiration from this, generating two descriptions $S_1 : \{\theta, \psi_1, \hat{y}_1\}$ and $S_2 : \{\theta, \psi_2, \hat{y}_2\}$. The encoding minimizes a cost function, that takes into consideration the difference between the descriptions in the latent spaces. Depending on the number of received descriptions, we use either \hat{y}_1 or \hat{y}_2 for reconstruction. If all the descriptions are received, the interlaced \hat{y}_0 is used for reconstruction. The images reconstructed with f_θ are denoted as \hat{x}_1, \hat{x}_2 , and \hat{x}_0 , respectively. The distortion metric of each reconstruction relative to the original is the Mean Squared Error (MSE) and is defined as follows:

$$D_j = \frac{1}{C \times W \times H} \sum_{i=1}^{C \times W \times H} (\hat{x}_{i|j} - x_i)^2 \quad (1)$$

where $\hat{x}_{i|j} \in \hat{x}_j$ and $\forall j \in \{0, 1, 2\}$

with C represents the number of image channels (for example 3 for RGB, 1 for gray image), W is the image width, H is the image height, and i is the position of the pixels in raster-scan order. Thus, we denote distortions in MSE of side reconstructions with D_1, D_2 and central reconstruction with D_0 .

In each set S_j with $j \in \{1, 2\}$, the latent space is entropy-coded for efficient transmission. This requires estimating the probability distribution p of each value from the unknown signal probability distribution q . The entropy coding algorithm

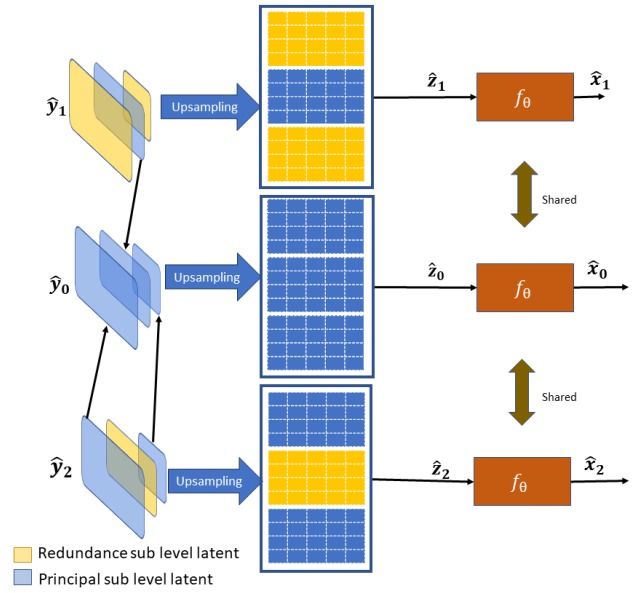


Fig. 2. **Synthesis model**: In this example, with three decomposition levels ($N=3$), the central latent set \hat{y}_0 is created by interleaving two side latent sets \hat{y}_1 and \hat{y}_2 as showed in the figure. Then they are upsampled to create the sets \hat{z}_0, \hat{z}_1 and \hat{z}_2 respectively. Finally, each up-sampled set is fed into a shared MLP for reconstruction.

can asymptotically achieve the rate of the signal's cross-entropy $H(\hat{y}_j)$, which is given by:

$$H(\hat{y}_j) = -E_{\hat{y}_j \sim q}[\log_2 p(\hat{y}_j)], \forall j \in \{1, 2\} \quad (2)$$

To estimate the distribution p , the autoregressive model f_{ψ_j} estimated the entropy with the input \hat{y}_j : $p_{\psi_j} = f_{\psi_j}(\hat{y}_j)$. Therefore, we can establish the global MDC cost function as:

$$J = D_0 + \alpha \sum_{j=1}^2 D_j + \sum_{j=1}^2 \lambda_j (R(\hat{y}_j) + R(\psi_j) + R(\theta)) \quad (3)$$

where $\alpha \in [0, 1]$ is redundancy factor, $R(\hat{y}_j)$ is the rate for the latent space \hat{y}_j from p_{ψ_j} and will be defined in section II-C. The rates $R(\theta)$ and $R(\psi_j)$ are estimated for the model parameters and will be described in section II-D. As the selected MLP is small in size, the bitrate costs for the parameters θ and ψ_j are considered negligible during training and only comes into play in the post-training optimization process. Therefore, from 3 the MDC cost function J_t used in the training process is defined as:

$$J_t = D_0 + \alpha \sum_{j=1}^2 D_j + \sum_{j=1}^2 \lambda_j R(\hat{y}_j) \quad (4)$$

Our training objective involves minimizing the cost function J_t , as given in Equation (4). The minimization process is represented below:

$$\underset{\theta, \psi_1, \psi_2, \hat{y}_1, \hat{y}_2}{\text{minimize}} J_t \quad (5)$$

B. Multiple description synthesis model

First we defined the uniform scalar quantization Q as:

$$\hat{s} = Q(s, \Delta s) \quad (6)$$

with s is the element to quantize and Δs is its associated quantization step. We define $\mathbf{y}_{k|j}$ as the 2D latent space at level k of description j . Each of these has a unique quantization step, and their quantized version $\hat{\mathbf{y}}_{k|j}$ is defined as:

$$\hat{\mathbf{y}}_{k|j} = Q(\mathbf{y}_{k|j}, \Delta \mathbf{y}_{k|j})$$

The quantized set of latent spaces $\hat{\mathbf{y}}_j$ for each description j is defined as:

$$\hat{\mathbf{y}}_j = \{\hat{\mathbf{y}}_{k|j} \in \mathbb{Z}^{H_k \times W_k}, k = 0, \dots, N-1\}$$

where $H_k = \frac{H}{2^k}$, $W_k = \frac{W}{2^k}$, and N represents the total number of hierarchical levels of $\hat{\mathbf{y}}_j$. When transmission is achieved without any loss of packets, we can fully receive both $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_2$, and the central latent space $\hat{\mathbf{y}}_0$ is the product of interleaving between $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_2$ and it is defined as:

$$\hat{\mathbf{y}}_0 = \{\hat{\mathbf{y}}_{2k'+1|1}, \hat{\mathbf{y}}_{2k'+1|2}, k' = 0, 1, \dots, \lfloor N/2 \rfloor\}$$

We design the MDC synthesis model as shown in Figure 2, each level of latent space will be up-sampled to the image size of $[H \times W]$ using bi-cubic interpolation. For each level in the hierarchy we have their up-sampled version:

$$\hat{\mathbf{z}}_{k|j} = \text{upsampled}(\hat{\mathbf{y}}_{k|j})$$

In the end, the shape of the upsampled latent space $\hat{\mathbf{z}}_j$ is $[N \times H \times W]$. Then, the synthesis model (f_θ) presents each pixel in the reconstructed image as a function of the up-sampled latent space as follows:

$$\hat{\mathbf{x}}_j = f_\theta(\hat{\mathbf{z}}_j) \text{ with } \hat{\mathbf{z}}_j = \{\hat{\mathbf{z}}_{k|j}, k = 0..N-1\}$$

Inspired from LMDC [11], during training, the three up-sampled sets $\hat{\mathbf{z}}_1, \hat{\mathbf{z}}_2, \hat{\mathbf{z}}_0$ are fed into a shared synthesis model. The network's goal is to minimize the cost function (4), with the differences in distortion between the side and central reconstructions being dependent on the redundancy factor α , which ranges from 0 to 1. The setup of the cost function (4) compels the Synthesis model to partition the image information into two sets of latent descriptions, $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2$, under rate constraints. However, because the latent space is discrete and the quantization process isn't differentiable, uniform noise is introduced, based on [7]. This introduction of noise allows for a differentiable operation, thereby enabling gradient-based optimization. The process is detailed as:

$$\hat{\mathbf{y}}_j = \begin{cases} \mathbf{y}_j + u, u \sim \mathcal{U}[-0.5, 0.5] & \text{during training} \\ Q(\mathbf{y}_j) & \text{otherwise} \end{cases}$$

where u : the uniform noise and $\forall j \in \{1, 2\}$

C. Auto-regressive probability model

The auto-regressive probability model so called f_{ψ_j} , implemented as MLP, aims to closely estimate the image's unknown latent distribution as p_{ψ_j} . Since the distribution of each pixel in

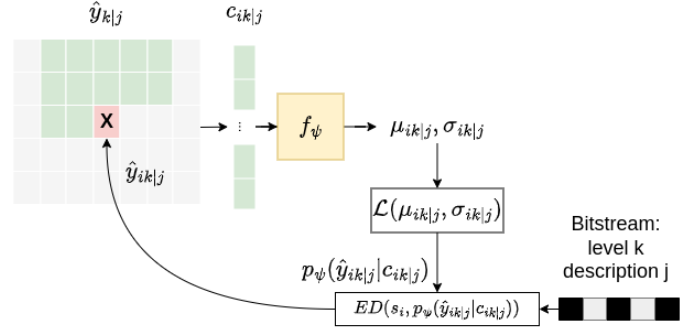


Fig. 3. **Autoregressive model:** In this example, the model uses 12 pixels, $c_{ik|j}$, to yield $\mu_{ik|j}$ and $\sigma_{ik|j}$, modeling a Laplacian distribution. The symbol probability is calculated, and an entropy decoder estimates the latent pixel, $\hat{y}_{ik|j}$, from a bitstream.

the latent space is conditioned by their neighbor, according to [15], the probability of the pixels is determined by a factorized model:

$$p_{\psi_j}(\hat{\mathbf{y}}_j) = \prod_{i,k} p_{\psi_j}(\hat{y}_{ik|j} | c_{ik|j}) \quad (7)$$

With $\hat{y}_{ik|j}$ is the latent pixel at the position i of level k of description j and $c_{ik|j}$ are the set of decoded neighbors pixels of $\hat{y}_{ik|j}$. Therefore, $c_{ik|j} \in \mathbb{Z}^C$ where C is the set of causal spatially neighboring pixels.

The discrete distribution $p_{\psi_j}(\hat{\mathbf{y}}_j)$ of quantized latent variables is modeled by integrating the continuous distribution of the non-quantized latent $g(y_i)$, modeled as a Laplace distribution. The MLP f_{ψ_j} learns to estimate proper expectation ($\mu_{ik|j}$) and scale ($\sigma_{ik|j}$) parameters for Laplacian distribution g of the set of context pixels $c_{ik|j}$. Consequently, the probability of a latent pixel is modeled as:

$$p_{\psi_j}(\hat{y}_{ik|j} | c_{ik|j}) = \int_{\hat{y}_{ik|j}-0.5}^{\hat{y}_{ik|j}+0.5} g(y) dy$$

where $g \sim \mathcal{L}(\mu_{ik|j}, \sigma_{ik|j})$ and $\mu_{ik|j}, \sigma_{ik|j} = f_{\psi_j}(c_{ik|j})$.

As the p_{ψ_j} approximates the real probability of latent space. From the article [14], by using the factorized model equation (7), the rate defined in equation (2) can be expressed as:

$$\begin{aligned} R(\hat{\mathbf{y}}_j) &= -\log_2(p_{\psi_j}(\hat{\mathbf{y}}_j)) = -\log_2 \prod_{i,k} p_{\psi_j}(\hat{y}_{ik|j} | c_{ik|j}) \\ &= -\sum_{i,k} \log_2 p_{\psi_j}(\hat{y}_{ik|j} | c_{ik|j}) \end{aligned} \quad (8)$$

In our MDC scheme, we aim to quantize with coarser grains in redundant latent levels and finer grains in principal ones. Given the multi-resolution latent organization, the smallest resolution latent space that captures low-frequency information is more critical and must be quantized finely. Therefore, we introduce a spatial resolution coefficient β_k to avoid excessive quantization. From equation (8), our final MDC weighted rate function becomes:

$$R(\hat{\mathbf{y}}_j) = -\sum_{i,k} \beta_k \log_2 p_{\psi_j}(\hat{y}_{ik|j} | c_{ik|j}) \text{ where } \beta_k = \frac{W_k \times H_k}{2^{2k}}$$

D. Model parameters quantization

Compressing the INR-MDSQC model parameter consists of compressing $\{\psi_1, \psi_2, \theta\}$. During the training phase, 32-bit floating-point precision was used. However, once the training is finished, such high-precision representation is not required. From equation (6), we use three separate quantization steps, $\Delta_{\psi_1}, \Delta_{\psi_2}$ and Δ_{θ} , to produce $\hat{\psi}_1, \hat{\psi}_2$ and $\hat{\theta}$, respectively. The entropy coder needs a probability model for each quantized model symbol $\hat{s}_i \in \hat{s}$, where $\hat{s} \in \{\hat{\psi}_1, \hat{\psi}_2, \hat{\theta}\}$, in order to encode it. Empirically, the distribution of model parameters is usually best approximated by a Laplace distribution centered at 0. Therefore, we employ a Laplacian model to estimate the entropy of \hat{s}_i :

$$p(\hat{s}_i) = \int_{\hat{s}_i-0.5}^{\hat{s}_i+0.5} g(s) ds$$

where $g \sim \mathcal{L}(0, \sigma_{\hat{s}})$, $\sigma_{\hat{s}}$ is the standard deviation. Same as function (8), the estimated rate function of \hat{s} can be expressed as:

$$R(\hat{s}) = - \sum_{\hat{s}_i \in \hat{s}} \log_2(p(\hat{s}_i))$$

We denote $\hat{\mathbf{y}}_j$ as the quantized latent space, which is fixed after training. Drawing from the cost function in (3), our MDC post-training J_p is defined as:

$$J_p = D_0 + \alpha \sum_{j=1}^2 D_j + \sum_{j=1}^2 \lambda_j (R(\hat{\mathbf{y}}_j) + R(\hat{\theta}) + R(\hat{\psi}_j))$$

As post-training quantization involves quantizing the model parameters, the optimization depends solely on $\hat{\theta}$, $\hat{\psi}_1$, and $\hat{\psi}_2$. Therefore, we can represent the optimization as follows:

$$\underset{\hat{\theta}, \hat{\psi}_1, \hat{\psi}_2}{\text{minimize}} \quad J_p \quad (9)$$

The solution to the minimization problem 9 above is achieved by finding the best set $\{\Delta_{\psi_1}, \Delta_{\psi_2}, \Delta_{\theta}\}$ within a predefined range (e.g., from 10^{-1} to 10^{-5}). To identify the optimal set, we independently conduct a linear search for each module. This procedure involves incrementing the quantization step sequentially and locating the corresponding step of minimum cost for each module. We apply the discovered quantization step to its respective module before moving on to the next. In our approach, we initiate this linear search with Δ_{θ} , followed by $\Delta_{\psi_1}, \Delta_{\psi_2}$.

E. Bitstream structure

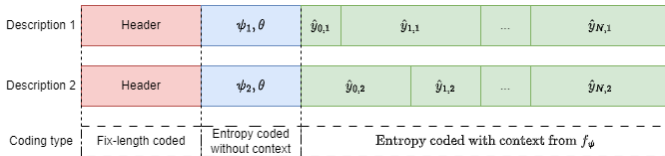


Fig. 4. Bit-stream structure

From the trained and quantized model, the transmitted data, formatted as depicted in Figure 4, starts with a header detailing

the decoder's configuration parameters, including image size, layer count, model parameter quantizer steps $\Delta_{\psi_1}, \Delta_{\psi_2}, \Delta_{\theta}$, and context pixel count. \hat{s}_i where s can be either ψ_1, ψ_2 or θ are entropy-coded using respective probabilities derived from $g \sim \mathcal{L}(0, \sigma_{\hat{s}})$. Finally, each latent pixel is entropy-coded using estimated probabilities from p_{ψ_j} as discussed earlier.

At decoding, the header is decoded first, followed by the network parameters. Latent pixels are then decoded from the bitstream using an entropy coder initialized with source statistics estimated by the decoded auto-regressive model. Depending on the number of received descriptions, the image is reconstructed from the decoded latent space directly or via an interlacing operation between the latent space levels of the two descriptions.

III. EXPERIMENTAL RESULT

A. Implementation detail

The framework is optimized using Synthesis MLP and Auto-regressive MLP, both featuring two equal-sized hidden layers with 12 units each and the non linear layer used is a ReLU. Our solution is evaluated firstly with two images extracted from the SET4¹ dataset, Lena (4.png) and the boat (1.png), and secondly, with all the images from the DIV2k² dataset. We set the number of hierarchical levels to 6 for Lena and boat images, and to 8 for the DIV2K dataset. The Adam optimizer is used for optimization, with an initial learning rate of $l_r = 0.1$ and 10000 iterations. For the high-resolution DIV2K dataset, we only tested with $\alpha = 0.1$, while for the other datasets, we tested with $\alpha = 0.1$ and $\alpha = 1.0$. The distortion is estimated through Peak Signal-to-Noise Ratio (PSNR) and Multi-Scale Structural Similarity Index (MS-SSIM) [16], while bit-per-pixel (bpp) is used to determine the compression rate. The entropy codec used is Range Coder from [17].

The image's central reconstruction is the result of the interlacing process between two descriptions of different hierarchical levels, hence the quantity of information captured in the latent spaces $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_2$ is unequal. Indeed, because the way the two descriptions are defined in the latent space, the proposed MDC is unbalanced. Therefore, for the same value of λ_j between two descriptions, the description which contain the lowest latent resolution as the principal latent will exhibit higher quality than the other. This property is beneficial because each description is transmitted over distinct, independent channels, each one possessing unique characteristics. The description of higher quality is dispatched via a less noisy channel, while the description of lower quality is conveyed through a channel with more noise. Due to the unequal rate, there is a consequent unequal distortion between the two descriptions. The side distortion curve showed in the figure 5 represents the PSNR of the average of the MSE across all side descriptions.

To ensure the validity of our solution, the performance of our INR-MDSQC at central reconstruction should not surpass

¹https://github.com/mdcnn/MDCNN_test40/tree/master/SET4

²https://github.com/mdcnn/MDCNN_test40/tree/master/DIV2K-10

the upper limit of the SDC set by the original coder COOL-CHIC, nor should it fall below the SDC at double rate. As showcased in the Lena and boat images (see Figure 5), with $\alpha = 0.1$, the solution approaches the upper bound limit of the single SDC. When full redundancy is applied, $\alpha = 1.0$, the performance of our MDC coincides with the lower limit SDC at double rate.

B. Rate distortion study

We benchmarked our solution against LMDC [11]. Additionally, we compared our approach with classic MDC methods such as HEVC-MDC (HMDC)[4]. The results for LMDC can be found in [11], and the results for HMDC were obtained from a re-implementation of the method.

With the small images of lena and of the boat from the SET4, with redundancy $\alpha = 0.1$, our solution shows an improvement in PSNR at high bit-rates compared to LMDC and overperform the HMDC method. This can be attributed to the fact that, at low bit-rates, the cost of coding θ, ψ_j becomes significant. However, at high bit-rates, our solution adapts more effectively to the image characteristics, thus enhancing the quality of the reconstruction. In terms of the MS-SSIM metric, our method also surpasses the LMDC and HMDC methods. Figure 6 presents a comparison of the visual outcomes from HMDC and our solution. Our method tends to produce less blocking artifacts. When full redundancy ($\alpha = 1.0$) is applied, our method achieves higher performance in side reconstruction, with improved central reconstruction performance compared to LMDC

In the high-resolution dataset DIV4K-10, traditional MDC strategies such as HMDC still outperform INR-MDSQC methods in terms of PSNR for side reconstruction. However, INR-MDSQC achieves nearly identical PSNR performance for central reconstruction. Moreover, in terms of MS-SSIM, our solution outperforms the HMDC method. Importantly, our method maintains superior central reconstruction performance compared to LMDC while achieving similar distortion levels for side reconstruction.

IV. CONCLUSION

We introduce INR-MDSQC, an Implicit Neural Representation Multiple Description Scalar Quantizer Codec, which is built based on the COOL-CHIC framework. By overfitting the neural network for each image, INR-MDSQC can capture more details, thereby enhancing performance compared to the traditional Autoencoder MDC approach. Furthermore, this framework allows for more flexible redundancy tuning. When compared to conventional MDC frameworks, our solution delivers superior reconstruction quality in term of MS-SSIM at almost the same central PSNR. From our perspective, a study aimed at reducing complexity is necessary to enhance the method's efficiency. Moreover, an evaluation of the system's performance under noisy channel conditions will be required in our future works.

Acknowledgement

We would like to extend our gratitude to Jérémy Mateos for his innovative ideas, suggestions that improved the perfor-

mance of our model.

REFERENCES

- [1] Vivek K. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 74–93, 2001.
- [2] Vinay A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 821–834, 1993.
- [3] Manuela Pereira, Marc Antonini, and Michel Barlaud, "Multiple description coding for internet video streaming," in *ICIP (3)*. 2003, pp. 281–284, IEEE.
- [4] Trung Hieu Le, Marc Antonini, Marc Lambert, and Karima Alioua, "Multiple description video coding for real-time applications using HEVC," *arXiv preprint arXiv:2303.05843*, 2023.
- [5] Tammam Tillo, Marco Grangetto, and Gabriella Olmo, "Redundant slice optimal allocation for H.264 multiple description coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 59–70, 2008.
- [6] Feifeng Wang, Jing Chen, Huanqiang Zeng, and Canhui Cai, "Spatial-frequency HEVC multiple description video coding with adaptive perceptual redundancy allocation," *J. Vis. Commun. Image Represent.*, vol. 88, pp. 103614, 2022.
- [7] Johannes Ballé, Valero Laparra, and Eero P. Simoncelli, "End-to-end optimization of nonlinear transform codes for perceptual quality," in *2016 Picture Coding Symposium, PCS 2016, Nuremberg, Germany, December 4-7, 2016*. 2016, pp. 1–5, IEEE.
- [8] Lucas Theis, Wenzhe Shi, Andrew Cunningham, and Ferenc Huszár, "Lossy image compression with compressive autoencoders," in *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. 2017, OpenReview.net.
- [9] Fabian Mentzer, Eirikur Agustsson, Michael Tschannen, Radu Timofte, and Luc Van Gool, "Conditional probability models for deep image compression," in *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*. 2018, pp. 4394–4402, Computer Vision Foundation / IEEE Computer Society.
- [10] Lijun Zhao, Huihui Bai, Anhong Wang, and Yao Zhao, "Multiple description convolutional neural networks for image compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 8, pp. 2494–2508, 2019.
- [11] Lijun Zhao, Jinjing Zhang, Huihui Bai, Anhong Wang, and Yao Zhao, "LMDC: learning a multiple description codec for deep learning-based image compression," *Multim. Tools Appl.*, vol. 81, no. 10, pp. 13889–13910, 2022.
- [12] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein, "Implicit neural representations with periodic activation functions," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, Eds. 2020, vol. 33, pp. 7462–7473, Curran Associates, Inc.
- [13] Yannick Strümpfer, Janis Postels, Ren Yang, Luc Van Gool, and Federico Tombari, "Implicit neural representations for image compression," in *Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part XXVI*, Shai Avidan, Gabriel J. Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, Eds. 2022, vol. 13686 of *Lecture Notes in Computer Science*, pp. 74–91, Springer.
- [14] Théo Ladune, Pierrick Philippe, Félix Henry, Gordon Clare, and Thomas Leguay, "COOL-CHIC: Coordinate-based low complexity hierarchical image codec," *arXiv preprint arXiv:2212.05458*, 2023.
- [15] David Minnen, Johannes Ballé, and George Toderici, "Joint autoregressive and hierarchical priors for learned image compression," in *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett, Eds., 2018, pp. 10794–10803.
- [16] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
- [17] Robert Bamler, "Understanding entropy coding with asymmetric numeral systems (ans): a statistician's perspective," *arXiv preprint arXiv:2201.01741*, 2022.

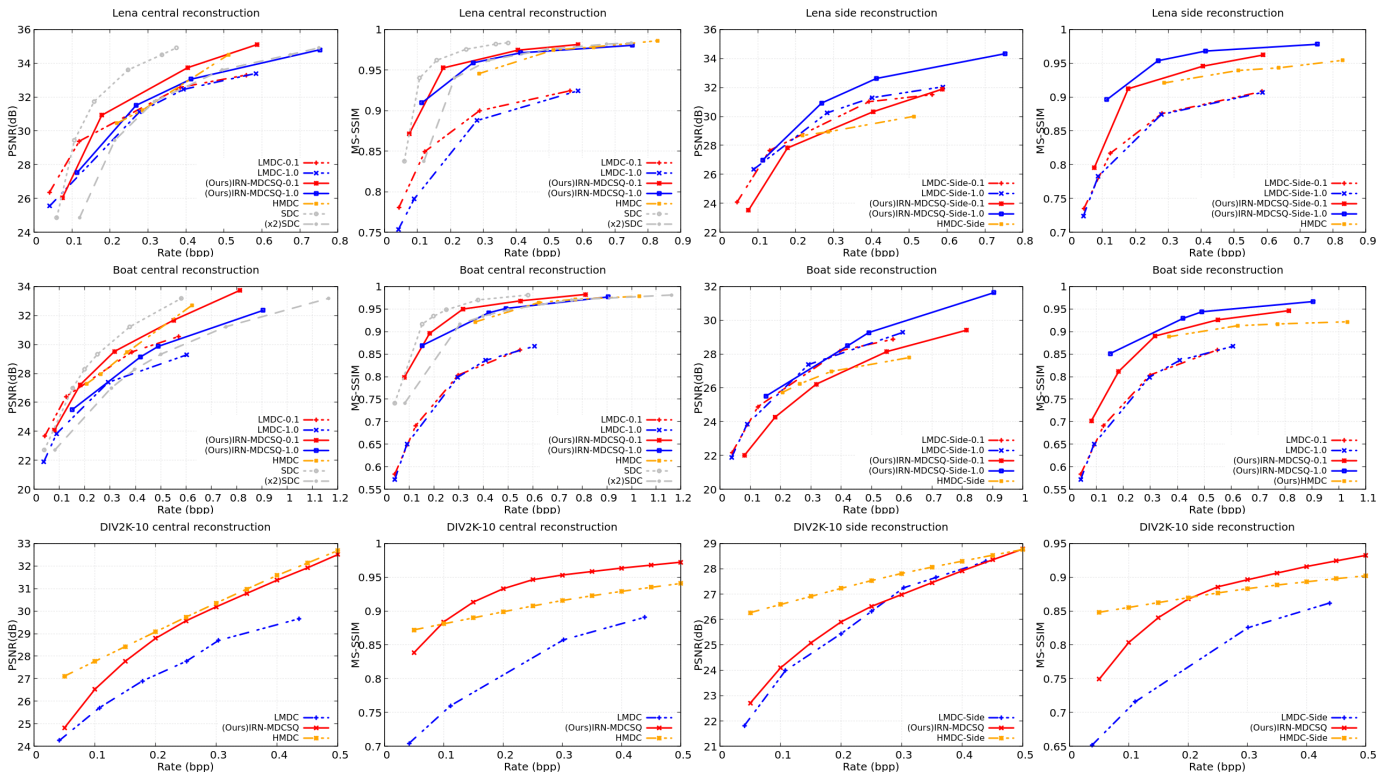


Fig. 5. The rate distortion performance was evaluated on Lena and Boat images (512x512 pixels) for two values of redundancy ($\alpha = 0.1$ and $\alpha = 1$) and the The DIV2K-10 dataset (1920x1080 px) for one redundancy factor ($\alpha = 0.1$)

(a) INR-MDSQC



Central: 31.79dB-0.54bpp

Side1: 27.15dB-0.21bpp

Side2: 28.95dB-0.33bpp

(b)HMDC



Central:31.25dB-0.58bpp

Side1:27.63dB-0.27bpp

Side2:27.96dB-0.31bpp

Fig. 6. The visual results of the image "Boat", shows that the blocking artifact caused by high quantization is more noticeable on the HMDC than INR-MDSQC at a "nearly" identical bitrate.