

# Learnable Non-uniform Quantization With Sampling-based Optimization for Variable-Rate Learned Image Compression

Shaohui Li, *Member, IEEE*, Wenrui Dai, *Member, IEEE*, Nuowen Kan, *Member, IEEE*, Chenglin Li, *Member, IEEE*, Junni Zou, *Member, IEEE*, and Hongkai Xiong, *Fellow, IEEE*

**Abstract**—Variable-rate coding is challenging but indispensable for learned image compression (LIC) that is in nature characterized by nonlinear transform coding (NTC). Existing methods for variable-rate LIC are restricted by the non-smooth quantization process with zero gradients almost everywhere, and consequently, suffer from training-test gap and degraded rate-distortion (R-D) performance. To address this problem, in this paper, we propose sampling-based optimization for training NTC models along with non-uniform quantizers. Different from gradient-based optimization, the proposed sampling-based optimization first randomly samples the parameters from Gaussian distributions with progressively reduced variance and then selects the optimal parameters with a R-D indicator. On the basis of sampling-based optimization, we develop a learnable non-uniform dead-zone quantizer by adaptively refining the quantization steps for variable-rate coding with nonlinear transforms. Furthermore, we incorporate the learnable dead-zone quantizer to achieve a variable-rate LIC model with enhanced R-D performance and design rate and distortion control algorithms to adapt to dynamic network conditions. Experimental results show that the proposed method achieves state-of-the-art R-D performance in variable-rate image compression. It obtains an average 8.82% BD-rate reduction compared to latest versatile video compression (VVC) standard, and simultaneously achieves precise rate and distortion control with an average variation of 0.0087 bpp in bit-rates and 0.1265 dB in distortion on the *Kodak* dataset.

**Index Terms**—Learned image compression, variable-rate image compression, rate control, learnable dead-zone quantizer.

## I. INTRODUCTION

DEEP learning techniques have been widely considered in lossy image compression in recent years [1]–[10]. In learned image compression (LIC), the main components of traditional handcrafted transform coding (*i.e.*, transform, quantization, and entropy coding) are realized with nonlinear neural

This work was supported in part by the National Natural Science Foundation of China under Grants 62301299, 62371288, 62401366, U24A20251, 62320106003, 62431017, 62125109, 62401357, 62120106007, in part by the Program of Shanghai Science and Technology Innovation Project under Grant 24BC3200800, in part by China Postdoctoral Innovation Talents Support Program under Grant BX20230184, and in part by China Postdoctoral Science Foundation under Grant 2024M751976. (*Corresponding author:* Wenrui Dai)

Shaohui Li is with the Shenzhen International Graduate School, Tsinghua University, Shenzhen, Guangdong 518055, China. E-mail: lishao-hui@sz.tsinghua.edu.cn.

Wenrui Dai and Junni Zou are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China. E-mail: daiwenrui@sjtu.edu.cn; zoujunni@sjtu.edu.cn.

Nuowen Kan, Chenglin Li, and Hongkai Xiong are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China. E-mail: kannw\_1230@sjtu.edu.cn; lcl1985@sjtu.edu.cn; xionghongkai@sjtu.edu.cn.

networks and optimized in an end-to-end fashion [15]. Recent LIC models [68], [70], [72], [73] are shown to outperform the latest manufactured compression standard (*i.e.*, Versatile Video Coding [45]) in rate-distortion (R-D) performance.

However, these fixed-rate LIC models cannot accommodate to real-world scenarios, especially image transmission under dynamic network conditions. Fixed-rate models cannot fully utilize the bandwidth due to limited number of available R-D points. In addition, the process of adapting to different bandwidths requires extensive computations to find the optimal model and parameters [61]. In contrast, variable-rate models [12]–[23], [28]–[36] can achieve finer rate adaptation with a single model. They can dramatically improve the utility of varying bandwidth and reduce the storage burden of compression models. Recent works focus on using a naive scaling method with fixed non-linear transforms and uniform quantizer to enable continuous rate and distortion control. Unfortunately, naive scaling can hardly be optimal in R-D performance. In [21], [64], [66], non-uniform quantizers such as dead-zone quantizers have been shown to improve the R-D performance of variable-rate compression methods. However, widely adopted gradient-based optimization methods (*e.g.*, stochastic gradient descent (SGD) and adaptive moment estimation (Adam)) cannot achieve efficient optimization on non-uniform quantizers due to the vanishing gradient problem.

As the core component of lossy image compression, quantization transforms precise input signals into coarse output symbols and is usually incorporated with transform coding to significantly reduce bit-rates at tolerable loss of reconstruction quality. In theory, any lossy compressor can be viewed as a quantizer. However, it is challenging to introduce a sophisticated quantizer into LIC due to the fact that its gradients are zero almost everywhere with the exception of non-differentiable discontinuities at the quantization boundaries. Existing LIC models commonly adopt proxy functions for quantizers to enable end-to-end optimization via gradient backpropagation [5], [13], [38], [44], [79]. However, they inevitably suffer from the training-test gap caused by the divergence between the proxy functions and quantizers, and lead to evidently degraded R-D performance in comparison to fixed-rate compression [57], [60]. Existing variable-rate LIC models are usually inferior to VVC.

In this paper, we focus on variable-rate learned image compression with more sophisticated learnable non-uniform quantizers beyond widely adopted simple uniform quantizers.

It is nontrivial to learn the non-uniform quantizer along with nonlinear transforms for LIC, due to the challenging issue of non-smooth gradients of quantizers. Existing LIC models approximate the quantizer with techniques like additive uniform noise (AUN) and straight-through estimator (STE) but suffer from the training-test gap caused by unavoidable approximation error. Alternative methods employ uniform quantizers with predetermined quantization steps that cannot be optimized for LIC and lead to degraded R-D performance.

To overcome these problems, we propose a sampling-based optimization method that leverages inference-only computation for learned image compression to circumvent the non-smooth gradients for backpropagation. The proposed sampling-based optimization enables learning to refine the quantization steps of non-uniform quantizers for variable-rate coding with nonlinear transforms. Furthermore, we incorporate the proposed learnable non-uniform quantizer into variable-rate learned image compression and achieve state-of-the-art R-D performance with a guarantee of continuous and precise rate and distortion control. The contributions of this paper are summarized as below.

- We propose a sampling-based optimization method for nonlinear transform coding to avoid gradient-based optimization and optimize quantizers without backpropagation during end-to-end training of LIC models.
- We develop the dead-zone quantizer with learnable quantization steps based on the proposed sampling-based optimization to address the limitations of uniform quantizer in variable-rate learned image compression.
- We employ the learnable dead-zone quantizer in variable-rate learned image compression to simultaneously achieve state-of-the-art R-D performance and allow precise rate and distortion control<sup>1</sup>.

To our best knowledge, this paper is the first successful attempt to achieve variable-rate comporession with learnable dead-zone quantizer for nonlinear transform coding (NTC). It breaks through the constraints of non-smooth gradient for backpropagation in end-to-end learning. Contrary to prevailing uniform quantizers based on gradient-based optimization on proxy functions [5], [13], [38], [44], [79], the proposed method resorts to sampling-based optimization based on inference-only computation without gradient backpropagation. Different from existing methods that employ dead-zone quantizers with pre-defined quantization steps [21], [66], the proposed method learns the quantization steps and optimizes non-uniform quantization to improve R-D performance for variable-rate learned image compression with nonlinear transform coding (NTC).

Experimental results demonstrate that the proposed method enables variable-rate compression with precise rate control. It achieves state-of-the-art R-D performance in variable-rate image compression and outperforms most fixed-rate LIC models. Remarkably, the proposed method achieves an average 8.82% BD-rate reduction compared with the latest VVC standard. Furthermore, it can simultaneously achieve precise rate and

<sup>1</sup>In this paper, we use the informal term “distortion control” to describe rate control with target distortion (or reconstruction quality) for brevity.

distortion control with an average deviation within 0.0087 bpp and 0.1265 dB on the *Kodak* dataset.

The rest of this paper is organized as follows. Section II briefly reviews the related work. Section III demonstrates the sampling-based optimization method and Section IV presents the framework of optimizing quantization steps for dead-zone quantizer. Section V elaborates the variable-rate learned image compression model using the proposed sampling-based optimization. Section VI provides the experiments results. Finally, we draw conclusions in Section VII.

## II. RELATED WORK

This section overviews related work on variable-rate learned image compression (LIC) models and quantizers for LIC.

### A. Variable-Rate Learned Image Compression

Variable-rate coding is considered as a fundamental demand in developing learned image compression methods [1], [3]. Here, we review variable-rate image compression via deep learning techniques, including multi-scale decomposition, conditional transform, and adjustable quantization.

1) *Multi-Scale Decomposition*: Early works [1], [3], [19], [31] realized variable-rate coding by incrementally enhancing recovered image with residuals but rendered additional computational complexity due to the multi-pass iterative residual estimation through the encoder and decoder. To avoid extra computational complexity, one-pass encoder was adopted to achieve multi-scale decomposition over the latent representations and enable rate-determined representation. Rippel and Bourdev [4] designed binary entropy coding to progressively encode latent representations from the highest bit-plane to the lowest one. Cai *et al.* [32] and Nakanishi *et al.* [28] used multi-scale latent representations of deep convolutional neural networks to realize variable-rate representations. Yang *et al.* [20] employed a slimmable neural network adapt the representation and computational complexity to variable-rate coding.

2) *Conditional Transform*: In addition to explicit multi-scale decomposition, rate-determined latent representations can also be realized by conditional transform where the encoder and decoder depend on or are modulated by hyper-parameters [13]–[15], [62] or importance maps [35], [58]. Choi *et al.* [13] designed a conditional autoencoder by nonlinearly modulating the kernels of its convolutional and deconvolutional layers with the Lagrangian multiplier for rate-distortion trade-off. Yang *et al.* [14] modulated the convolved latent representations with approximate equivalence to [13], due to the linear convolution operations. Ballé *et al.* [15] enhanced the conditional encoder and decoder with modulated generative division normalization (GDN) layers and aggregated the nonlinearities with exponentially growing number of modulation parameters. Sun *et al.* [62] employed R-D optimization with interpolated Lagrangian multiplier for fine-grainularity rate adaptation. Qin *et al.* [56] adapted transformer-based image compression methods with visual prompt tuning techniques.

3) *Adjustable Quantization*: As a main component of transform coding, quantization has been explored for variable-rate learned image compression. Different from conditional transform, adaptive quantization performs scalar scaling [21]–[23], [57], [60] or channel-wise scaling [2], [12]–[18] on the latent representations only. Moreover, efforts on element-wise scaling shows the state-of-the-art R-D performance [81]. Channel-wise scaling has the potential to achieve competitive performance with independently learned fixed-rate models using a wider network [36] or an improved entropy model [12], while enabling continuous variable-rate coding [12] via interpolation. Scalar scaling provides monotonic R-D curves for versatile variable-rate coding but could lead to performance loss compared to fixed-rate models [57]. Recently, Arezki *et al.* [59] alternatively realized variable-rate image compression by uniformly quantizing the raw image.

It is worth mentioning that multi-scale decomposition and conditional transform build an implicit relationship between rate and distortion and resort to complex network structures for variable-rate coding. Channel-wise scaling works as a switch to activate different channels for varying bit-rates but is not universal to explicit rate-distortion modeling. Therefore, in this paper, we focus on optimizing scalar scaling for adaptive quantization to achieve precise rate control.

### B. Quantizers for Learned Image Compression

1) *Training Proxies for Quantizers*: Proxy functions have been widely studied to address zero-gradient quantizer in end-to-end optimization of LIC models by allowing gradient back-propagation and preventing vanishing gradients. Early works employed stochastic quantization [1]–[3] that randomly quantized inputs with a Bernoulli probability. Recent works commonly employ additive uniform noise (AUN) [5]–[7], [10], [11] and straight-through estimator (STE) [9], [69], [70], [72], [73] as proxies. They are equivalent to introducing an identity function for gradient back-propagation. In [79] and [44], AUN and STE were further combined with annealing algorithm to improve R-D performance. Agustsson *et al.* [38] achieved smooth approximation of the non-smooth quantization function to reduce the training-test gap. Specifically, dithered quantizers [12], [13] were leveraged to approximate the quantizer with a stochastic offset for variable-rate coding. Recently, Presta *et al.* [82] employed a differentiable proxy to achieve adjustable quantization for variable-rate compression. However, proxy functions inevitably cause the training-test gap on quantizers. On the contrary, the proposed sampling-based optimization employs forward-only processes to avoid training-test gap on quantizers.

2) *Non-uniform Quantizers*: Non-uniform quantizers have been utilized to enhance the approximation ability of LIC models. The soft quantizer [77], [78] realized vector quantization in the latent space to achieve more compressible representations. Zhou *et al.* [21] incorporated the dead-zone quantizer with the trained orthonormal transform in LIC. Lu *et al.* [64] introduced dead-zone quantizer in progressive image compression, and showed advantages of non-uniform quantizer in variable-rate compression. However, existing methods leverage dead-zone quantizers with pre-designed quantization steps, while

the proposed method optimizes the dead-zone quantizer with learnable quantization steps.

### III. SAMPLING-BASED OPTIMIZATION FOR NONLINEAR TRANSFORM CODING

In this section, we develop a sampling-based optimization framework for NTC models to address the zero-gradient problem in existing gradient-based optimization. We first formulate the rate-distortion (R-D) optimization problem for NTC, and then propose the sampling-based optimization framework along with a toy example for demonstration.

#### A. Problem Formulation

Consider an NTC model with a nonlinear analysis transform  $g_a(\cdot)$  and a nonlinear synthesis transform  $g_s(\cdot)$ . Given the input  $x$ , the latent representation is extracted by  $y = g_a(x)$  and is then quantized into  $\hat{y} = Q(y)$  using a quantizer  $Q(\cdot)$ . Supposing that  $\hat{y}$  follows a discrete distribution  $p_{\hat{y}}(\hat{y})$ , the coding budget of  $\hat{y}$  is  $-\log_2 p_{\hat{y}}(\hat{y})$ . The nonlinear synthesis transform  $g_s(\cdot)$  reconstructs the input from  $\hat{y}$ , *i.e.*,  $\hat{x} = g_s(\hat{y})$ . Let  $\theta$  denote the learnable parameters of the NTC model. The sampling-based optimization targets minimizing the average R-D loss on a given training set  $\mathcal{X} = \{x_i\}$ .

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta} \sim p_{\boldsymbol{\theta}}(\boldsymbol{\theta})} L_{\text{R-D}}(\boldsymbol{\theta}) \quad (1)$$

with the average R-D loss defined by

$$L_{\text{R-D}}(\boldsymbol{\theta}) = R + \lambda \cdot D \\ = \mathbb{E}_{\hat{y} \sim p(\hat{y}|x), x \in \mathcal{D}} [-\log_2 p_{\hat{y}}(\hat{y})] + \lambda \cdot \mathbb{E}_{x \in \mathcal{D}} [\|x - \hat{x}\|_2^2]. \quad (2)$$

Here,  $\lambda$  is the Lagrangian multiplier, and mean squared error (MSE) is adopted to measure the distortion  $x$  and  $\hat{x}$ . We further clarify the contribution of each individual components from the perspectives of dead-zone quantizer and sampling-based optimization as below.

*Dead-zone quantizer*: The dead-zone quantizer has been widely adopted in traditional compression methods (with linear transforms) to achieve a superior R-D performance in variable-rate compression. In [21], [64], [66], dead-zone quantizers are incorporated with nonlinear transforms (*i.e.*, the foundation of current learned image compression models) for learned image compression and are shown to improve the R-D performance of variable-rate compression. However, predefined dead-zone quantizers are adopted in [21], [64], [66], since it is challenging to optimize the dead-zone quantizer for nonlinear transform. In this paper, we present a novel method that successfully achieves learnable dead-zone quantizer in learned image compression.

*Sampling-based optimization*: We propose the sampling-based optimization for the non-differentiable quantizer. The proposed method originates from Diffuser [65], which is an offline reinforcement learning method that can achieve trajectory optimization. Different from Diffuser that employs a learned guidance, the proposed method uses an analytic guidance with R-D loss to accommodate to learned image compression. We have demonstrated that the proposed sampling-based optimization method achieves improved performance on both variable-rate and fixed-rate compression.

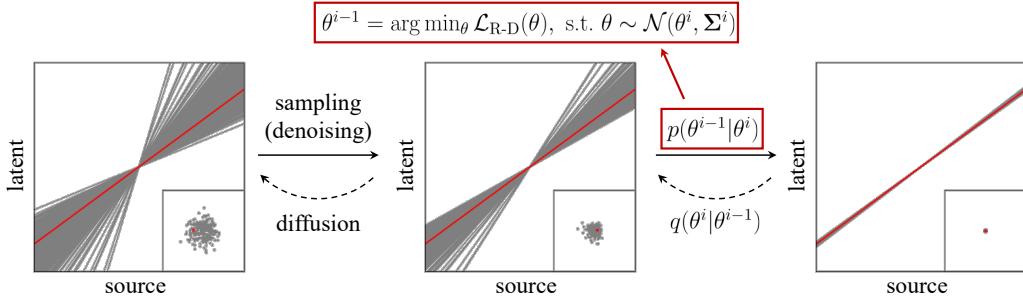


Fig. 1. Illustration of the proposed sampling-based optimization for nonlinear transform coding (NTC) models. The red curve is the optimal sample (analysis transform) with the minimum rate-distortion (R-D) loss in each sampling step, while the grey curves are other samples. The sampling process follows the ancestral sampling where the optimal sample in the previous step is adopted to initialize the samples in the current step. The distribution of the learnable parameters of the NTC model is visualized at the bottom right in each step.

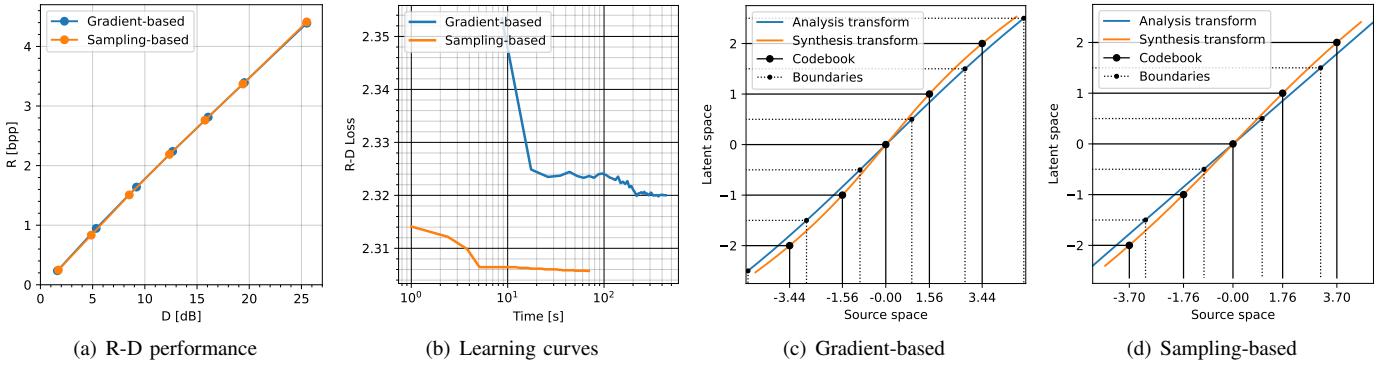


Fig. 2. Comparison between the gradient-based optimization and the proposed sampling-based optimization for a toy NTC model. (a) presents the R-D performance for compressing a standard Laplacian source with two optimization frameworks. The original NTC model reproduces those presented by Bellé *et al.* [15]. The BD-rate difference is less than 0.1%. (b) shows the learning curve of two optimization frameworks with identical nonlinear transforms (*i.e.*, a single GDN or  $i$ -GDN layer). (c) and (d) illustrate the learned nonlinear transforms of NTC and SamplingQ. (b), (c), and (d) are experimental results optimized with  $\lambda = 2.83$ .

### B. Sampling-based Optimization

Without loss of generality, we assume that the learnable parameters  $\theta$  are sampled from an unknown underlying distribution  $p_{\theta}(\theta)$ , *i.e.*,  $\theta \sim p_{\theta}(\theta)$ . We estimate the unknown distribution using  $T$  sequential sampling steps with each step corresponding to a conditional probability  $p(\theta_t | \theta_{t-1})$ . In the  $t$ -th step, the learnable parameters  $\theta_t$  follow a distribution that

$$p_{\theta_t}(\theta_t) = \prod_{i=1}^t p(\theta_i | \theta_{i-1}) \cdot p(\theta_0). \quad (3)$$

The initial learnable parameters  $\theta_0$  are commonly supposed to obey a Gaussian distribution. Thus, the conditional probability  $p(\theta_t | \theta_{t-1})$  equals a Gaussian distribution perturbed by an R-D indicator. For any  $t = 1, \dots, T$ ,

$$p(\theta_t | \theta_{t-1}) = \mathcal{N}(\theta_{t-1}, \Sigma_t) \cdot h(\theta_t), \quad (4)$$

where the R-D indicator  $h(\theta_t)$  selects the parameters achieving the minimum R-D loss:

$$h(\theta_t) = \begin{cases} 1, & \text{when } \theta = \arg \min_{\theta_t \sim p_{\theta_t}} L_{R-D}(\theta) \\ 0, & \text{otherwise} \end{cases}. \quad (5)$$

The proposed sampling-based optimization can be regarded as a variation of the recent diffusion-based trajectory optimizer Diffuser [65], where the posterior distributions in our work

are Gaussian distributions instead of implicit distributions established by neural networks, as explained in Section III-C.

### C. Perspective from Diffusion Process

Ballé *et al.* [15] show that nonlinear analysis and synthesis transforms in recent LIC models can be regarded as mappings that bridge the quantizers in the latent and source spaces. The quantizer in the latent space is commonly configured as a uniform quantizer, while the quantizer in the source space is optimized by learning nonlinear transforms via large-scale training. The nonlinear analysis transform determines the boundaries for quantization and the nonlinear synthesis transform connects codewords. These models are summarized in the scope of nonlinear transform coding (NTC) [15].

Let  $\mathcal{C}_{\text{latent}} = \{c'_i\}$  be the codebook consisting of codewords  $c'_i$  in the latent space, and  $\mathcal{C}_{\text{source}} = \{c_i\}$  the codebook of codeword  $c_i$  in the source space. Similarly, we denote  $\mathcal{B}_{\text{latent}} = \{b'_j\}$  and  $\mathcal{B}_{\text{source}} = \{b_j\}$  as quantization boundaries in the latent and source spaces, respectively. The mapping between the latent and source spaces can be represented by

$$g_a : b_j \rightarrow b'_j \text{ and } g_s : c'_i \rightarrow c_i. \quad (6)$$

Note that the quantization boundaries in high-dimensional spaces are hyperplanes. These hyperplanes partition the spaces

into cells and discretize the input within the cell to corresponding codeword  $c_i$ . The NTC model is optimized by minimizing the R-D loss presented in (2) via backpropagation. From the perspective of the entropy-constrained quantizer, it equals to the optimizing the codebook  $\mathcal{C}_{\text{source}}$  and boundaries  $\mathcal{B}_{\text{source}}$  for the quantizer in the source space. However, the quantizer  $Q(\cdot)$  in the latent space has a zero gradient almost everywhere. This results in vanished gradients and hinders end-to-end optimization. As an alternative, existing methods usually introduce a proxy function with a smooth gradient to approximate  $Q(\cdot)$  but suffer from training-test gap.

#### D. Toy Example: Compressing 1-D Laplacian Source

We exemplify the proposed sampling-based optimization for NTC with a toy example to show its efficiency. In the toy example, we optimize an NTC model for compressing the standard Laplacian source  $x \sim p_x$  with the probability density function (PDF) defined as

$$p_x(x) = \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|x|\right). \quad (7)$$

We develop a single-layer NTC model to compress the Laplacian source. The NTC model employs a single generalized divisive normalization (GDN) layer [26] for the analysis transform, and an inverse GDN ( $i$ -GDN) layer for the synthesis transform. The employed GDN and  $i$ -GDN layers have one-dimensional (1-D) input and output and are parameterized by two scalars, respectively. Specifically, the GDN layer for analysis transform works as a local normalization.

$$y = g_a(x) = \frac{x}{\sqrt{\beta_a + \gamma_a \cdot x^2}}. \quad (8)$$

The  $i$ -GDN layer for synthesis transform works inversely as

$$\hat{x} = g_s(\hat{y}) = \hat{y} \cdot \sqrt{\beta_s + \gamma_s \cdot \hat{y}^2}. \quad (9)$$

A uniform scalar quantizer  $Q(\cdot)$  with step size  $q = 1$  is adopted to quantize the 1-D latent representation  $y$  into  $\hat{y}$ . The R-D loss function for optimizing the learnable parameters  $\theta = \{\beta_a, \gamma_a, \beta_s, \gamma_s\}$  is

$$L_{\text{R-D}}^{\text{toy}}(\theta) = \mathbb{E}_{x \sim p_x} [-\log_2 p_{\hat{y}}(Q(g_a(x))) + \lambda \cdot \|x - \hat{x}\|_2^2], \quad (10)$$

Subsequently, we compare the proposed sampling-based optimization and the commonly adopted gradient-based optimization using back-propagation and proxy uniform noise. The configurations are elaborated below.

1) *Gradient-based Optimization*:  $p_{\hat{y}}(\hat{y})$  is modeled via a factorized entropy model. The additive uniform noise is introduced as the proxy of the quantizer. Adam optimizer with a learning rate of  $10^{-3}$  is adopted.

2) *Sampling-based Optimization*: We draw  $N = 5,000$  samples from the conditional distribution in each sampling step.  $p_{\hat{y}}$  is estimated statistically using the batch samples. The whole optimization process includes  $T = 50$  steps. In the  $t$ -th step, the scale  $\Sigma_t$  of the Gaussian distribution  $p(\theta_t | \theta_{t-1})$  is a diagonal matrix whose diagonal elements  $\sigma_t^{(i)}$ ,  $i = 1, \dots, N$  are defined by

$$\sigma_t^{(i)} = \left(1 - \frac{t+1}{T}\right) \theta_{t-1}^{(i)}, \quad (11)$$

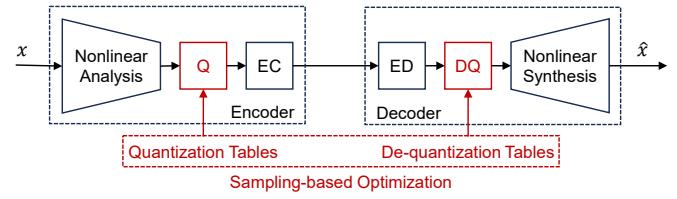


Fig. 3. Schematic illustration of the proposed variable-rate image coding framework with learnable dead-zone quantizers.

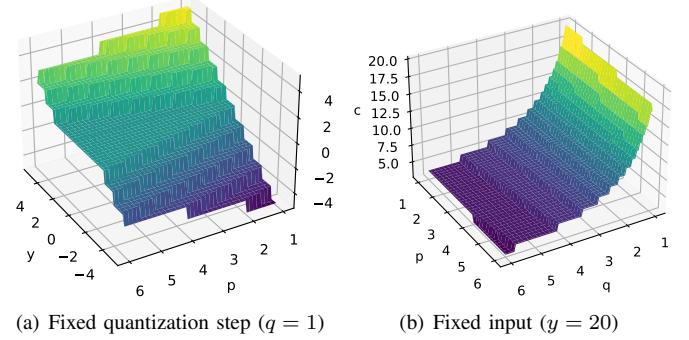


Fig. 4. Illustration of the index mapping process of a dead-zone quantizer with fixed quantization step or input, indicating that the gradients are almost zero everywhere.

where  $\theta_{t-1}^{(i)}$  is the  $i$ -th element in  $\theta_{t-1}$ .

The Lagrangian multiplier  $\lambda$  for optimizing the NTC models takes the values of 0.56, 1.26, 2.83, 6.35, 14.25, 32.00, and 128.00. All the experiments are implemented on a single NVIDIA GeForce RTX 4070Ti with 12GB memory. Fig. 2 shows that the proposed sampling-based optimization achieves faster convergence in training in terms of R-D performance.

## IV. VARIABLE-RATE IMAGE CODING WITH SAMPLING-BASED OPTIMIZATION

In this section, we elaborate the proposed variable-rate coding method realized by the sampling-based optimization to employ adjustable quantization and de-quantization tables, as presented in Fig. 3. The quantization tables obtained using the sampling-based optimization are compatible with non-uniform quantizers such as the dead-zone quantizer.

### A. Overview

The proposed method employs adjustable quantization and de-quantization tables to achieve variable-rate coding. The two tables are learned to control the granularity of quantization and optimize rate-distortion tradeoff for image compression. During the quantization process, the quantization table determines the index mapping that transforms the continuous coefficients to discrete indices, whereas the de-quantization table controls the reconstruction rule that recovers the coefficients from the discrete indices with an expected minimum reconstruction loss. Consequently, we elaborate the index mapping and reconstruction rule of uniform and dead-zone quantizers and introduce learnable parameters for dead-zone quantizers.

*1) Uniform Quantizers:* Uniform quantizers are prevailing in LIC. Index mapping  $C_q^{\text{UQ}}(\cdot)$  and reconstruction rule  $R_q^{\text{UQ}}(\cdot)$  are formulated as follows.

$$c = C_q^{\text{UQ}}(x) = \lceil x/q \rceil, \quad (12)$$

$$\hat{x} = R_q^{\text{UQ}}(c) = c \cdot q, \quad (13)$$

where  $\lceil \cdot \rceil$  returns the closest integer of the input and  $q$  is the quantization step. Note that, in LIC models,  $q$  is commonly set to 1 for the uniform quantizer and the reconstruction process of the quantizers is omitted.

*2) Dead-zone Quantizers:* For dead-zone quantizers, index mapping  $C_{p,q}^{\text{DZQ}}(\cdot)$  and reconstruction rule  $R_{p,q}^{\text{DZQ}}(\cdot)$  are

$$c = C_{p,q}^{\text{DZQ}}(x) = \text{sign}(x) \cdot \max \left( 0, \left\lfloor \frac{|x|}{q} - \frac{p}{2q} + 1 \right\rfloor \right), \quad (14)$$

$$\hat{x} = R_{p,q}^{\text{DZQ}}(c) = \begin{cases} 0, & c = 0 \\ \text{sign}(c) \left( |c| \cdot q + \frac{p-q}{2} \right), & c \neq 0 \end{cases}, \quad (15)$$

where  $p$  is the size of dead-zone and  $q$  the size of other bins. Note that the uniform quantizer is a special case of the dead-zone quantizer with  $p = q$ .

*3) Adjustable Quantization & De-quantization Tables:* Without loss of generality, we specify the quantization and de-quantization tables for dead-zone quantizers as four 3-D tensors for a 3-D latent representation  $\mathbf{y} \in \mathbb{R}^{H \times W \times C}$ :

- Quantization tables: size of dead-zone  $\mathbf{p}_a \in \mathbb{R}^{H \times W \times C}$  and size of other bins  $\mathbf{q}_a \in \mathbb{R}^{H \times W \times C}$  for obtaining the codeword  $\mathbf{c}$  from the latent representation  $\mathbf{y}$ ;
- De-quantization: size of dead-zone  $\mathbf{p}_s \in \mathbb{R}^{H \times W \times C}$  and size of other bins  $\mathbf{q}_s \in \mathbb{R}^{H \times W \times C}$  for recovering the quantized latent representation  $\hat{\mathbf{y}}$  from the codeword  $\mathbf{c}$ .

Therefore, the learnable parameters  $\theta = \{\mathbf{p}_a, \mathbf{q}_a, \mathbf{p}_s, \mathbf{q}_s\}$  are optimized by minimizing the R-D loss function  $L_{\text{R-D}}(\theta)$ .

$$L_{\text{R-D}}(\theta) = \mathbb{E}_{\mathbf{x} \in \mathcal{D}} \left[ -\log_2 p_{\hat{\mathbf{y}}}(\mathbf{Q}(\mathbf{g}_a(\mathbf{x}))) + \lambda \cdot \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \right]. \quad (16)$$

It should be noted that existing dead-zone quantizers in LIC models are manually designed before gradient-based optimization [21], [64], [66] without considering learnable parameters. Thus, they cannot be adjusted for diverse distributions. Different from existing dead-zone quantizers, we learn both quantization and de-quantization tables by addressing the non-trivial problem of optimizing (16). In fact, it is challenging to optimize the R-D loss function in (16) with gradient-based methods. As shown in Fig. 4, for a dead-zone quantizer, the gradients of index mapping are zero almost everywhere. In this paper, we propose sampling-based optimization rather than resort to gradient-based methods to solve the problem.

### B. Toy Example of Optimizing Dead-Zone Quantizers

We further employ a toy example on the standard Laplacian source to evaluate the efficiency of the proposed sampling-based optimization for dead-zone quantizers. In this toy example, we employ a pre-trained NTC model with the uniform quantizer as an anchor, and further substitute the uniform quantizer with a dead-zone quantizer and adjust the quantization steps to achieve variable-rate coding. We evaluate

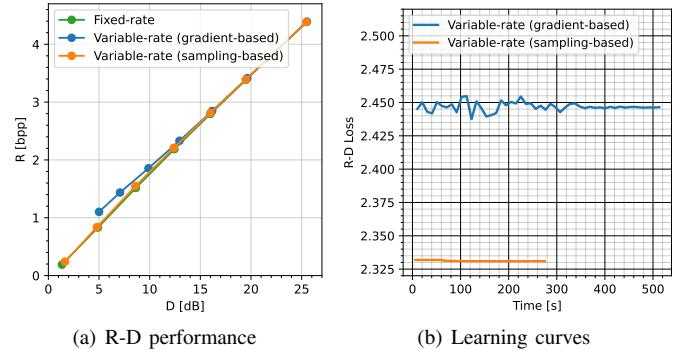


Fig. 5. R-D performance and learning curves of the variable-rate coding dead-zone quantizers optimized with gradient-based method (blue) and sampling-based method (orange).

gradient-based and sampling-based optimization methods under the same setting.

We develop a pre-trained NTC model by reproducing the fully connected neural network introduced by Ballé *et al.* [15] for nonlinear analysis and synthesis transforms. The nonlinear analysis transform contains fully connected layers and SoftPlus operators. In the experiments, the pre-trained model is optimized with  $\lambda = 128.00$ . We optimize the quantization steps of the dead-zone quantizer by minimizing the R-D loss function in (16) using Lagrangian multipliers  $\lambda = \{0.56, 1.26, 2.83, 6.35, 14.25, 32.00\}$ . Fig. 5 compares the R-D performance of the following variable-rate coding methods: a) independent fixed-rate NTC models (*i.e.*, fixed-rate); b) variable-rate coding model optimized using gradients (*i.e.*, gradient-based); c) variable-rate coding model optimized by sampling (*i.e.*, sampling-based). The gradient-based method performs worse than the proposed sampling-based optimization in optimizing the dead-zone quantizer, especially in the lower bit-rate regions. This result indicates that the gradient-based method with proxy function is inefficient. Moreover, Fig. 5 provides the learning curves of the commonly adopted gradient-based optimization and the proposed sampling-based method and shows a smoother and faster convergence of the sampling-based method.

We further provide supporting evidence for employing a sampling-based method for dead-zone quantizer optimization by studying the optimality of R-D optimization. Fig. 6 visualizes the R-D loss with respect to the quantization steps. The optimality of the R-D optimization problem is obvious. This fact implies that the proposed sampling-based optimization can converge to the global optimum with the conditional probability  $p(\theta_t | \theta_{t-1})$  presented in (3).

### C. Compacting Sampling Space for Image Compression

We further enhance the efficiency of sampling-based optimization by reducing the sampling space for learned image compression based on prior knowledge. Specifically, we consider nonlinear transforms, initialization of quantization steps, sharing quantization steps for latent representation, and entropy model, as elaborated below.

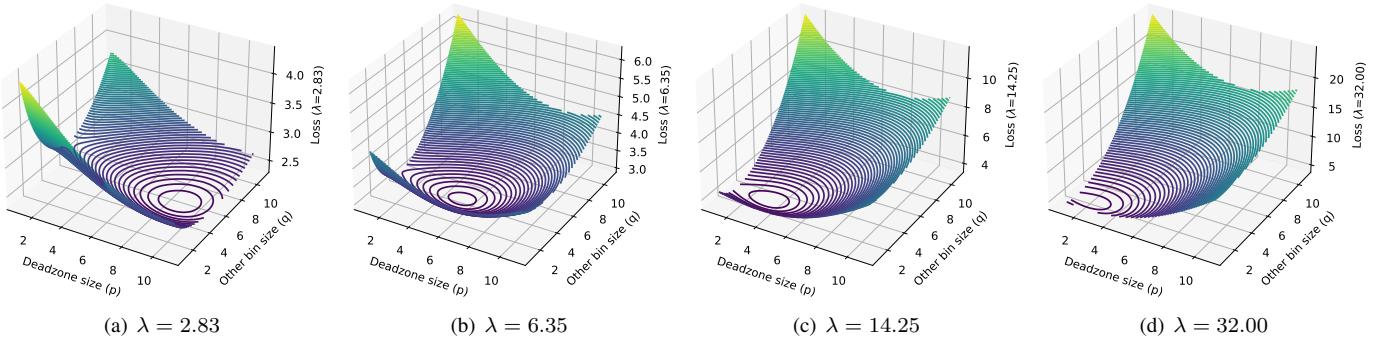


Fig. 6. Rate-distortion loss with respect to the dead-zone size  $p$  and other bin size  $q$  of the dead-zone quantizer combined with pretrained nonlinear transforms. The nonlinear transforms are from a pretrained nonlinear transform coding model for compressing a standard Laplacian source, with the Lagrangian multiplier  $\lambda$  set to 128.00.

1) *Nonlinear Transforms*: We employ nonlinear transforms of pre-trained LIC models to reduce the computational complexity. In this paper, we trained a variable-rate model to achieve a competitive R-D performance in a wide rate and distortion range.

2) *Initialization of Quantization Steps*: We employ a variable-rate model with a uniform quantizer of adjustable quantization steps as the pre-trained model for further optimization. The pre-trained LIC models are with uniform quantizers during training. We leverage the quantization learned in pre-trained models as the initial quantization steps for the size of dead-zone  $p$  and size of other bins  $q$ . A precise initialization can dramatically accelerate the convergence of the proposed sampling-based optimization.

3) *Cross-Element Sharing of Learned Quantization Steps*: We share the quantization steps across all elements of the latent representations. Moreover, we force the quantization steps in the quantization and de-quantization tables to be identical to further reduce the sampling space.

4) *Entropy Model*: This paper uses the pre-trained entropy model in the pre-trained LIC models to estimate the distribution of the quantized latent representations. Existing entropy models could be classified into hyperprior model and autoregression model. The hyperprior model offers unbiased estimation with a compressed prior, while the autoregression model provides conditional distribution based on the decoded elements. Therefore, there exists an R-D performance loss due to the involved autoregression model, since the estimated distribution is biased and results in coding overhead.

## V. PROPOSED MODEL FOR VARIABLE-RATE LEARNED IMAGE COMPRESSION

In this paper, we propose a variable-rate learned image compression model based on the learnable dead-zone quantizer via sampling-based optimization. The proposed model is constructed with two steps, *i.e.*, i) training a variable-rate model with uniform quantizer, and ii) optimizing the dead-zone quantizer. In step i), we first employ the learned image compression model based on frequency-aware transformer [68] as the base model for training a variable-rate model with the uniform quantizer using adjustable quantization steps. The variable-rate model with the uniform quantizer is trained via the multi-

objective optimization algorithm [63], similar to Kamisli *et al.* [57]. In step ii), we substitute the uniform quantizer with a dead-zone quantizer, and employ the proposed sampling-based optimization to refine the quantization steps for the dead-zone quantizer. Furthermore, we propose rate and distortion control algorithms to achieve precise and continuous adjustment of the quantization steps.

### A. Variable-Rate Coding Model

1) *Training Base Variable-rate Model*: We employ a training process similar to that of Kamisli *et al.* [57] to optimize the pre-trained model based on frequency-aware transformer [68]. The latent representation  $\mathbf{y} = g_a(\mathbf{x})$  is obtained from the input image  $\mathbf{x}$  using neural network-based encoder  $g_a$ . To achieve variable-rate compression, during encoding,  $\mathbf{y}$  is first scaled with a forward scaling factor  $s$  and then quantized with a uniform quantizer  $Q(\cdot) = \lfloor \cdot \rceil$ , *i.e.*,  $\hat{\mathbf{y}}_s = \lfloor s \cdot \mathbf{y} \rceil = \lfloor s \cdot \mathbf{y} \rceil$ , with  $\lfloor \cdot \rceil$  returning the closest integer for each coefficient of the input.  $\hat{\mathbf{y}}_s$  is used for lossless coding to transmit from the encoder to the decoder. When decoded, the image is recovered from the quantized latent representation via rescaling (with the inverse scaling factor  $s'$ ) and neural network-based decoder  $\hat{\mathbf{x}} = g_s(s' \cdot \hat{\mathbf{y}}_s)$  in a sequence. Note that the product of  $s$  and  $s'$  is maintained as a constant  $c$  for all achievable rates to enable continuous R-D modeling. Without loss of generality, we assume  $c = 1$  and  $s = 1/s'$  for simplicity.

We define the expectation of rate  $R$  and distortion  $D$  of the proposed model as the functions associated with  $s$ .

$$R(s) = \mathbb{E}_{P(\mathbf{x})P(\hat{\mathbf{y}}_s|\mathbf{x})} [-\log P(\hat{\mathbf{y}}_s)], \quad (17)$$

$$D(s) = \mathbb{E}_{P(\mathbf{x})} [\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2], \quad (18)$$

where  $P(\mathbf{x})$  is the underlying probability distribution of natural images and  $P(\hat{\mathbf{y}}_s) = \int P(\mathbf{x})P(\hat{\mathbf{y}}_s|\mathbf{x})d\mathbf{x}$  is the marginal distribution of quantized latent representation  $\hat{\mathbf{y}}_s$  using  $s$ . The parameters of base model  $\theta$  are optimized along with the scaling factor  $s$ . Thus, rate-distortion optimization for the proposed model on a single R-D point is formulated as

$$\min_{\theta, s} L(\theta, s) = \min_{\theta, s} R(\theta, s) + \lambda \cdot D(\theta, s), \quad (19)$$

where  $\lambda$  is the Lagrangian multiplier and  $\theta$  includes the learnable parameters of encoder, decoder, and entropy model.

To achieve variable-rate compression for diverse rate-distortion tradeoff with different Lagrangian multipliers, the proposed model employs a set of optimized scaling factors  $s_1, \dots, s_N$  under the given values of  $\lambda$  as the key points for interpolating continuous scale-rate and scale-distortion models. Let  $L_i(\boldsymbol{\theta}, s_i) = R(\boldsymbol{\theta}, s_i) + \lambda_i \cdot D(\boldsymbol{\theta}, s_i)$  the  $i$ -th R-D loss using  $s_i$ . The training process solves the multi-objective optimization problem with a vector of loss functions  $L_i(\boldsymbol{\theta}, s_i)$ ,  $i = 1, \dots, N$ .

$$\min_{\boldsymbol{\theta}, \{s_i\}_{i=1, \dots, N}} [L_1(\boldsymbol{\theta}, s_1), L_2(\boldsymbol{\theta}, s_2), \dots, L_N(\boldsymbol{\theta}, s_N)]. \quad (20)$$

We use recent multi-objective optimization algorithm [63] to solve (20). The training strategy is elaborated in Section VI.

*2) Transferring from Uniform Quantizer to Dead-Zone Quantizer:* We further replace the uniform quantizer with a dead-zone quantizer with the learnable quantization steps  $p$  and  $q$ . Instead of solving the multi-objective optimization problem to achieve a Pareto optimum directly, we optimize the quantization steps for each objective function with sampling-based optimization. For the  $i$ -th R-D loss function, we have

$$\min_{p_i, q_i} L_i(p_i, q_i). \quad (21)$$

To solve (21),  $p_i$  and  $q_i$  are initialized as  $s_i$ , and are then optimized with sequential sampling steps and the R-D indicator. When optimized, we fit a nonlinear projection between the scaling factor  $s$  and the quantization steps  $p$  and  $q$  for the dead-zone quantizer. Thus, rate and distortion control can be achieved using a single scaling factor  $s$ .

### B. Rate and Distortion Control

We employ a scaling factor  $s$  to adjust quantization steps for the uniform quantizer and compress the raw images with target rates. The encoder scales the latent representation  $\mathbf{y} \in \mathbb{R}^{H \times W \times C}$  by  $s$  and quantizes  $s \cdot \mathbf{y}$  into the codeword  $\mathbf{c} \in \mathbb{N}^{H \times W \times C}$  with the uniform quantizer, whereas the decoder obtains the recovered latent representation  $\hat{\mathbf{y}}$  from  $\mathbf{c}$  using the inverse scaling factor  $s' = 1/s$ .

$$\mathbf{c} = \lfloor s \cdot \mathbf{y} \rfloor \text{ and } \hat{\mathbf{y}} = \lfloor s' \cdot \mathbf{c} \rfloor. \quad (22)$$

The proposed method leverages the scalar scaling to yield monotonically decreasing distortion with the growth of rate. The relationship between the scaling factor  $s$  and rate  $r$  (or distortion  $d$ ) can be represented with empirical models. Different from previous works that employ  $\lambda$ -domain models [61], we establish the scale-rate and scale-distortion models to directly describe the relationship between the scaling factor and the target rate (and distortion).

*1) Scale-Rate Model:* We formulate the scale-rate model for the scaling factor  $s$  and rate  $r$  as a quadratic model characterized by three parameters, *i.e.*,  $\alpha$ ,  $\beta$ , and  $\gamma$ .

$$s = \alpha \cdot r^2 + \beta \cdot r + \gamma \quad (23)$$

Here, the rate  $r$  refers to the rate of latent representation without including hyperprior. In (23),  $\alpha$ ,  $\beta$ , and  $\gamma$  can be

estimated using three distinctive samples, *i.e.*, pairs of  $s$  and  $r$ . Given any  $(r_1, s_1)$ ,  $(r_2, s_2)$ , and  $(r_3, s_3)$ , we obtain that

$$\begin{aligned} \alpha &= \frac{r_3(r_2^2 - r_1^2) + r_2(r_1^2 - r_3^2) + r_1(r_3^2 - r_2^2)}{s_3(r_2^2 - r_1^2) + s_2(r_1^2 - r_3^2) + s_1(r_3^2 - r_2^2)}, \\ \beta &= \frac{(s_3 - s_2) - \alpha \cdot (r_3 - r_2)}{r_3^2 - r_2^2}, \\ \gamma &= s_3 - \alpha \cdot r_3 - \beta \cdot r_3^2. \end{aligned} \quad (24)$$

To reduce the number of samples required for estimating the parameters, we further propose a fast approximation for the quadratic model by employing a fixed linear relationship between  $\alpha$  and  $\beta$  (*i.e.*,  $\beta = k \cdot \alpha + t$ ) in (23).

$$s = \alpha \cdot r^2 + (k \cdot \alpha + t)r + \gamma. \quad (25)$$

According to (25), we can first estimate  $\alpha$  and  $\gamma$  using two pairs of  $s$ -rate samples (*i.e.*,  $(r_1, s_1)$  and  $(r_2, s_2)$ ) and then calculate  $\beta$  using the linear relationship.

$$\begin{aligned} \alpha &= \frac{(s_2 - s_1) - t(r_2 - r_1)}{(r_2 - r_1)(r_2 + r_1) + k(r_2 - r_1)} \\ \beta &= k \cdot \alpha + t \\ \gamma &= s_1 - \alpha \cdot r_1^2 - b \cdot r_1 \end{aligned} \quad (26)$$

*2) Scale-Distortion Model:* The relationship between the scaling factor  $s$  and the distortion  $d$  is formalized using three parameters  $\zeta$ ,  $\eta$ , and  $\iota$  as below.

$$s = \zeta \cdot d^{-1} + \eta \cdot d^{-2} + \iota. \quad (27)$$

Here, distortion refers to mean squared error (MSE) and we do not consider quality losses like multi-scale structural similarity (MS-SSIM) and perceptual losses like Fréchet inception distance (FID) and learned perceptual image patch similarity (LPIPS). Let  $(s_1, d_1)$ ,  $(s_2, d_2)$ , and  $(s_3, d_3)$  be the three pairs of  $s$  and  $d$  taken as the samples for estimating three parameters. We obtain that

$$\begin{aligned} \zeta &= \frac{d_3^{-1}(d_2^{-2} - d_1^{-2}) + d_2^{-1}(d_1^{-2} - d_3^{-2}) + d_1^{-1}(d_3^{-2} - d_2^{-2})}{s_3(d_2^{-2} - d_1^{-2}) + s_2(d_1^{-2} - d_3^{-2}) + s_1(d_3^{-2} - d_2^{-2})}, \\ \eta &= \frac{(s_3 - s_2) - \zeta \cdot (d_3^{-1} - d_2^{-1})}{d_3^{-2} - d_2^{-2}}, \\ \iota &= s_3 - \zeta \cdot d_3^{-1} - \eta \cdot d_3^{-2}. \end{aligned} \quad (28)$$

*3) Efficient Sample Extraction for Parameter Estimation:* To estimate the parameters of the scale-rate and scale-distortion models, we acquire the codewords  $\mathbf{c}$  of the latent representations  $\mathbf{y}$  with different quantization steps and corresponding probability  $p_c(\mathbf{c})$ . The analysis transform is conducted for only once in the proposed method, since  $\mathbf{y}$  is identical and only the scaling factor  $s$  changes for all the quantization processes. On the contrary, fixed-rate models and variable-rate models via conditional transforms have to perform multiple times of analysis transforms to search for rate adaptation. Thus, we just sample the scaling factors to fit the two models and the operations can be dramatically reduced. Algorithms 1 and 2 elaborate the proposed algorithms for rate and distortion control, respectively. Let us denote  $r_T$  as the target bit-rate and  $d_T$  the target distortion. Given the latent representation  $\mathbf{y}$  and scaling factors  $\mathcal{S} = \{s_1, s_2, s_3\}$ , we can

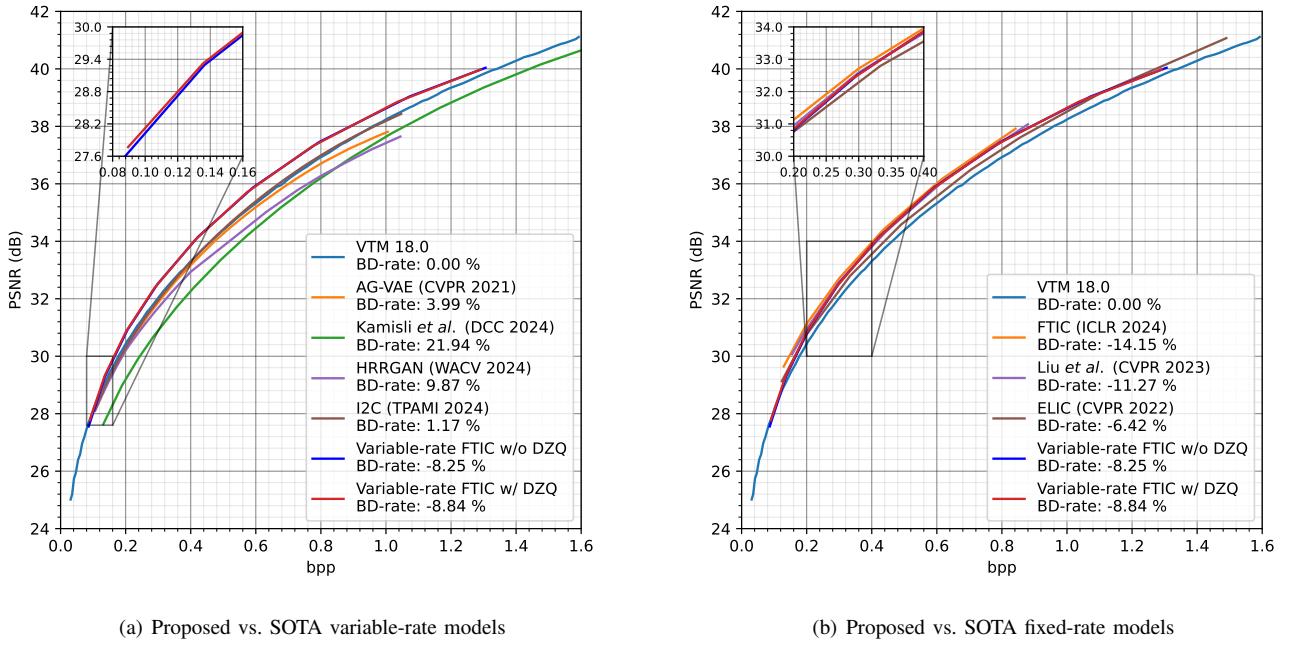


Fig. 7. Comparison of the proposed method with state-of-the-art learned image compression models over the *Kodak* dataset.

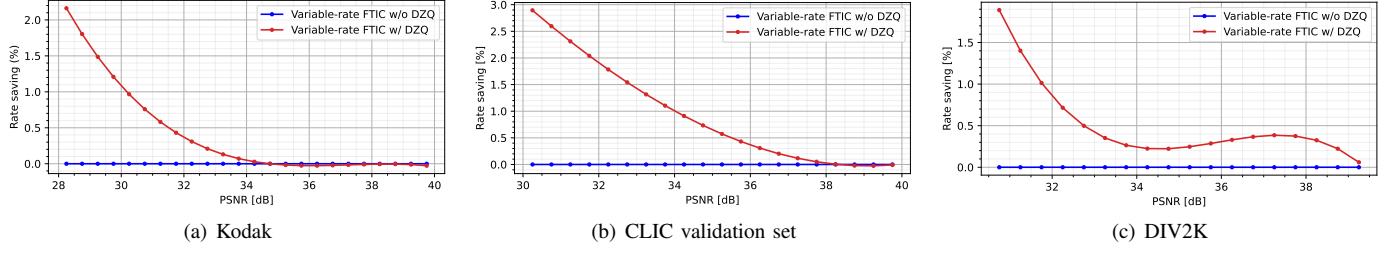


Fig. 8. Rate savings offered by the optimized dead-zone quantizer on *Kodak*, *CLIC validation*, and *DIV2K* datasets.

### Algorithm 1 Rate control algorithm.

**Input:** Target rate  $r_T$ , latent  $\mathbf{y}$ , index  $i \leftarrow 0$ , and scaling factors  $\mathcal{S} = \{s_1, s_2, s_3\}$ ;  
**Output:** Estimated scaling factor  $s^*$ ;

- 1: **for**  $s$  in  $\mathcal{S}$  **do**
- 2:    $i \leftarrow i + 1$
- 3:    $\mathbf{c}_i \leftarrow [\mathbf{y} \cdot s]$
- 4:    $p(\mathbf{c}_i) \leftarrow \text{EntropyModel}(\mathbf{y}, s)$
- 5:    $r_i \leftarrow -\log_2(p(\mathbf{c}_i))$
- 6: **end for**
- 7:  $\alpha, \beta, \gamma \leftarrow \text{Eq. (24)}$  with  $\{s_1, s_2, s_3\}, \{r_1, r_2, r_3\}$
- 8:  $s^* \leftarrow \text{Eq. (23)}$  with  $r_T$
- 9: **return**  $s^*$

estimate the parameters for the scale-rate and scale-distortion models using (24) and (28).

Different from conditional transforms and channel-wise or content-adaptive scaling, the proposed model achieves variable-rate compression via determined transforms and scalar scaling. This produces a continuous and monotonic R-D

### Algorithm 2 Distortion control algorithm.

**Input:** Target distortion  $d_T$ , latent  $\mathbf{y}$ , index  $i \leftarrow 0$ , and scaling factors  $\mathcal{S} = \{s_1, s_2, s_3\}$ ;  
**Output:** Estimated scaling factor  $s^*$ ;

- 1: **for**  $s$  in  $\mathcal{S}$  **do**
- 2:    $i \leftarrow i + 1$
- 3:    $\mathbf{c}_i \leftarrow [\mathbf{y} \cdot s]$
- 4:    $d_i \leftarrow g_s(\mathbf{c}_i / s)$
- 5: **end for**
- 6:  $\zeta, \eta, \iota \leftarrow \text{Eq. (28)}$  with  $\{s_1, s_2, s_3\}, \{d_1, d_2, d_3\}$
- 7:  $s^* \leftarrow \text{Eq. (27)}$  with  $d_T$
- 8: **return**  $s^*$

curve, with an explicit relationship between the quantization steps and rendered distortion. Consequently, precise rate and distortion control can be achieved with feasible models.

## VI. EXPERIMENTAL RESULTS

In this section, we perform extensive experiments to validate the proposed method. The base variable-rate LIC model (*i.e.*,

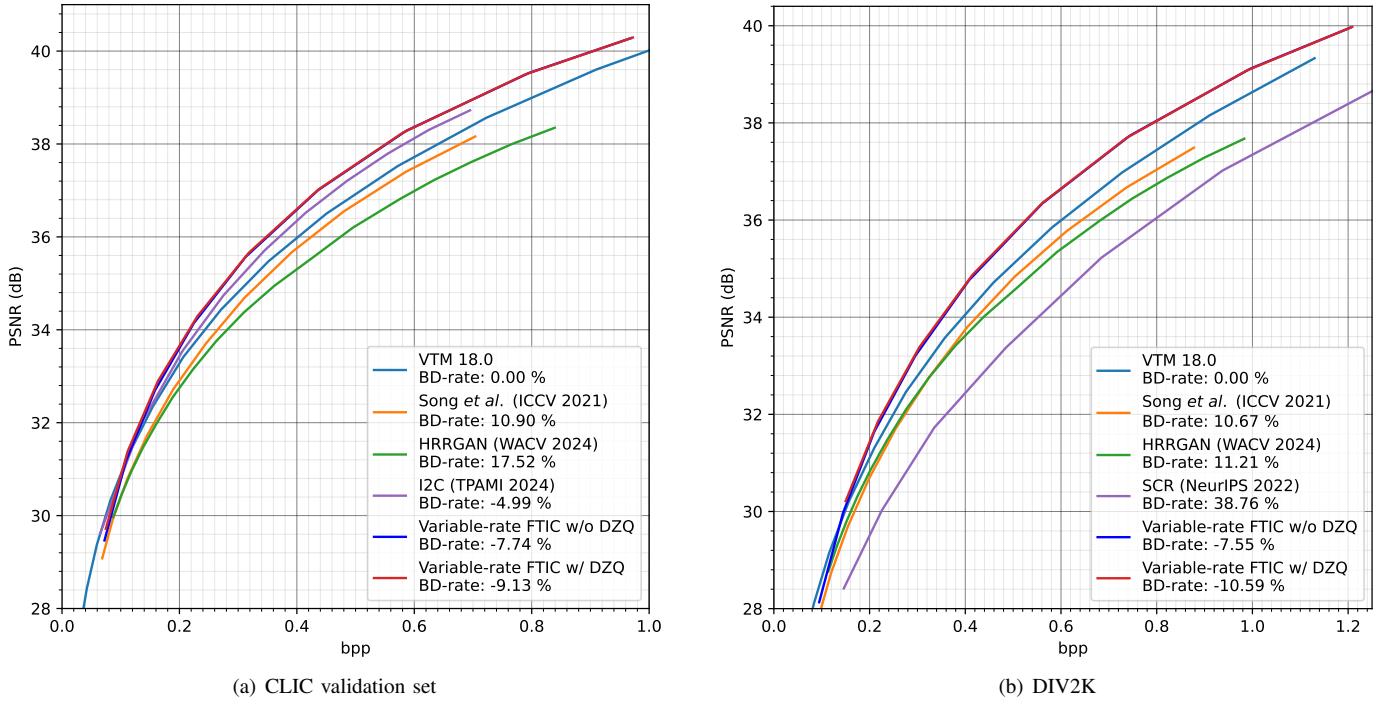


Fig. 9. R-D performance of the proposed method on *CLIC* validation set and *DIV2K*.

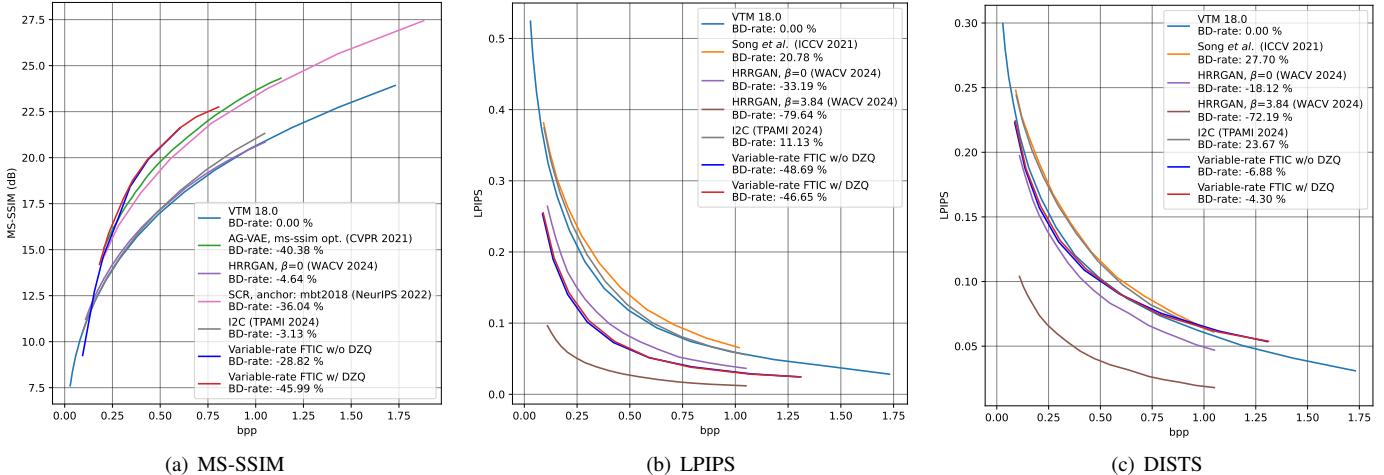


Fig. 10. R-D performance of the proposed method measured by MS-SSIM, LPIPS, and DISTS on the *Kodak* dataset.

variable-rate FTIC w/o DZQ) in the experiments is trained on the Flicker dataset [80], where the images are randomly cropped into  $256 \times 256$  patches without resizing. We optimize the variable-rate LIC model using fast adaptive multitask optimization (FAMO) algorithm [63]. The model is fine-tuned from a fixed-rate model optimized with  $\lambda = 0.0483$  for 100K steps with a batch size of 3, and validated every 2K steps. The learning rate is initialized as  $3 \times 10^{-5}$  and decayed to 30% with a patience of 5. The training process is implemented on a workstation with an NVIDIA Tesla V100 GPU of 32 GB memory. The Lagrangian multiplier  $\lambda$  for R-D optimization takes its value from 0.0009, 0.0018, 0.0035, 0.0067, 0.0130, 0.0250, 0.0483, 0.0932, and 0.1800, and the scaling factor  $s$  is correspondingly initialized as 0.10000, 0.1931, 0.2692,

0.3724, 0.5188, 0.7195, 1.0000, 1.3891, and 1.9305.

#### A. Rate-Distortion Performance

We evaluate the rate-distortion performance obtained by the proposed method, several state-of-the-arts variable-rate models [12], [35], [57], [58], [62], [67], [81], and fixed-rate models [68], [70], [72]. The results of other baseline methods are obtained from their official implementation or original papers, where all the models are optimized for MSE. Fig. 7(a) shows that the proposed method is superior to all the existing variable-rate image compression models (including variable-rate LIC models and VVC). Remarkably, it achieves an average BD-rate reduction of 8.82% in comparison to VTM 18.0, the test model of latest VVC. Furthermore, Fig. 7(b)

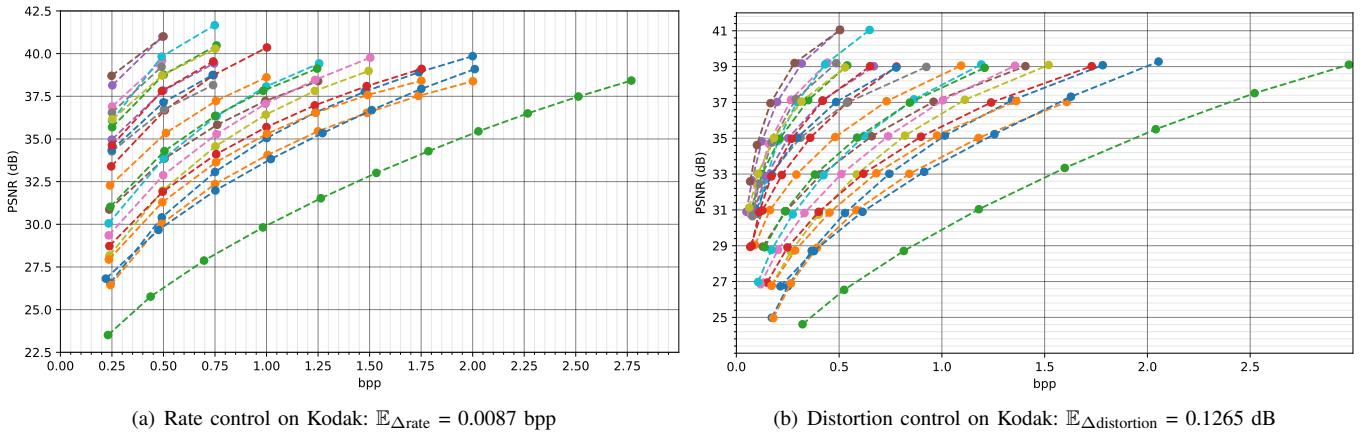


Fig. 11. Illustration of rate and distortion control of the proposed method on *Kodak* with given target rates or target distortions. In (a), the target rates are 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50, and 2.75 bpp. In (b), the target distortions are 25, 27, 29, 31, 33, 35, 37, 41 dB.

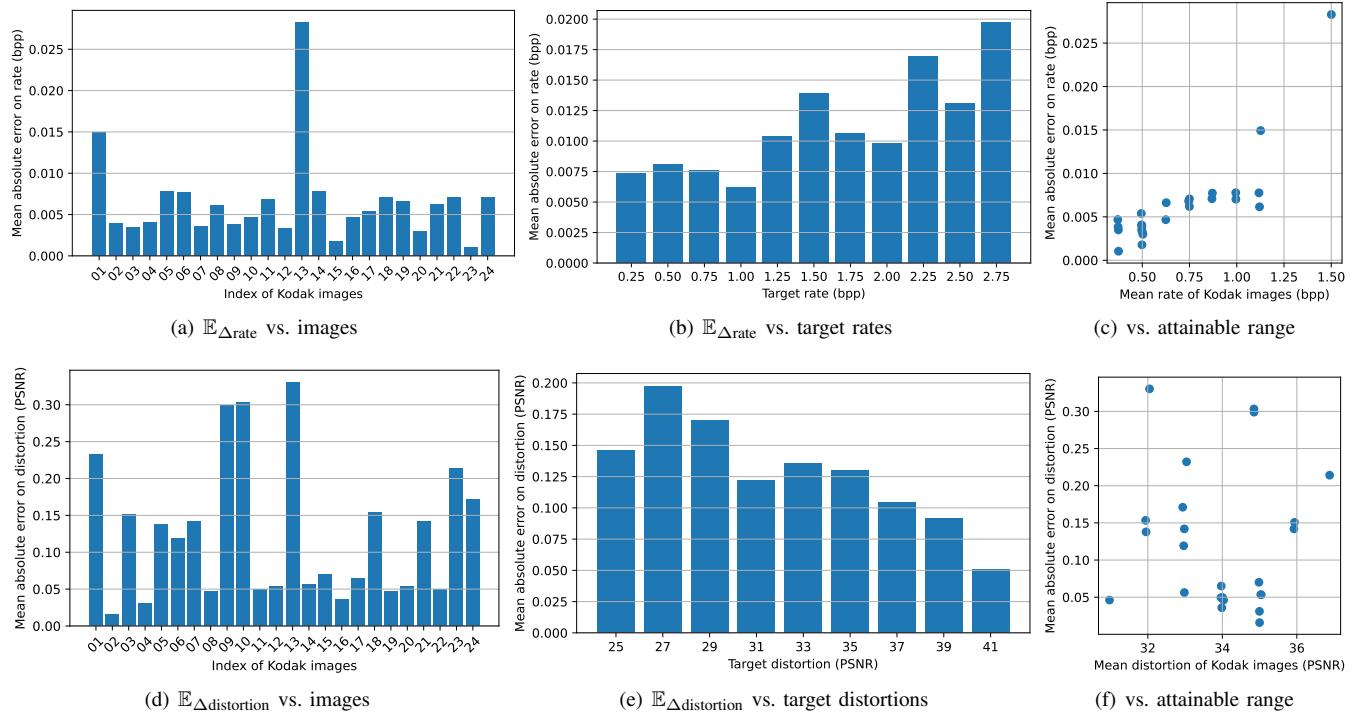


Fig. 12. Characteristics of rate and distortion control of the proposed method on different images and target rates and distortions.

shows that the proposed method outperforms most fixed-rate LIC models, while enabling variable-rate compression and enhanced rate adaptation.

We also evaluate the proposed method on two high-resolution image dataset, *i.e.*, the CLIC validation set and the DIV2K datasets. The CLIC validation set includes 41 images with resolutions up to  $1370 \times 2048$ , while the DIV2K dataset contains 100 images of 2K resolution. The experimental results are provided in Fig. 9, showing 9.13% BD-rate reduction on CLIC validation set and 10.59% BD-rate reduction on DIV2K dataset in comparison to VTM 18.0.

Fig. 8 presents more detailed rate savings of the optimized dead-zone quantizer (*i.e.*, variable-rate FTIC w/ DZQ) over the base model (*i.e.*, variable-rate FTIC w/o DZQ). The results

show that the proposed method achieves up to 2.16% rate saving on *Kodak*, 2.89% rate saving on CLIC validation set, and 1.89% rate saving on DIV2K dataset.

In addition to PSNR, we also evaluate the R-D performance in terms of MS-SSIM, LPIPS, and DISTs on *Kodak*. The results are presented in Fig. 10. We have employed a pre-trained fixed-rate FTIC [68] model optimized for MS-SSIM as the base model and employ the proposed sampling-based optimization to improve the performance. Experimental results in Fig. 10(a) suggest that the proposed method can further improve the R-D performance on MS-SSIM. As for LPIPS and DISTs, we use identical models presented in Fig. 7 instead of directly optimizing the models with LPIPS or DISTs, since the proposed method focus on variable-rate compression in

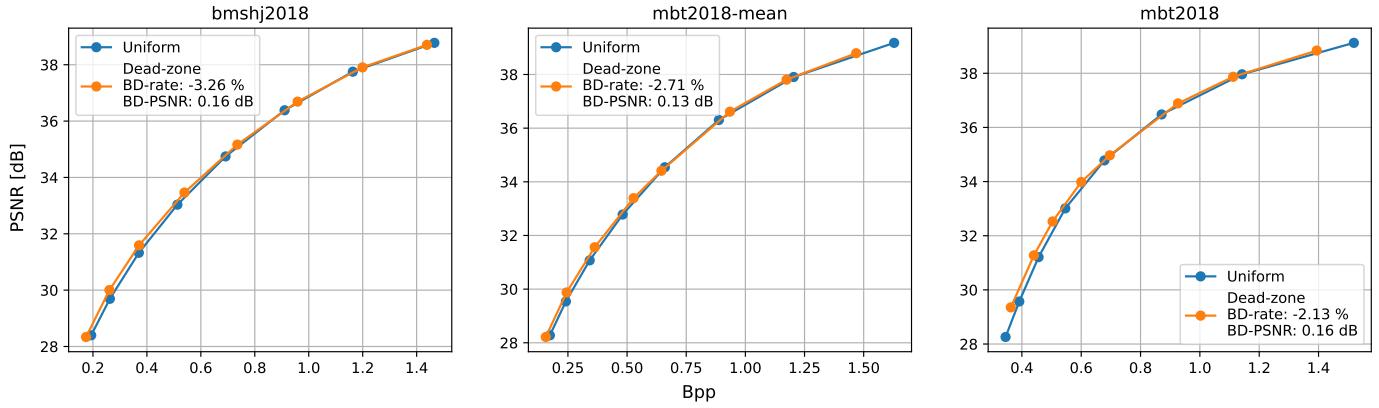


Fig. 13. Comparison on R-D performance of the proposed method over the baseline method with uniform quantizer on the *Kodak* dataset.

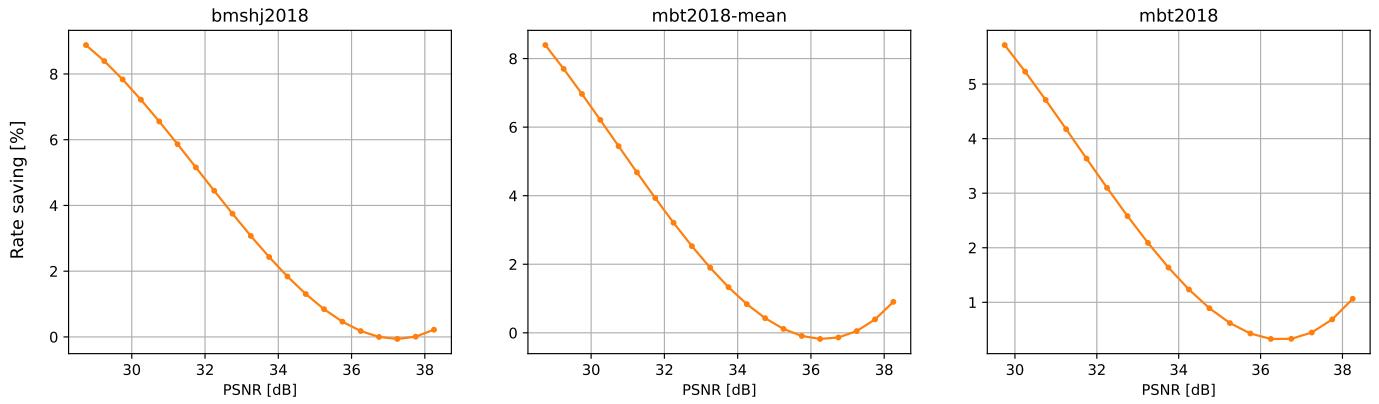


Fig. 14. Rate savings of the learnable dead-zone quantizer over the baseline models (*i.e.*, models with a scalar scaling strategy) on the *Kodak* dataset.

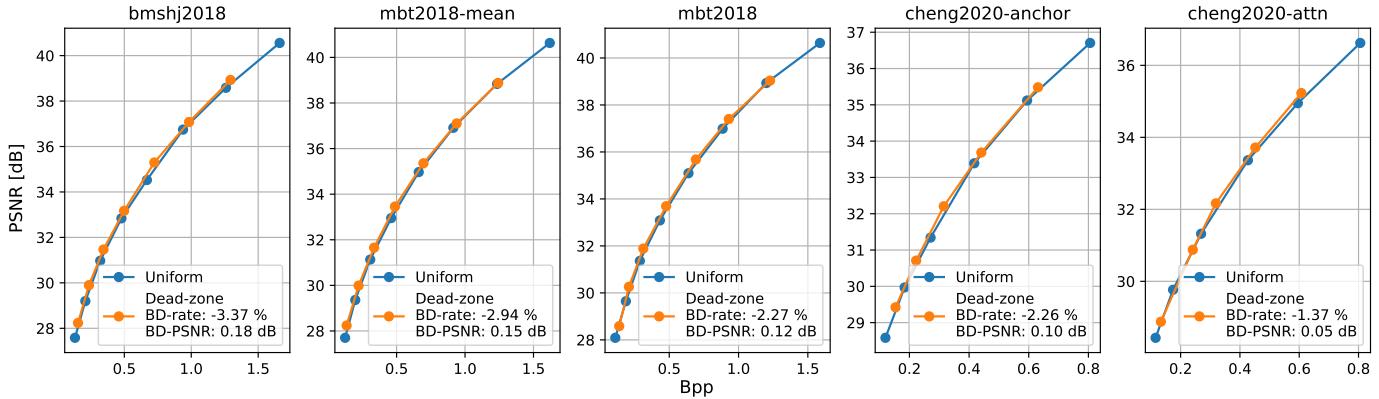


Fig. 15. Illustration on the R-D performance improvement of the sampling-based optimized dead-zone quantizers on the fixed-rate compression models on the *Kodak* dataset. The BD-rate reduction and BD-PSNR gain of the dead-zone quantizers over the original uniform ones are provided in the legends.

general scenarios. The results in Fig. 10(b) and Fig. 10(c) indicate that the variable-rate FTIC models achieve satisfactory performance and outperforms most compared models.

#### B. Performance of Rate and Distortion Control

Fig. 11 presents the performance of the proposed algorithms for rate and distortion control. The experiments are conducted on the *Kodak* dataset, where the scaling factor are limited within the range of pre-trained ones to prevent from unattainable target rate (or distortion). The proposed method achieves

an average error of 0.0087 bpp for rate control and an average error of 0.1265 dB for distortion control. Note that the fast implementation for rate control in (26) can also obtain an average error of 0.0278 bpp (which is not included in Fig. 8) when  $k$  and  $t$  are estimated on the *Kodak* dataset. The average rate error would increase to 0.0708 bpp when we use the *Tecnick* dataset for parameter estimation. This suggests that the fast rate approximation could be affected by the distribution shift, but it still works in the cases that aim to rapidly control the rate without strictly requiring a high precision.

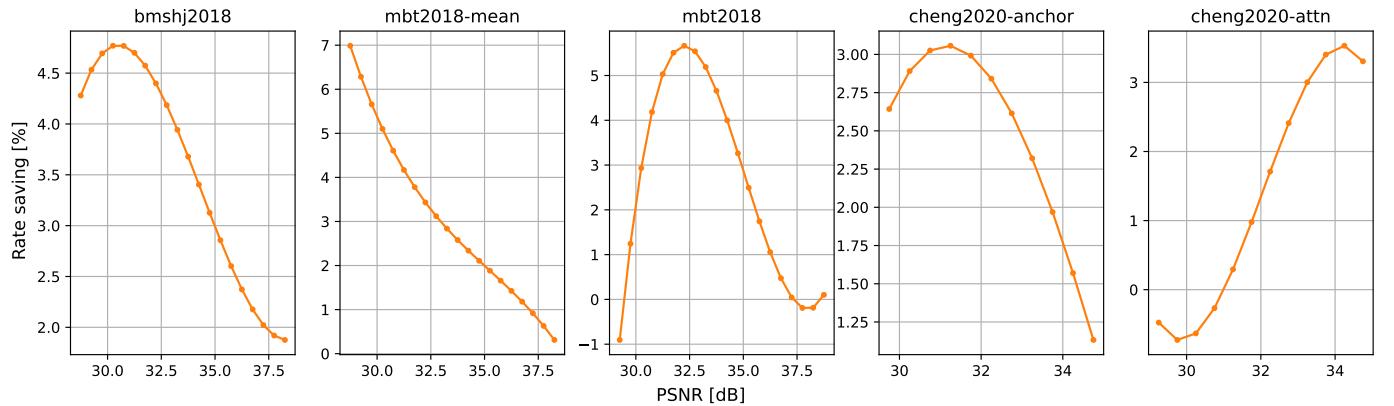


Fig. 16. Rate saving of the proposed method over the pre-trained fixed-rate models on the *Kodak* dataset.

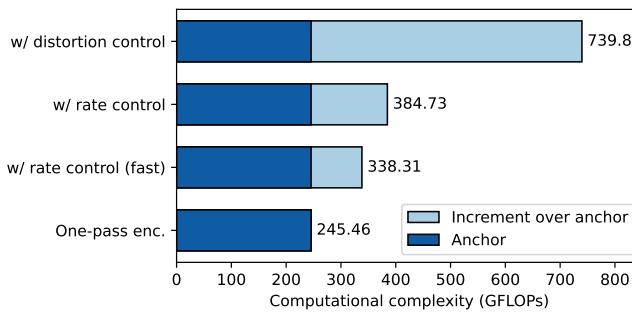


Fig. 17. Comparison of the proposed method with other SOTA learned image compression models over the *Kodak* dataset.

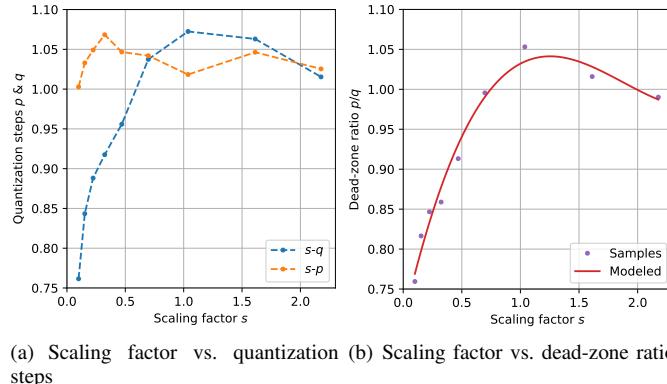


Fig. 18. Illustration on the characteristics of the quantization steps of the dead-zone quantizer in the proposed variable-rate image compression model.

Fig. 12 illustrates the characteristics of the proposed algorithms for rate and distortion control. Fig. 12 shows that the variation of rate control for most images can be limited in 0.010 bpp, and the variation of distortion control for most images can be limited within 0.15 dB, which shows the generalization of the proposed algorithms. Fig. 12(c) also indicates that the proposed algorithms render higher mean absolute error on those images with higher attainable rates (*i.e.*, hard-to-compress images). This fact indicates that the scale-rate models for the hard-to-compress images might be different from those for easy-to-compress ones. It is potential

to design an image-specific scale-rate model to improve the quadratic model for all images.

### C. Complexity of Rate and Distortion Control

Fig. 17 compares the computational complexity of the proposed rate-control algorithm with the one-pass encoding on a  $256 \times 256$  image. The rate-control algorithm with target rate renders about 66.2% additional computational complexity over original one-pass encoding, and the rate control algorithm with target distortion yields 217.4% increment on computational complexity. The results indicate that the proposed method can achieve reduced computational complexity, compared to multiple-pass encode required by fixed-rate models and conditional transform-based variable-rate models.

### D. Quantization Steps of Dead-Zone Quantizers

Fig. 18 illustrates the characteristics of dead-zone quantizers for different target rate and distortion. Fig. 18(a) presents the quantization steps of the dead-zone quantizers obtained by solving (21) with the sampling-based optimization. Fig. 18(b) presents that the relationship between the dead-zone ratio  $p/q$  and the scaling factor  $s$  can be approximately modeled using a third-order polynomial. The dead-zone ratio is shown to change for different R-D points, especially in the lower-rate regions. This result indicates the sub-optimality of the uniform quantizer for the variable-rate compression model.

### E. Additional Validations on Sampling-based Optimization

To show the potential of the proposed sampling-based optimization and learnable dead-zone quantizer, we further conduct experiments on fixed-rate and variable-rate models. Fig. 13 and Fig. 14 show the proposed method achieves R-D performance gains on the variable-rate models with three base models, *i.e.*, scale-only hyperprior [6], mean-scale hyperprior [7], and spatial-autoregression [7]. Fig. 15 and Fig. 16 demonstrate the R-D performance gain of the proposed method on the fixed-rate models with five different base models, *i.e.*, scale-only hyperprior [6], mean-scale hyperprior [7], spatial-autoregression [7], and two base models provided by Cheng *et al.* [10]. These results show that the proposed method achieve

an average 2%–3% bit-rate reduction compared to the models with uniform quantizers. Remarkably, the proposed method yields more evident gains in the lower rate regions, *i.e.*, up to 8% bit-rate savings of for the variable-rate models and 7% for the fixed-rate models. Moreover, the proposed method can be applied in the other learning-based visual coding scenarios (such as, coding for machines and point cloud coding) by adapting the proposed method to the specific scenarios of varying downstream tasks or irregular distribution in the 3-D space.

## VII. CONCLUSIONS

In this paper, we propose a sampling-based optimization method for variable-rate learned image compression to prevent gradient approximation error in gradient-based optimization methods. We apply the proposed sampling-based optimization method to optimize the quantization steps of dead-zone quantizer, and obtain improved rate-distortion performance for variable-rate compression using NTC models. Furthermore, the sampling-based optimization of non-uniform dead-zone quantizer can be incorporated to achieve precise rate and distortion control. Experiments show that the proposed model achieves state-of-the-art R-D performance in variable-rate image compression. In future, we will explore to reduce the parameters of nonlinear transforms for LIC models to extend the applications of the proposed sampling-based optimization method. Convergence properties could be also established for the sampling-based optimization to further guarantee the R-D performance in theory.

## REFERENCES

- [1] G. Toderici *et al.*, “Variable rate image compression with recurrent neural networks,” in *4th Int. Conf. Learn. Rep. (ICLR)*, 2016.
- [2] L. Theis, W. Shi, A. Cunningham, and F. Huszár, “Lossy image compression with compressive autoencoders,” in *5th Int. Conf. Learn. Rep. (ICLR)*, 2017.
- [3] G. Toderici *et al.*, “Full resolution image compression with recurrent neural networks,” in *2017 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 5306–5314.
- [4] O. Rippel and L. Bourdev, “Real-time adaptive image compression,” in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 2922–2930.
- [5] J. Ballé, V. Laparra, and E. Simoncelli, “End-to-end optimized image compression,” in *5th Int. Conf. Learn. Rep. (ICLR)*, 2017.
- [6] J. Ballé, D. Minnen, S. Singh, S. Hwang, and N. Johnston, “Variational image compression with a scale hyperprior,” in *6th Int. Conf. Learn. Rep. (ICLR)*, 2018.
- [7] D. Minnen, J. Ballé, and G. Toderici, “Joint autoregressive and hierarchical priors for learned image compression,” in *Adv. Neural Inf. Process. Syst. 31*, 2018, pp. 10771–11780.
- [8] J. Lee, S. Cho, and S.-K. Beack, “Context-adaptive entropy model for end-to-end optimized image compression,” in *7th Int. Conf. Learn. Rep. (ICLR)*, 2019.
- [9] D. Minnen and S. Singh, “Channel-wise autoregressive entropy models for learned image compression,” in *2020 IEEE Int. Conf. Image Process. (ICIP)*, 2020, pp. 3339–3343.
- [10] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, “Learned image compression with discretized Gaussian mixture likelihoods and attention modules,” in *2020 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 7939–7948.
- [11] T. Chen, H. Liu, Z. Ma, Q. Shen, X. Cao, and Y. Wang, “End-to-end learnt image compression via non-local attention optimization and improved context modeling,” *IEEE Trans. Image Process.*, vol. 30, pp. 3179–3191, 2021.
- [12] Z. Cui *et al.*, “Asymmetric gained deep image compression with continuous rate adaptation,” in *2021 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 10532–10541.
- [13] Y. Choi, M. El-Khamy, and J. Lee, “Variable rate deep image compression with a conditional autoencoder,” in *2019 IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 3146–3154.
- [14] F. Yang, L. Herranz, J. van de Weijer, J. A. I. Gutiérrez, A. M. López, and M. G. Mozerov, “Variable rate deep image compression with modulated autoencoder,” *IEEE Signal Process. Lett.*, vol. 27, pp. 331–335, 2020.
- [15] J. Ballé *et al.*, “Nonlinear transform coding,” *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 2, pp. 339–353, 2021.
- [16] T. Chen and Z. Ma, “Variable bitrate image compression with quality scaling factors,” in *2020 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2020, pp. 2163–2167.
- [17] Y. Gao, Y. Wu, Z. Guo, Z. Zhang, and Z. Chen, “Perceptual friendly variable rate image compression,” in *2021 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPR-W)*, 2021, pp. 1916–1920.
- [18] M. Akbari, J. Liang, J. Han, and C. Tu, “Learned multi-resolution variable-rate image compression with octave-based residual blocks,” *IEEE Trans. Multimedia*, vol. 23, pp. 3013–3021, 2021.
- [19] R. Su, Z. Cheng, H. Sun, and J. Katto, “Scalable learned image compression with a recurrent neural networks-based hyperprior,” in *2020 IEEE Int. Conf. Image Process. (ICIP)*, 2020, pp. 3369–3373.
- [20] F. Yang, L. Herranz, Y. Cheng, and M. G. Mozerov, “Slimmable compressive autoencoders for practical neural image compression,” in *2021 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPR-W)*, 2021, pp. 4998–5007.
- [21] J. Zhou, A. Nakagawa, K. Kato, S. Wen, K. Kazui, and Z. Tan, “Variable rate image compression method with dead-zone quantizer,” in *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPR-W)*, 2020, pp. 624–628.
- [22] H. Ma, D. Liu, N. Yan, H. Li, and F. Wu, “End-to-end optimized versatile image compression with wavelet-like transform,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 3, pp. 1247–1263, 2022.
- [23] T. Dumas, A. Roumy and C. Guillemot, “Autoencoder based image compression: can the learning be quantization independent?” in *2018 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2018, pp. 1188–1192.
- [24] D. Taubman, “High performance scalable image compression with EBCOT,” *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1158–1170, 2000.
- [25] D. He, Y. Zheng, B. Sun, Y. Wang, and H. Qin, “Checkerboard context model for efficient learned image compression,” in *2021 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 14771–14780.
- [26] J. Ballé, “Efficient nonlinear transforms for lossy image compression,” in *Picture Coding Symp. (PCS)*, 2018, pp. 248–252.
- [27] 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*, [Online]. Available: <https://arxiv.org/abs/2011.03029>.
- [28] K. M. Nakanishi, S. Maeda, T. Miyato, and D. Okanohara, “Neural multi-scale image compression,” in *Asian Conf. Comput. Vis. (ACCV)*, 2018, pp. 718–732.
- [29] C. Aytekin *et al.*, “Block-optimized variable bit rate neural image compression,” in *2018 IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPR-W)*, 2018, pp. 2551–2554.
- [30] Z. Zhang, Z. Chen, J. Lin, and W. Li, “Learned scalable image compression with bidirectional context disentanglement network,” in *2019 IEEE Int. Conf. Multimedia Expo (ICME)*, 2019, pp. 1438–1443.
- [31] C. Jia, Z. Liu, Y. Wang, S. Ma, and W. Gao, “Layered image compression using scalable auto-encoder,” in *2019 IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, 2019, pp. 431–436.
- [32] C. Cai, L. Chen, X. Zhang, and Z. Gao, “Efficient variable rate image compression with multi-scale decomposition network,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 12, pp. 3687–3700, 2019.
- [33] M. Akbari, J. Liang, J. Han, and C. Tu, “Learned Variable-Rate Image Compression With Residual Divisive Normalization,” in *2020 IEEE Int. Conf. Multimedia Expo (ICME)*, 2020, pp. 1–6.
- [34] C. Han, Y. Duan, X. Tao, M. Xu, and J. Lu, “Toward Variable-Rate Generative Compression by Reducing the Channel Redundancy,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 7, pp. 1789–1802, 2020.
- [35] M. Song, J. Choi, and B. Han, “Variable-rate deep image compression through spatially-adaptive feature transform,” in *2021 IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 2380–2389.
- [36] A. Dosovitskiy and J. Djolonga, “You only train once: Loss-conditional training of deep networks,” in *8th Int. Conf. Learn. Rep. (ICLR)*, 2020.
- [37] C. Jia, Z. Ge, S. Wang, S. Ma, and W. Gao, “Rate distortion characteristic modeling for neural image compression,” in *2022 Data Compression Conf. (DCC)*, 2022, pp. 202–211.

- [38] E. Agustsson and L. Theis, "Universally quantized neural compression," in *Adv. Neural Inf. Process. Syst.* 33, 2020, pp. 12 367–12 376.
- [39] Y.-H. Ho, C.-C. Chan, W.-H. Peng, and H.-M. Hang, "End-to-end learned image compression with augmented normalizing flows," in *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPR-W)*, 2021, pp. 1931–1935.
- [40] L. Helminger, A. Djelouah, M. Gross, and C. Schroers, "Lossy image compression with normalizing flows," in *Int. Conf. Learn. Rep. Workshops*, 2021.
- [41] Y. Xie, L.-C. Ka, and Q. Chen, "Enhanced invertible encoding for learned image compression," in *Proc. 29th ACM Int. Conf. Multimedia*, 2021, pp. 162–170.
- [42] Y. Wang, M. Xiao, C. Liu, S. Zheng, and T.-Y. Liu, "Modeling lost information in lossy image compression," *arXiv preprint arXiv:2006.11999*, 2020. [Online]. Available: <https://arxiv.org/abs/2006.11999>.
- [43] K. Tsubota and K. Aizawa, "Comprehensive comparisons of uniform quantizers for deep image compression," in *2021 IEEE Int. Conf. Image Process. (ICIP)*, 2021, pp. 2089–2093.
- [44] Z. Guo, Z. Zhang, R. Feng, and Z. Chen, "Soft then hard: Rethinking the quantization in neural image compression," in *Proc. 38th Int. Conf. Mach. Learn.*, 2021, pp. 3920–3929.
- [45] B. Bross et al., "Overview of the Versatile Video Coding (VVC) standard and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3736–3764, 2021.
- [46] R. Franzen, "Kodak lossless true color image suite," 2013. [Online]. Available: <http://r0k.us/graphics/kodak>
- [47] N. Asuni and A. Giachetti, "TESTIMAGES: A large-scale archive for testing visual devices and basic image processing algorithms (SAMPLING 1200 RGB set)," in *STAG: Smart Tools Apps Graphics*, 2014. [Online]. Available: [https://sourceforge.net/projects/testimages/files/OLD\\_OLD\\_SAMPLING/testimages.zip](https://sourceforge.net/projects/testimages/files/OLD_OLD_SAMPLING/testimages.zip)
- [48] J. Ziv, "On universal quantization," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 344–347, 1985.
- [49] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.*, vol. 4, no. 3, pp. 247–269, 1998.
- [50] O. Sener and V. Koltun, "Multi-task learning as multi-objective optimization," in *Adv. Neural Inf. Process. Syst.* 31, 2018, pp. 525–536.
- [51] S. Bhadane, A. B. Wagner, and J. Acharya, "Principal bit analysis: Autoencoding with schur-concave Loss," in *Proc. 38th Int. Conf. Mach. Learn.*, 2021, pp. 852–862.
- [52] S. Li, W. Dai, Z. Zheng, C. Li, J. Zou, and H. Xiong, "Reversible autoencoder: A CNN-based nonlinear lifting scheme for image reconstruction," *IEEE Trans. Signal Process.*, vol. 69, pp. 3117–3131, 2021.
- [53] A. N. Gomez, M. Ren, R. Urtasun, and R. B. Grosse, "The reversible residual network: Backpropagation without storing activations," in *Adv. Neural Inf. Process. Syst.* 30, 2017, pp. 2214–2224.
- [54] L. Dinh, D. Krueger, and Y. Bengio, "NICE: Non-linear independent components estimation," in *Int. Conf. Learn. Rep. Workshops*, 2015.
- [55] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd Int. Conf. Learn. Rep. (ICLR)*, 2015.
- [56] S. Qin et al., "Progressive learning with visual prompt tuning for variable-rate image compression," in *2024 IEEE Int. Conf. Image Process. (ICIP)*, 2024, pp. 1767–1773.
- [57] F. Kamislı, F. Racapé, and H. Choi, "Variable-rate learned image compression with multi-objective optimization and quantization-reconstruction offsets," in *2024 Data Compression Conf. (DCC)*, 2024, pp. 193–202.
- [58] S. Cai et al., "I2C: Invertible continuous codec for high-fidelity variable-rate image compression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 6, pp. 4262–4279, 2024.
- [59] B. Arezki, F. Feng, and A. Mokraoui, "Universal end-to-end neural network for lossy image compression," in *EUSIPCO Eur. Conf. Signal Process.*, 2024, pp. 626–630.
- [60] K. Tong, Y. Wu, Y. Li, K. Zhang, L. Zhang, and X. Jin, "QVRF: A quantization-error-aware variable rate framework for learned image compression," in *2023 IEEE Int. Conf. Image Process. (ICIP)*, 2023, pp. 1310–1314.
- [61] P. Jia et al., "Bit rate matching algorithm optimization in JPEG-AI verification model," in *Picture Coding Symp. (PCS)*, 2024, pp. 1–5.
- [62] Z. Sun, et al., "Interpolation variable rate image compression," in *Proc. 29th ACM Int. Conf. Multimedia*, 2021, pp. 5574–5582.
- [63] B. Liu, Y. Feng, P. Stone, and Q. Liu, "FAMO: Fast adaptive multitask optimization," in *Adv. Neural Inf. Process. Syst.* 36, 2023, pp. 57 226–57 243.
- [64] Y. Lu, Y. Zhu, Y. Yang, A. Said and T. S. Cohen, "Progressive neural image compression with nested quantization and latent ordering," in *2021 IEEE Int. Conf. Image Process. (ICIP)*, 2021, pp. 539–543.
- [65] M. Janner, Y. Du, J. Tenenbaum, and S. Levine, "Planning with diffusion for flexible behavior synthesis," in *Proc. 39th Int. Conf. Machine Learn.*, 2022, pp. 9902–9915.
- [66] S. Li, H. Li, W. Dai, C. Li, J. Zou, and H. Xiong, "Learned progressive image compression with dead-zone quantizers," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 6, pp. 2962–2978, 2023.
- [67] S. Iwai, T. Miyazaki, and S. Omachi, "Controlling rate, distortion, and realism: Towards a single comprehensive neural image compression model," in *IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2024, pp. 2900–2909.
- [68] H. Li, S. Li, W. Dai, C. Li, J. Zou, and H. Xiong, "Frequency-aware transformer for learned image compression," in *12th Int. Conf. Learn. Rep. (ICLR)*, Vienna, Austria, May 2024.
- [69] W. Jiang, J. Yang, Y. Zhai, F. Gao, and R. Wang, "MLIC++: Linear complexity multi-reference entropy modeling for learned image compression," in *Int. Conf. Machine Learn. Workshops*, Honolulu, HI, USA, Jul. 2023.
- [70] 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 *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 5718–5727.
- [71] R. Zou, C. Song, and Z. Zhang, "The devil is in the details: Window-based attention for image compression," in *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 17471–17480.
- [72] J. Liu, H. Sun, and J. Katto, "Learned image compression with mixed Transformer-CNN architectures," in *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Vancouver, Canada, Jun. 2023, pp. 14388–14397.
- [73] Y. Zhu, Y. Yang, and T. Cohen, "Transformer-based transform coding," in *10th Int. Conf. Learn. Rep. (ICLR)*, Virtual Event, Apr. 2022.
- [74] H. Fu, F. Liang, J. Liang, B. Li, G. Zhang, and J. Han, "Asymmetric learned image compression with multi-scale residual block, importance scaling, and post-quantization filtering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 8, pp. 4309–4321, Aug. 2023.
- [75] A. B. Koyuncu, P. Jia, A. Boev, E. Alshina, and E. Steinbach, "Efficient contextformer: Spatio-channel window attention for fast context modeling in learned image compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 8, pp. 7498–7511, Aug. 2024.
- [76] C. Li, S. Yin, C. Jia, F. Meng, Y. Tian, and Y. Liang, "Multirate progressive entropy model for learned image compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 8, pp. 7725–7741, Aug. 2024.
- [77] E. Agustsson et al., "Soft-to-hard vector quantization for end-to-end learning compressible representations," in *Adv. Neural Inf. Process. Syst.* 30, Long Beach, CA, USA, Dec. 2017, pp. 1141–1151.
- [78] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. Van Gool, "Conditional probability models for deep image compression," in *2018 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, Jun. 2018, pp. 4394–4402.
- [79] Y. Yang, R. Bamler, and S. Mandt, "Improving inference for neural image compression," in *Adv. Neural Inf. Process. Syst.* 33, Virtual, Dec. 2020, pp. 573–584.
- [80] J. Liu, G. Lu, Z. Hu, and D. Xu, "A unified end-to-end framework for efficient deep image compression," *arXiv preprint arXiv:2002.03370*, 2020. [Online]. Available: <https://arxiv.org/abs/2002.03370>.
- [81] J. Lee, S. Jeong, and M. Kim, "Selective compression learning of latent representations for variable-rate image compression," in *Adv. Neural Inf. Process. Syst.* 35, New Orleans, LA, USA, Nov. 2022, pp. 13146–13157.
- [82] A. Presta, E. Tartaglione, A. Fiadrotti, and M. Grangetto, "STanH: Parametric quantization for variable rate learned image compression," *IEEE Trans. Image Process.*, vol. 34, pp. 639–651, Jan. 2025.



**Shaohui Li** (Member, IEEE) received the B.Eng. degree in electronic engineering from Dalian University of Technology, Dalian, China, in 2016. He received the M.Eng. degree and the Ph.D. degree in information and communication engineering from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2019 and 2022, respectively. He is currently a post-doc at Shenzhen International Graduate School, Tsinghua University. His research interests include image reconstruction and image/video coding.



**Wenrui Dai** (Member, IEEE) received B.S., M.S., and Ph.D. degree in Electronic Engineering from Shanghai Jiao Tong University, Shanghai, China in 2005, 2008, and 2014. He is currently a full professor at the Department of Computer Science and Engineering, Shanghai Jiao Tong University (SJTU). From 2019 to 2024, he was an associate professor with Department of Computer Science and Engineering, Shanghai Jiao Tong University (SJTU). Before joining SJTU, he was with the faculty of the University of Texas Health Science Center at Houston from 2018 to 2019. He was a postdoctoral scholar with the Department of Biomedical Informatics, University of California, San Diego from 2015 to 2018 and a postdoctoral scholar with the Department of Computer Science and Engineering, SJTU from 2014 to 2015. His research interests include multimedia signal processing, image and video coding, and machine learning. He was the co-author of a Top 5% Paper in IEEE ICIP 2024 and a Top Paper Award in ACM Multimedia 2022.



**Junni Zou** (Member, IEEE) received the M.S. degree and the Ph.D. degree in communication and information system from Shanghai University, Shanghai, China, in 2004 and 2006, respectively. From March 2006 to Jan. 2017, she was with the School of Communication and Information Engineering, Shanghai University, Shanghai, where she is a full Professor. Since Feb. 2017, she has been a full Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, China. From June 2011 to June 2012, she was

with the Department of Electrical and Computer Engineering, University of California, San Diego (UCSD), as a visiting Professor.

Her research interests include multimedia networking and communications, machine learning and network resource optimization. She has published over 100 international journal/conference papers on these topics. She was the co-author of a Top 10% Paper Award in VCIP 2016. Dr. Zou was granted National Science Fund for Outstanding Young Scholar in 2016. She was the recipient of Shanghai Young Rising-Star Scientist in 2011. She is the associate editor of Digital Signal Processing.



**Nuowen Kan** (Member, IEEE) received the B.S. degree in information engineering from the Nanjing University of Aeronautics and Astronautics, China, in 2017, the M.S. degree in electronics and communications engineering from Shanghai Jiao Tong University (SJTU), China, in 2020, and the Ph.D. degree from computer science from SJTU in 2024. Now he is a postdoc with the Department of Electronic Engineering, SJTU. His research interests include multimedia communications, and theories and applications of reinforcement learning and graph learning in multimedia communication systems. He was awarded the ACM Multimedia Top Paper Award in 2022.



**Hongkai Xiong** (Fellow, IEEE) is a Cheung Kong Professor and a Distinguished Professor at Shanghai Jiao Tong University (SJTU). Currently, he is with both Dept. Electronic Engineering and Dept. Computer Science and Engineering. From 2018-2020, he acted as the Vice Dean of Zhiyuan College in SJTU. He received the Ph.D. degree from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2003. Since then, he has been with the Department of Electronic Engineering, SJTU. He was an Associate Professor from 2005-2011 and an Assistant Professor from 2003-2005. From 2007 to 2008, he was a Research Scholar in the Department of Electrical and Computer Engineering, Carnegie Mellon University (CMU), Pittsburgh, PA, USA. During 2011-2012, he was a Scientist with the Division of Biomedical Informatics at the University of California (UCSD), San Diego, CA, USA.

Dr. Xiong's research interests include signal representation and wavelet analysis, image and video coding, multimedia communication and networking, computer vision, and machine learning. He published over 360 referred journal and conference papers. He was the co-author of the TOP 1% Paper Award of the 2022 ACM International Conference on Multimedia (ACM Multimedia'22), the Top 5% Award at the 2024 IEEE International Conference on Image Processing (IEEE ICIP'24), the Top 10% Paper Award at the 2016 IEEE Visual Communication and Image Processing (IEEE VCIP'16), the Best Student Paper Award at the 2014 IEEE Visual Communication and Image Processing (IEEE VCIP'14), the Best Paper Award at the 2013 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (IEEE BMSB'13), and the Top 10% Paper Award at the 2011 IEEE International Workshop on Multimedia Signal Processing (IEEE MMSP'11).

In 2022, Dr. Xiong was granted the First Prize of the Shanghai Science and Technology Progress Award. In 2021, Dr. Xiong was awarded the First Prize of Natural Science of the Chinese Institute of Electronics. In 2018, he was the recipient of the Shanghai Youth Science and Technology Distinguished Accomplishment Award. In 2017, he received both the Science and Technology Innovative Leader Talent Award, and Shanghai Academic Research Leader Talent Award. In 2014, he received both National Science Fund for Distinguished Young Scholar Award from Natural Science Foundation of China (NSFC) and Shanghai Youth Science and Technology Talent Award. In 2013, he obtained Shanghai Shu Guang Scholar Award. In particular, he was given the recipients of Baosteel Excellent Faculty Award in both 2021 and 2017, and the SJTU SMC-A Excellent Young Faculty Awards in both 2010 and 2013. In both 2017 and 2011, he has been granted the First Prizes of the Shanghai Technology Innovation Awards. In 2009, he obtained the New Century Excellent Talent Award from Ministry of Education of China. From 2018-2021, he was the Associate Editor of IEEE Transactions on Circuits and Systems for Video Technology (TCSVT). He was elevated to a Fellow of the Chinese Institute of Electronics in 2023. He is a Fellow of the IEEE.



**Chenglin Li** (Member, IEEE) received the B.S., M.S. and Ph.D. degrees in Electronic Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2007, 2009 and 2015, respectively. From 2015 to 2017, he was a Postdoctoral Researcher with the Signal Processing Laboratory (LTS4) at École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. From 2017 to 2018, he was a Senior Researcher with the Chair of Media Technology (LMT) at Technical University of Munich (TUM), Germany, supported by the Hildegard Maier Research Fellowship of the Alexander von Humboldt Foundation for postdoctoral researchers.

He is currently a Full Professor with the Department of Electronic Engineering at Shanghai Jiao Tong University. His current research interests include multimedia signal processing and communications, adaptive video streaming, and theories and applications of reinforcement learning and federated learning in multimedia communication systems. He was awarded the Microsoft Research Asia (MSRA) Fellow in 2011, the Alexander von Humboldt Research Fellow in 2017, and the Top Paper Award of ACM Multimedia in 2022. He is the Associate Editor of the IEEE Transactions on Signal and Information Processing over Networks, the Youth Editor of the Chinese Journal of Electronics, and the Area Chair of 2022 ACM International Conference on Multimedia.