Rate-controllable Learned Image Compression Using Channel Attention

Zhiwei Xie XIEZW@XS.USTB.EDU.CN
Wenyu Zhang[*] WYZHANG@USTB.EDU.CN
Hua Shao SHAOHUA@USTB.EDU.CN
Xianze Yang YANGXZ@XS.USTB.EDU.CN
Xiao Zhang ZHANGXIAO@XS.USTB.EDU.CN
University of Science and Technology Beijing, Beijing, 10083, China
[*]*Corresponding author*

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

Classical learned image compression (LIC) methods usually require training multiple models to achieve the best compression performances at different rates, which greatly increases their training and deployment cost. Though existing methods can realize rate variation by using channel scaling factors or transform of the Lagrange multiplier, they are not able to adaptively control the compression process with desired rates, which causes additional trial cost if we want to obtain results with given compression ratios. In this paper, we address this issue by employing channel attention modules that use the desired target bit-rate as side information to adjust the distributions of feature channels, and a new rate-distortion loss function that integrates the target bit-rate into the rate-distortion optimization framework is proposed to train the model to realize continuous rate control. Additionally, a two-stage training strategy is utilized to ensure that the network can adaptively adjust the bit-rates, at the same time achieving the best rate-distortion performance. Experimental results demonstrate that our method achieves effective rate control over a wide range of bit-per-pixels (BPPs).

Keywords: Learned image compression, Channel attention, Rate-distortion optimization.

1. Introduction

Image compression has been a long-standing and prominent topic with the goal of achieving efficient storage and transmission of high-quality image data. Conventional image compression methods, such as JPEG (Wallace, 1992), JPEG2000 (Rabbani and Joshi, 2002), and BPG (Bellard, 2014), generally utilize modular architectures incorporating transformation, quantization, and entropy coding to realize the compression and reconstruction of images. The three modules are designed and optimized in an independent way, which limits the potential in improving the compression capability since global optimal cannot be achieved.

Learned image compression (LIC) employs an end-to-end learning architecture to achieve joint rate-distortion optimization by minimizing $R + \lambda D$, where R denotes the bit rate, D represents the distortion of image reconstruction, and λ is the Lagrange multiplier controlling the trade-off between R and D. Most LIC methods are realized by CNN or Transformer networks under the framework of variational auto-encoder (VAE) architecture, and recent advances have shown that the compression capabilities of LIC models can be improved by using advanced network structures (Liu et al., 2023), attention mechanisms (Cheng et al., 2020; Zou et al., 2022), precise entropy models

(Qian et al., 2022; Jiang et al., 2023), and generative learning methods (Relic et al., 2024; Jia et al., 2024). However, all of them necessitate training multiple models with different λ values to cover different compression ratios, which requires substantial computational and storage resources.

To realize rates variation in LIC, Song et al. (2021) leverages quality maps as conditional inputs to achieve different spatial feature transformation. Lee et al. (2022) and Li et al. (2025) adjust bit rate by controlling the amount of transmitted information. Though these methods can implicitly realize compression with variable or continuous rates, they can not achieve controllable compression with desired rates.

In this paper, our goal is realizing rate control in LIC, and it has not been considered in existing work. We propose to use channel attention modules to control the distributions to realize LIC with controllable compression rate. More specifically, to perceive the desired target bit-rates, we add channel attention modules into the network, and the target bit-rate is used as the side information of each channel attention module. We employ a two-stage training strategy to realize rate control at the same time guaranteeing the R-D performance. In the first stage, a compression model without attention is trained with the classical rate-distortion loss that minimizes $R+\lambda D$. In the second stage, we propose a new rate-distortion loss function to fine-tune the model, and rate part is modified as a rate control part that minimizes the differences between the real rates and desired rates. We evaluate the performances of the proposed approach on the Kodak dataset, experimental results show that the proposed method can realized effective rate-controllable with one LIC model without causing evident performance loss compared with the model without rate control mechanism.

2. Related Work

2.1. Learned Image Compression

In recent years, LIC has been emerged as a new way to realize the transform process, and recent advances have demonstrated that LIC can achieve stronger compression performances compared with conventional methods. Ballé et al. (2016) pioneered an end-to-end learned image compression framework based on Variational Auto-Encoder (VAE), achieving global R-D joint optimization. Subsequently, Ballé et al. (2018) enhanced the framework by introducing a hyperprior model to capture spatial dependencies in latent representations and improving the entropy model for enhanced performance. Minnen et al. (2018) proposed an autoregressive model into the existing hyperprior framework to achieve more precise modeling and estimation of the probability distribution of latent representations. Following these foundational works, numerous extensions have emerged to further enhance the compression capabilities of LIC models. Qian et al. (2022) and Jiang et al. (2023) focused on refining the entropy model to reduce redundancy. Liu et al. (2023) utilized Transformer-based network architectures to extract more compact and efficient latent representations. However, these approaches typically require training multiple models to achieve optimal compression performance at different bit rates.

2.2. Variable Rate Compression

In order to enhance the flexibility in the rate variation of LIC, several methods have been proposed. Song et al. (2021) used variable importance mapping features corresponding to λ as side information to joint training with latent features. Cui et al. (2021) model the inter-channel importance of latent representations, leveraging different channel gains to adjust quantization losses. Lee et al.

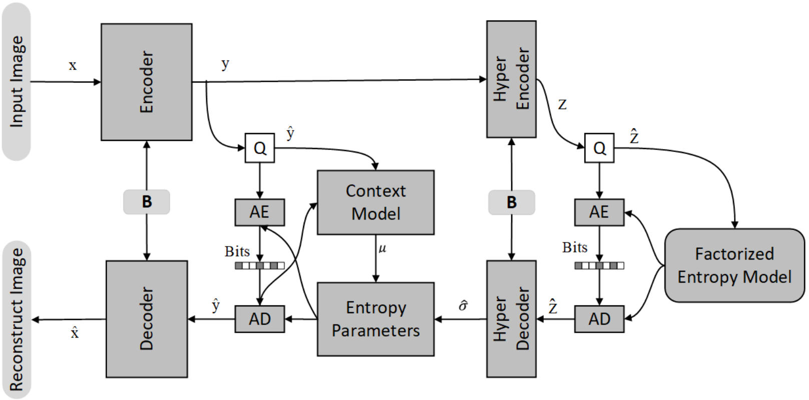

Figure 1: The proposed Rate-controllable LIC framework, in which the target desired bit-rate B is used as the side information of the encoders and decoders to guide the model to realize rate-controllable LIC process.

(2022) employed the Trit-Planes (DPICT) algorithm to decompose image features into multiple trit planes. By regulating the number of transmission iterations per plane, the overall bitrate can be adaptively varied. Li et al. (2025) designed a scalable encoder–decoder architecture coupled with a heterogeneous, scalable entropy model to support flexible bitrate conversion. All of the above methods indirectly influence the compression rate through feature scaling and clipping, but do not enable control of bit-rates as desired, which limits the flexibility of LIC.

2.3. Channel Attention

In this paper, we use channel attention modules to realize rate-controllable LIC. Channel attention mechanism (Hu et al., 2018) was first proposed to explicitly model channel-wise dependencies, it first employs global average pooling to condense spatial information into latent channel-wise features which followed by two fully connected layers and a Sigmoid activation function to learn the interdependencies between channels and generate channel-specific weights. Then, an attention mechanism is applied along the channel dimension to scale the latent features by the corresponding weight coefficients. Due to the characteristic of lightweight and flexible, it has been widely adopted for channel contribution adjustment and various conditionally adaptive scenes.

3. Proposed Method

3.1. Problem Formulation

As Fig. 1 shows, the proposed rate-controllable LIC method follows the mainstream LIC framework proposed by (Ballé et al., 2018), and the main difference is that we consider a scenario that the user want to conduct the image compression process with a desired bit-rate B. To achieve this goal, B is used as the side information guide the encoder and decoder models to realize rate-controllable image compression processes, which can be formulated by:

$$\begin{aligned} \boldsymbol{y} &= f_e(\boldsymbol{x}; \boldsymbol{\theta}_e, B), \\ \hat{\boldsymbol{y}} &= Q(\boldsymbol{y}), \\ \hat{\boldsymbol{x}} &= g_d(\hat{\boldsymbol{y}}; \boldsymbol{\theta}_d, B), \end{aligned} \tag{1}$$

where $\boldsymbol{x}, \hat{\boldsymbol{x}}, \boldsymbol{y}$, and $\hat{\boldsymbol{y}}$ represent the raw image, reconstructed image, latent features, and quantized latent features respectively. Models f_e, g_d represent encoder model and decoder model respectively, and $\boldsymbol{\theta}_e$ and $\boldsymbol{\phi}_d$ are the corresponding model parameters. We will illustrate the model structures in the following subsection to explain how to realize rate-controllable LIC. Function Q means quantization process, and to make the compression architecture learnable, in the training stage it is approximated by adding a additive uniform noise $\mathcal{U}\left(-\frac{1}{2}, \frac{1}{2}\right)$. In the inference stage, $Q(\boldsymbol{y})$ is realized as a rounding-based quantization of $\boldsymbol{y}$.

With the quantized latent feature $\hat{\boldsymbol{y}}$, entropy coding is used to generate the bitstream, at the same time realizing lossless compression of $\hat{\boldsymbol{y}}$. In the entropy parameter estimation process, a hyperprior model and a context model are used for obtaining the input features of the entropy parameter model. The hyperprior model is used to capture spatial dependencies of $\boldsymbol{y}$, and it is formulated as:

$$\begin{aligned} \boldsymbol{z} &= f_{he}(\boldsymbol{y}; \boldsymbol{\theta}_{he}, B), \\ \hat{\boldsymbol{z}} &= Q(\boldsymbol{z}), \\ \boldsymbol{\psi} &= g_{hd}(\hat{\boldsymbol{z}}; \boldsymbol{\theta}_{hd}, B), \end{aligned} \tag{2}$$

where f_{he} and g_{hd} denote the hyper-encoder and hyper-decoder respectively, $\boldsymbol{\theta}_{he}$ and $\boldsymbol{\phi}_{hd}$ represent the corresponding model parameters, $\boldsymbol{\psi}$ means the obtained hyperprior, $\boldsymbol{z}$ and $\hat{\boldsymbol{z}}$ means the hyper latent features and quantized hyper latent features respectively. Since the data amount of $\hat{\boldsymbol{z}}$ is quite small, a low-complexity factorized entropy model $\boldsymbol{\varphi}$ can be used for estimating the distributions of $\hat{\boldsymbol{z}}$, as given by

$$p_{\hat{\boldsymbol{z}}|\boldsymbol{\varphi}}(\hat{\boldsymbol{z}}|\boldsymbol{\varphi}) = \prod_i (p_{z_i|\boldsymbol{\varphi}}(\boldsymbol{\varphi}) * \mathcal{U}(-\frac{1}{2}, \frac{1}{2}))(\hat{z}_i), \tag{3}$$

where z_i denotes the i-th element of $\boldsymbol{z}$. Estimating high-quality distribution parameters of quantized latent features $\hat{\boldsymbol{y}}$ is critical for reducing the data redundancy in the following entropy coding process. Following the work of Cheng et al. (2020) we set that $\hat{\boldsymbol{y}}$ follow Gaussian mixture distributions, and a autoregressive context model g_{cm} is used to enhance the distribution estimation capability of the entropy parameter model g_{ep}, as given by

$$\begin{aligned} \boldsymbol{\phi}_i &= g_{cm}(\boldsymbol{y}_{<i}; \boldsymbol{\theta}_{cm}), \\ \omega_i, \mu_i, \sigma_i &= g_{ep}(\boldsymbol{\psi}, \boldsymbol{\phi}_i; \boldsymbol{\theta}_{ep}), \end{aligned} \tag{4}$$

where $\boldsymbol{\theta}_{cm}$ and $\boldsymbol{\theta}_{ep}$ denote the model parameters of the context model and entropy model, $\boldsymbol{\phi}_i$ means the causal context obtained from the context model, ω_i, μ_i, and σ_i denote the prior probability, the mean value, and the scaling factor of the i-th distribution respectively. We employ discretized Gaussian mixture likelihoods to achieve a more powerful entropy parameters model. With K groups of parameters of weights $\omega_{i,k}$, means $\mu_{i,k}$ and variance $\sigma^2_{i,k}$ for each elements $\hat{y}_i$, the entropy model can be denoted as:

$$p_{\hat{\boldsymbol{y}}|\hat{\boldsymbol{z}}}(\hat{y}_i|\hat{\boldsymbol{z}}, \boldsymbol{\theta}_{hd}, \boldsymbol{\theta}_{cm}, \boldsymbol{\theta}_{ep}) = \left(\sum_{k=1}^{K} \omega_{i,k} \mathcal{N}(\mu_{i,k}, \sigma^2_{i,k}) * \mathcal{U}(-\frac{1}{2}, \frac{1}{2}) \right) (\hat{y}_i). \tag{5}$$

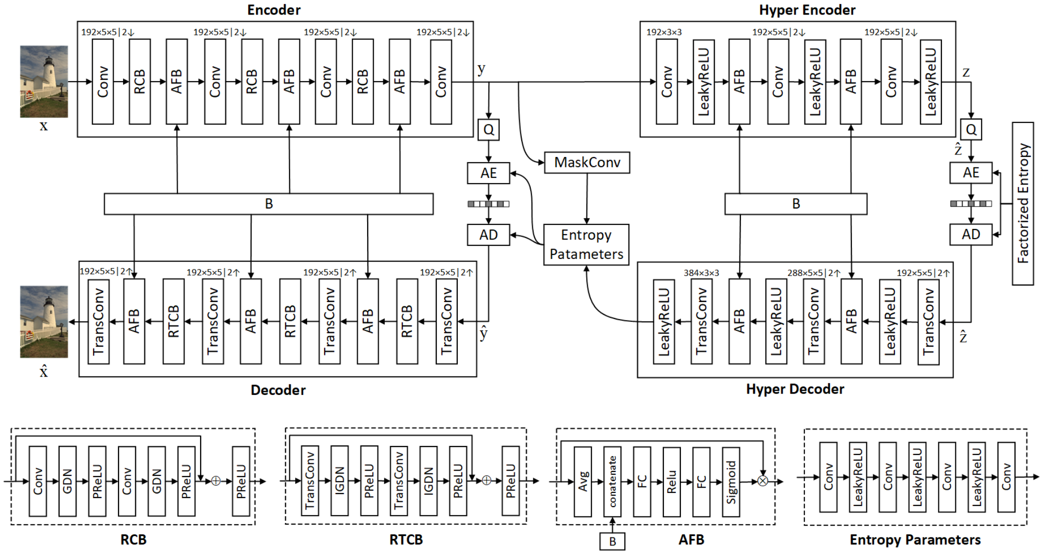

Figure 2: Rate controllable image compression network architecture

Then We can construct the following optimization model to realize the rate-controllable LIC:

$$\min_{\boldsymbol{\theta}_e,\boldsymbol{\phi}_d,\boldsymbol{\theta}_{he},\boldsymbol{\phi}_{hd},\boldsymbol{\theta}_{cm},\boldsymbol{\theta}_{ep},\boldsymbol{\theta}_{fep}} \quad \mathbb{E}\left\|\boldsymbol{x}-\hat{\boldsymbol{x}}\right\|_2^2,$$

$$\text{s.t.}\, \mathbb{E}\left[-\log_2(p_{\hat{\boldsymbol{y}}|\hat{\boldsymbol{z}}}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{z}}))\right]+\mathbb{E}\left[-\log_2(p_{\hat{\boldsymbol{z}}|\boldsymbol{\varphi}}(\hat{\boldsymbol{z}}|\boldsymbol{\varphi}))\right]=B. \tag{6}$$

In the above model, $\boldsymbol{\theta}_{fep}$ means the model parameters of the factorized entropy model φ. Then, by introducing the Lagrange multiplier λ, we can obtain the following rate-distortion loss function

$$\mathcal{L}(\boldsymbol{x},\hat{\boldsymbol{x}};\Omega)=\underbrace{\left\|\mathbb{E}[\log_2(p_{\hat{\boldsymbol{y}}|\hat{\boldsymbol{z}}}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{z}}))]+\mathbb{E}[\log_2(p_{\hat{\boldsymbol{z}}|\boldsymbol{\varphi}}(\hat{\boldsymbol{z}}|\boldsymbol{\varphi}))]+B\right\|_2^2}_{\text{rate-control}}+\lambda\cdot\underbrace{\mathbb{E}\left\|\boldsymbol{x}-\hat{\boldsymbol{x}}\right\|_2^2}_{\text{distortion}}, \tag{7}$$

where $\Omega=\{\boldsymbol{\theta}_e,\boldsymbol{\phi}_d,\boldsymbol{\theta}_{he},\boldsymbol{\phi}_{hd},\boldsymbol{\theta}_{cm},\boldsymbol{\theta}_{ep},\boldsymbol{\theta}_{fep}\}$ denote a set of model parameters. In the above rate-distortion function, the rate-control part means the difference between the bit-rate and the desired target bit-rate B. With the loss function, we can train the model to minimize the distortion between the input image and the reconstructed image, at the same time reducing the gap between the bit-rate and the desired bit-rate B.

3.2. Network Architecture

Fig. 2 illustrates the network architecture of the proposed rate-controllable LIC framework. For the encoder model, it is realized by four convolutional (Conv) layers, three residual convolutional blocks (RCB), and three attention feature blocks (AFB). A convolutional layer is designated as $c \times h \times w| \updownarrow s$, where c, h and w correspond to the number of output channels, height, and width of the convolutional kernel, respectively. The symbols ↑ and ↓ indicate upsampling and downsampling operations, and s denote the stride length. For the RCB, it follows a typical residual structure, and Generalized Divisive Normalization (GDN) is used as the normalization layer. The structure of the AFB is almost the same with channel attention module in Hu et al. (2018), except that the desired bit-rate B is also an input data. More specifically we first compute the channel-wise mean value of the image features, then concatenate it with the desired bit-rate B to obtain the input of the attention model. Sigmoid activation function is used for restricting the value of the learned weights in range $(0, 1)$, and each of them be used to adjust the values of feature channel. With the help of AFB, we expect that the model can adaptively adjust the compression process to realize rate-controllable LIC with desired bit-rate B. For the first three convolutional layers, each of them is followed by a RCB and a AFB, which are used for learning effective feature representations and control the distribution of feature channels.

For the decoder model, it conducts the image reconstructing process, and its structure can be regarded as a inverse process of the RCB. More specifically, the convolutional layers are all substituted as transposed-convolutional layers (TransConv), and the downsampling process becomes upsampling process. For the context model, a 5×5 masked convolution is used to capture the correlation among adjacent elements in the latent representation, which have been demonstrated useful for enhancing the capability of the entropy parameter model (Cheng et al., 2020). The output of context model will be combined with the hyperprior information, then fed into the entropy parameters model to generate K groups of mean and standard deviation parameters for the latent feature distributions. Gaussian mixture models and factorized entropy models are utilized to estimate the distributions of the latent features $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{z}}$, respectively. The network models of the hyperprior model, the entropy parameters model, and the factorized entropy parameters model are the same with the models in (Cheng et al., 2020).

3.3. Two-step Training

In our test, we observe that the best rate control capability cannot be learned by directing training the proposed model shown in Fig. 2. Therefore, the following two-stage training strategy is used to obtain rate-controllable LIC models:

Stage 1: We pre-train the network without the AFB module, and using the classical rate-distortion function $R + \lambda D$ as the loss function. In this stage, the Lagrange multiplier is required to be relatively large value, such as $\lambda = 8192$ for PSNR. In this way, the pretrained model will have a large bit-rate, such that we can scale it in a wide range of bit-rates.

Stage 2: The AFB module is introduced into the pre-trained model, and we further train the network by using the new rate-distortion loss defined in (7). For each training sample, its target bit-rate B is randomly generated in range $[B_{min}, B_{max}]$, where B_{min} and B_{max} denote the lower and upper bounds of the target bit rate. In this paper, we use bit-per-feature (BPF) to quantify the bit rate, and it has broader value range compared with the commonly used bit-per-pixel (BPP) metric.

4. Experiments

4.1. Implementation Details

We employed a subset of the ImageNet dataset (Deng et al., 2009) as the training dataset, which consists of 50000 images randomly selected from its training set. The training samples are randomly cropped as sizes of 256×256, and data augmentation techniques including horizontal and vertical flipping are adopted to enhance the generalization capability of trained model. We trained the model with channel number $c = 192$ and Gaussian mixture number $K = 3$. The model is optimized using the Adam optimizer with a batch size of 8. The learning rate is selected from the set $\left\{1 \times 10^{-4}, 5 \times 10^{-5}, 1 \times 10^{-5}, 5 \times 10^{-6}, 1 \times 10^{-6}\right\}$, and the learning rate schedule is set as follows:

- Stage 1: During the pre-training phase, the learning rate was initially fixed at 1×10^{-4} for the first 100 epochs followed by a decay every 60 epochs, and resulting in a total of 340 epochs.
- Stage 2: During the fine-tune phase, the learning rate was fixed at 1×10^{-4} for the first 100 epochs followed by a decay every 40 epochs.

To realize stable training process in stage 2, we first generate a staircase bit-rate $B \in [0.2, 2]$ with fixed interval 0.225, i.e. the bit-rate of the 8 samples are $0.2, 0.425, 0.65, ..., 2$. Then, uniformly distributed noise within range $[-0.1125, 0.1125]$ are added to ensure that the bit rates cover the continuous rate region. Mean squared error (MSE) was used as the distortion metric.

We evaluat our method on the Kodak dataset (Eastman Kodak Company, 1993), which consists of 24 lossless images with resolution of 768×512. To facilitate comparison with other compression methods, the inferred BPF values are converted to the BPP domain, which is considered as a standard metric for rate evaluation.

4.2. Performance

We evaluate the rate control performance of the proposed model, and comparing its performances to several classical compression standards: JPEG, JPEG2000, BPG, and VTM-12.1 (Joint Video Experts Team, 2021) , as well as neural network-based fixed bit-rate models proposed by (Ballé et al., 2018; Minnen et al., 2018; Cheng et al., 2020), and the variable-rate adaptive compression method proposed by (Cui et al., 2021). Note that, our primary goal is not demonstrating that the proposed method can achieve superior compression capability compared with existing methods, but demonstrating that the proposed method can achieve effective rate control with one LIC model in a continuous domain.

Fig. 3 shows the rate control results on Kodak datasets. In this test, we conducted 45 independent tests on each image with randomly generated values of B, and record the results of the desired target BPPs and inference BPPs. For Kodak data, We can observe that, in high target BPP region, i.e. $B > 0.4$, the BPP of the inference process are quite close to the target BPP, and the MSE and mean absolute error (MAE) between the target BPPs and the corresponding inference BPPs are 0.0594 and 0.0083, respectively. In a small target BPP region, i.e. BPP≤ 0.4, the gap between the target BPP and inference BPP becomes more evident when the target BPP becomes smaller, and the MSE and MAE between the target BPPs and the corresponding inference BPPs become 0.0025 and 0.0347, respectively. This is because we use a relatively large value of λ for pre-training the model to cover a wide range of target BPPs, and its rate control ability will be decreased if the target BPP is too small. Overall, the MAE and MSE between the the target BPPs and the corresponding inference BPPs are 0.0594 and 0.0083, respectively, and the inference BPPs range from 0.1759 to 1.4279.

Next, we compare the achieved PNSR performances of the proposed method and other methods. Figs. 4 shows the results obtained on Kodak datasets. By comparing the results of the proposed method with AFB and the models without AFB, we can observe that the proposed method can realize a wide range of continuous rate control, and the PSNR performances achieved are almost the same with the models without AFB modules, demonstrating that the proposed method can achieve both continuous rate control and high-quality image reconstruction at the same time. It should be

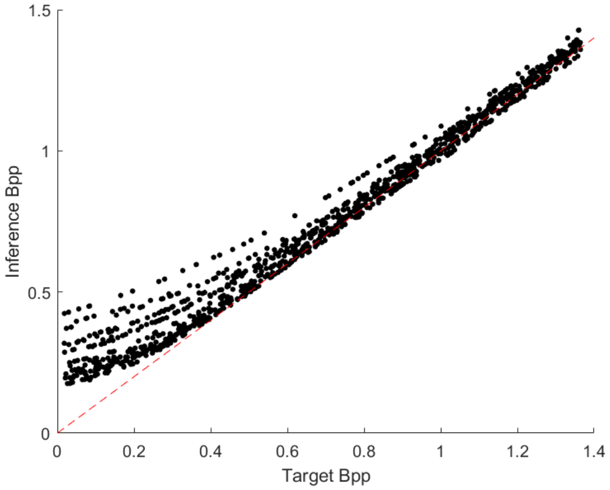

Figure 3: Rate control performance on Kodak.The dots represent the actual bit rate obtained during inference at the specified target BPP. The red line serves as a reference, with dots closer to the red line indicating a closer match between the inferred BPP and the target BPP.

note that when BPP > 0.9, the PSNR performances of the proposed method is slightly lower than the model without AFB modules. This result maybe caused the unstable model training process. More specifically, the model is trained to allocate more resources to minimize the rate control loss in high BPP region, and rate control is also more easy to be achieved in this situation, resulting a slight decrease of the distortion loss.

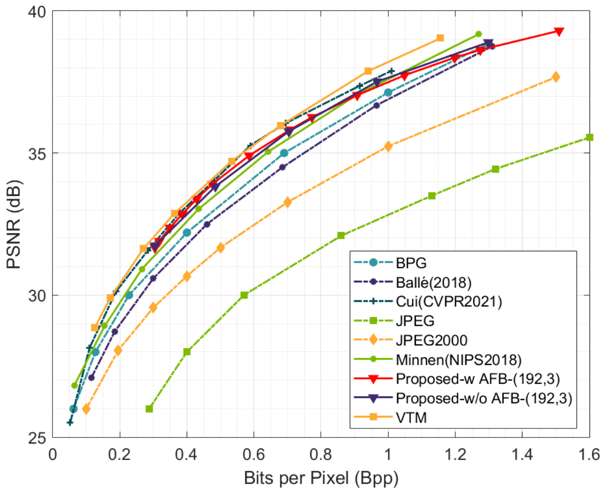

Figure 4: Comparison of rate-distortion Performances on Kodak dataset

For R-D performances, we can see that, in Fig. 4, the proposed method obtains competitive performance compared with Cui et al. (2021). When BPP < 0.6. For BPP > 0.6, the network shows a slight performance drop with the largest discrepancy at BPP $= 1.012$, where the PSNR is approximately 0.3 dB lower than that of Cui et al. (2021). This is because the complexity of the model in Cui et al. (2021) is much higher that the proposed method. More specifically, it uses 3 mask-conv models, the spatial attention module, and Asymmetric Gaussian Entropy Model, which enhance its image reconstruction capability in high BPP region. Compared with classical compression standards, the performance of proposed method is lower than VTM.

In conclusion, the above results demonstrate that the proposed model can achieve comparable performances compared with existing methods. Most importantly, the results show that the proposed method can realize effective rate control in LIC in a wide BPP region, and the performance is almost the same with the model without rate control capability.

5. Conclusion

In this paper, we proposed a rate-controllable LIC method by employing channel attention to adaptively adjust the distributions of feature channels according to target bit-rates. We introduce a new rate-distortion loss function to train the model to conduct the compression process with specified bit-rates, which enables a single model to realize continuous and controllable bit rates. We used a two-stage training strategy including model pretraining with fixed bit-rate, and model fine-tuning with randomly generated bit-rates. Experimental results demonstrate that our method achieves effective continuous rate control over a wide range of rates. In terms of compression performance, our approach surpasses most classical image compression codec and is comparable to advanced learned image coding methods. Our framework is the first one to enable direct control of continuous rates with a single model, which can enhance the convenience and flexibility of LIC models in implementation. In addtion, the channel attention-based rate control mechanism is applicable to other learned image compression frameworks. We will further investigate more complex network architectures to enhance its rate-distortion performance, and eliminating the performance loss of the proposed method in high BPP region.

References

Johannes Ballé, Valero Laparra, and Eero P Simoncelli. End-to-end optimized image compression. *arXiv preprint arXiv:1611.01704*, 2016.

Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston. Variational image compression with a scale hyperprior. *arXiv preprint arXiv:1802.01436*, 2018.

Bellard. Bpg image format. https://bellard.org/bpg/, 2014.

Zhengxue Cheng, Heming Sun, Masaru Takeuchi, and Jiro Katto. Learned image compression with discretized gaussian mixture likelihoods and attention modules. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7939–7948, 2020.

Ze Cui, Jing Wang, Shangyin Gao, Tiansheng Guo, Yihui Feng, and Bo Bai. Asymmetric gained deep image compression with continuous rate adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10532–10541, 2021.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

Eastman Kodak Company. Kodak lossless true color image suite (photocd pcd0992), 1993. URL http://r0k.us/graphics/kodak/.

Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

Zhaoyang Jia, Jiahao Li, Bin Li, Houqiang Li, and Yan Lu. Generative latent coding for ultra-low bitrate image compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26088–26098, 2024.

Wei Jiang, Jiayu Yang, Yongqi Zhai, Feng Gao, and Ronggang Wang. Mlic++: Linear complexity multi-reference entropy modeling for learned image compression. *arXiv preprint arXiv:2307.15421*, 2023.

Joint Video Experts Team. Vvc official test model vtm. https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tree/VTM-12.1, 2021.

Jae-Han Lee, Seungmin Jeon, Kwang Pyo Choi, Youngo Park, and Chang-Su Kim. Dpict: Deep progressive image compression using trit-planes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16113–16122, 2022.

Chao Li, Tianyi Li, Fanyang Meng, Qingyu Mao, Youneng Bao, Yonghong Tian, and Yongsheng Liang. One is all: A unified rate-distortion-complexity framework for learned image compression under energy concentration criteria. *IEEE Transactions on Multimedia*, 2025.

Jinming Liu, Heming Sun, and Jiro Katto. Learned image compression with mixed transformer-cnn architectures. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14388–14397, 2023.

David Minnen, Johannes Ballé, and George D Toderici. Joint autoregressive and hierarchical priors for learned image compression. *Advances in neural information processing systems*, 31, 2018.

Yichen Qian, Ming Lin, Xiuyu Sun, Zhiyu Tan, and Rong Jin. Entroformer: A transformer-based entropy model for learned image compression. *arXiv preprint arXiv:2202.05492*, 2022.

Majid Rabbani and Rajan Joshi. An overview of the jpeg 2000 still image compression standard. *Signal processing: Image communication*, 17(1):3–48, 2002.

Lucas Relic, Roberto Azevedo, Markus Gross, and Christopher Schroers. Lossy image compression with foundation diffusion models. In *European Conference on Computer Vision*, pages 303–319. Springer, 2024.

Myungseo Song, Jinyoung Choi, and Bohyung Han. Variable-rate deep image compression through spatially-adaptive feature transform. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2380–2389, 2021.

Gregory K Wallace. The jpeg still picture compression standard. *IEEE transactions on consumer electronics*, 38(1):xviii–xxxiv, 1992.

Renjie Zou, Chunfeng Song, and Zhaoxiang Zhang. The devil is in the details: Window-based attention for image compression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17492–17501, 2022.