

# Reinforced Rate Control for Neural Video Compression via Inter-Frame Rate-Distortion Awareness

Wuyang Cong<sup>1</sup>, Junqi Shi<sup>1</sup>, Lizhong Wang<sup>2</sup>, Weijing Shi<sup>2</sup>, Ming Lu<sup>1\*</sup>, Hao Chen<sup>1</sup>, and Zhan Ma<sup>1\*</sup>

<sup>1</sup>School of Electronic Science and Engineering, Nanjing University

<sup>2</sup>Samsung Research China-Beijing

{congwuyang,junqishi}@smail.nju.edu.cn, {lz.wang,weijing...shi}@samsung.com, {minglu,chenhao1210,mazhan}@nju.edu.cn

## Abstract

Neural video compression (NVC) has demonstrated superior compression efficiency, yet effective rate control remains a significant challenge due to complex temporal dependencies. Existing rate control schemes typically leverage frame content to capture distortion interactions, overlooking inter-frame rate dependencies arising from shifts in per-frame coding parameters. This often leads to suboptimal bitrate allocation and cascading parameter decisions. To address this, we propose a reinforcement-learning (RL)-based rate control framework that formulates the task as a frame-by-frame sequential decision process. At each frame, an RL agent observes a spatiotemporal state and selects coding parameters to optimize a long-term reward that reflects rate-distortion (R-D) performance and bitrate adherence. Unlike prior methods, our approach jointly determines bitrate allocation and coding parameters in a single step, independent of group of pictures (GOP) structure. Extensive experiments across diverse NVC architectures show that our method reduces the average relative bitrate error to 1.20% and achieves up to 13.45% bitrate savings at typical GOP sizes, outperforming existing approaches. In addition, our framework demonstrates improved robustness to content variation and bandwidth fluctuations with lower coding overhead, making it highly suitable for practical deployment.

## 1 Introduction

Past years have witnessed the explosive growth of neural video compression (NVC) approaches (Lu et al. 2019; Liu et al. 2020; Li, Li, and Lu 2021, 2023, 2024; Lu et al. 2024; Jia et al. 2025), which leverages the powerful nonlinear modeling capabilities of deep neural networks (DNNs) and end-to-end optimization to surpass traditional video coding standards in compression efficiency (Bross et al. 2021). Despite these breakthroughs, rate control in NVC remains underexplored, with only a few recent efforts addressing this fundamental problem (Li et al. 2022; Zhang et al. 2023; Chen et al. 2023), although rate control is crucial for the practicality of NVC.

Similar to traditional codecs, rate control in NVC aims to meet bitrate constraints while maximizing reconstruction quality. This typically involves dynamically adjusting coding parameters of each coding unit<sup>1</sup> (e.g., the Lagrange multiplier

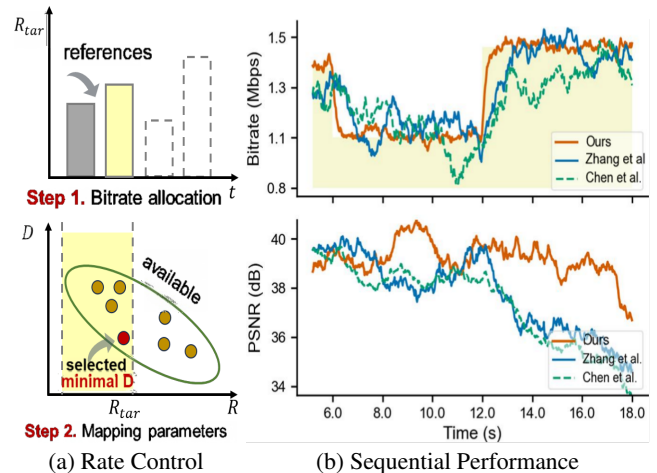


Figure 1: Rate Control in NVC. (a) It typically involves bitrate allocation and parameter mapping. (b) The effectiveness of rate control is reflected in its ability to accurately meet the target bitrate while incurring minimal quality degradation.

and/or resolution) (Sullivan 1998). Essentially, it requires learning a policy that allocates frame-level target bitrate and maps to the optimal coding parameters, as shown in Fig. 1(a). However, due to inter-frame dependencies, this process is affected by preceding frames and impact subsequent frames, making the global optimization problem NP-hard that computationally intractable via brute-force search.

A feasible solution is to approximate this global optimization problem using window-based schemes (Li et al. 2022; Zhang et al. 2023; Chen et al. 2023), where the target bitrate is first uniformly distributed across windows (e.g., GOPs), and then rule-based or heuristic strategies are applied within each window to allocate bitrate at the frame level. These methods follow the philosophy of traditional codecs—leveraging content complexity to model inter-frame distortion dependencies while assuming negligible rate dependencies (Hu et al. 2011; Wang et al. 2013b; Li et al. 2020). However, this assumption does not hold for NVCs. Due to their jointly optimized pixel-level, feature-based, and contextual references information, NVCs exhibit complex and tightly coupled rate and distortion dependencies (Sheng et al. 2024). These dependencies can-

\*Corresponding authors.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

<sup>1</sup>Coding unit includes frame, GOP (group of pictures), etc.

not be accurately modeled using only content characteristics without incorporating information from actual encoded references. Once rate control is introduced, even minor changes in coding parameters of reference frames can lead to substantial variations in inter-frame dependencies, causing long-term shifts in the rate-distortion (R–D) behavior of both current and subsequent frames. Hence, relying solely on distortion estimations based on frame content leads to suboptimal rate allocation and coding parameters, ultimately degrading both rate accuracy and R–D performance, as shown in Fig. 1(b).

To tackle this, we propose a reinforcement learning (RL)-based rate control framework for NVC, formulated as a *Constrained Markov Decision Process (CMDP)* (Altman 2021). Unlike prior schemes, our method learns a dynamic policy that jointly considers reference information, frame content and network bandwidth variations as input state. This enables the model to capture not only frame-level characteristics but also the rate and distortion dependencies induced by references. The policy directly maps states to frame-level coding parameter actions, with each action conditioned on preceding states and optimized with respect to its long-term impact. This sequential decision-making strategy enables globally-aware rate control that accounts for both frame content and sustained impacts of inter-frame dependencies in NVC.

Technically, we design an enhanced Actor-Critic architecture (Haarnoja et al. 2018, 2019) that integrates a neural spatiotemporal state extractor for capturing current frame contents and inter-frame dependencies, a distributed policy to support robust exploration across a diverse action space, and a tempered reward mechanism that delivers steady feedback. These components collectively enable our scheme to adapt rapidly to varying states while maintain high decision quality and efficiency as Fig. 1(b). Our main contributions are:

- We conduct both theoretical and empirical analyses of the rate control problem in NVC. In contrast to traditional codecs, our study highlights the critical role of inter-frame dependencies—spanning both distortion and bitrate—in shaping rate control behavior;
- We propose the first CMDP-based RL framework for rate control in NVC, which integrates spatiotemporal state modeling, robust distributed exploration, and adaptive reward tempering, formulating a dynamic policy directly deciding per-frame coding parameters of NVC;
- We demonstrate that our method consistently outperforms prior approaches across multiple NVC architectures, achieving superior rate accuracy, bitrate savings, and robustness with minimal complexity overhead.

## 2 Related Work

### 2.1 Neural Video Compression and Rate Control

NVC methods draw inspiration from traditional hybrid coding paradigms (Wiegand et al. 2003; Sullivan et al. 2012; Bross et al. 2021), designing various DNNs to implement key components such as intra-frame texture coding, inter-frame residual coding, and motion representation within an end-to-end learning framework (e.g., DVC (Lu et al. 2019)). These processes, including motion estimation/compensation,

intra/inter prediction and residual coding, can be performed either in the original pixel domain (Lu et al. 2019; Liu et al. 2020) or in a learned latent space (Hu, Lu, and Xu 2021; Liu et al. 2022), enabling greater flexibility in modeling. The R-D trade-off is typically optimized jointly using a Lagrangian multiplier, optionally combined with a resolution scaling factor to better control bitrate and reconstruction quality (Alexandre, Hang, and Peng 2022).

Following the success of the DVC series, conditional coding introduced in DCVC (Li, Li, and Lu 2021) further improved the efficiency of inter-frame feature representation and compression. Building on this, successors such as DCVC-DC (Li, Li, and Lu 2023) and DCVC-RT (Jia et al. 2025) have demonstrated significant gains over the latest VVC standard (Bross et al. 2021) in terms of compression performance. Despite their effectiveness, these models still operate under fixed or heuristically tuned bitrate settings during inference, limiting their adaptability to practical application scenarios.

To enable rate control in NVC, existing approaches often draw inspiration from traditional codecs (Li et al. 2014; Liang et al. 2013; Wang et al. 2013a; Li et al. 2020). They divide the video sequence into windows (e.g., GOPs) and strive to fit the distortion dependencies across and within the windows, either by applying fixed R–D– $\lambda$  models (Li et al. 2022; Chen et al. 2023; Zhang and Gao 2024), or rely on empirical adjustments based on historical coding statistics (Jia et al. 2025; Yang et al. 2025) or heuristic pre-analysis (Mandhane et al. 2022; Gu et al. 2024). Then according to this preset distortion dependencies relationship, they allocate bitrate and finally map it to per-frame coding parameters. However, these methods still abide by the traditional codecs’ characteristics, only consider the distortion relationships but ignore the coupled rate and distortion dependencies, which is not suitable for NVC due to its complex inter-frame references. Moreover, unlike prior overfitting methods (Lu et al. 2020; Tang et al. 2024; Chen et al. 2024), which are too slow, our scheme is designed to integrate a practical rate control tool into NVC.

### 2.2 Reinforcement Learning

RL focuses on learning a dynamic policy  $\pi$  that selects an action  $a_t$  based on the current state  $s_t$  to maximize the expected cumulative return. This policy interacts with the environment by iteratively updating states and sampling actions that balance immediate rewards  $r_t$  with the expected discounted sum of future rewards (discounted by factor  $\gamma$ ), optimized via a Q-value function  $Q^\pi(s_t, a_t)$ . In this framework, the global optimization problem is recast as a sequential decision-making process over a temporal horizon:

$$Q^\pi(s_t, a_t) = r(s_t, a_t) + \gamma \mathbb{E}_{s_{t+1}, a_{t+1}} [Q^\pi(s_{t+1}, a_{t+1})]. \quad (1)$$

At each timestep  $t$ , future outcomes are uncertain due to the non-deterministic state transition  $\mathcal{P}(s_{t+1}|s_t, a_t)$ . This uncertainty is analogous to rate control in NVC, where future content, network bandwidth, and codec behavior may change unpredictably. Fundamentally, this scenario reflects the exploration-exploitation dilemma central to RL. In this work, we propose an enhanced Actor-Critic framework (Konda and Tsitsiklis 1999; Haarnoja et al. 2018, 2019)

that improves state representation, action sampling, and reward shaping, yielding a more practical and effective trade-off between performance and adaptability.

While RL techniques have been explored in rate control for traditional codecs, they either build on traditional rule-based tools by using RL to explore better rules (Zhou et al. 2021; Ho et al. 2021b; Gadot et al. 2025), or rely on heuristic search methods that require multiple pre-codings (Mao et al. 2020; Mandhane et al. 2022). In essence, none of them explicitly account for the coupled inter-frame rate and distortion dependencies in NVC, causing suboptimal performance. Moreover, the lack of robust existing tools and dependence of multiple pre-codings further restrict their applicability to NVC.

### 3 Rate Control for NVC

In this section, we first provide a re-analysis of the rate control problem in NVC, followed by an in-depth analysis of how rate and distortion coupled dependencies impact rate control.

#### 3.1 Problem Formulation

Rate control for a video sequence  $\mathcal{X} = \{x_1, x_2, \dots, x_N\}$  of length  $N$  can be formulated as a constrained optimization problem. Given a target bitrate  $R_{tar}$  imposed at a specific level (e.g., sequence or GOP level), the goal is to determine an optimal set of frame-wise coding parameters  $\Pi^{(\mathcal{X})} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N\}$  that minimizes the total distortion:

$$\Pi^{(\mathcal{X})} = \arg \min_{\Pi^{(\mathcal{X})}} \sum_{t=1}^N D_t, \text{ s.t. } \frac{1}{N} \sum_{t=1}^N R_t \leq R_{tar}, \quad (2)$$

This constrained problem can be equivalently reformulated in an unconstrained form by introducing a global Lagrangian multiplier  $\Lambda$  that balances distortion minimization and rate constraint satisfaction:

$$\Pi^{(\mathcal{X})} = \arg \min_{\Pi^{(\mathcal{X})}} \sum_{t=1}^N D_t + \Lambda \left( \frac{1}{N} \sum_{t=1}^N R_t - R_{tar} \right). \quad (3)$$

Taking the derivative of Eq. (3) with respect to each  $\mathbf{a}_t$  yields the necessary condition for the optimal parameters set:

$$\frac{\partial \sum_{t=1}^N D_t}{\partial \mathbf{a}_t} + \frac{\Lambda}{N} \frac{\partial \sum_{t=1}^N R_t}{\partial \mathbf{a}_t} = 0, \quad t = 1, 2, \dots, N. \quad (4)$$

According to R-D theory, the R-D function is a convex, and its slope at each point is given by  $\lambda_t = \partial D_t / \partial R_t$ . In traditional video codecs, it is typically assumed that inter-frame bitrate dependencies are negligible, and distortion dependencies are approximated using fixed rules or heuristics (Hu et al. 2011; Wang et al. 2013b; Li et al. 2020). Under these assumptions, Eq. (4) can be transformed into the following optimality criterion:

$$\lambda_t = \frac{\Lambda}{N \cdot \frac{\partial \sum_{i=t}^N D_t}{\partial D_t}} = \omega_t \cdot \Lambda, t = 1, 2, \dots, N. \quad (5)$$

However, these assumptions do not hold in jointly-trained, non-linear NVCs. Their complex pixel-level, feature-based, and contextual dependencies not only introduce strong inter-frame distortion dependencies but also lead to tightly coupled

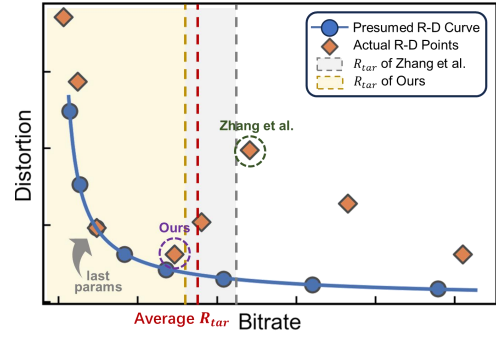


Figure 2: Rate Control Process for the 25-th Frame on *BasketballDrive*. Ignoring inter-frame rate dependencies leads to improper bitrate allocation. As a result, the coding decision still follows the pretrained R-D curve (blue), producing suboptimal parameters (green “Zhang et al.”).

bitrate dependencies. For example, temporal context modeling directly affects the estimated probability distributions of latent features, thereby influencing the actual bitrate. As a result, Eq. (4) must be revised to capture these interactions:

$$\sum_{i=t}^N \left( \frac{\Lambda}{N} \frac{\partial R_i}{\partial \mathbf{a}_t} - \frac{\partial D_i}{\partial \mathbf{a}_t} \right) = \sum_{i=t}^N \left( \left( \frac{\Lambda}{N} - \lambda_i \right) \frac{\partial R_i}{\partial \mathbf{a}_t} \right) = 0, \quad (6)$$

which implies that the optimal coding parameter  $\mathbf{a}_t$  must not only reflect the frame’s R-D behavior (i.e.,  $\lambda_t$ ) but also consider propagated rate and distortion impact on future frames.

#### 3.2 The Impact of Inter-frame Dependencies

Typically, a pretrained NVC is optimized to adapt to a specific R-D dependency under a fixed global  $\Lambda$  constraint. However, once rate control is introduced, frame-wise variations in  $\lambda$  induce new rate and distortion dependencies that diverge from those learned during pretraining.

To further investigate the impact of inter-frame dependencies-formulated in Eq. (6)-on rate control in NVC, we conduct a toy experiment on the *BasketballDrive* sequence using the DCVC-DC codec. Following the state-of-the-art rate control method for NVC (Zhang et al. 2023), we set the target bitrate for the entire sequence to match the average bitrate achieved by standard fixed-QP coding with  $QP = 32$ , where the quantization parameter (QP) corresponds one-to-one with  $\lambda$ . Fig. 2 illustrates the rate control behavior for the 25-th frame. The blue solid line represents the assumed R-D curve for the frame under the pretrained model, typically a smooth hyperbolic function that does not account for reference frame impacts. However, in the rate controlled setting, the 24-th reference frame may be encoded with a mismatched QP (e.g.,  $QP=12$ ), leading to altered inter-frame dependencies. This discrepancy shifts the R-D behavior of the 25th frame, causing it to deviate from the original trajectory—as shown by the scattered orange points.

Since prior methods do not account for reference frame information produced during actual encoding, they cannot capture this shift. As a result, they allocate an inappropriate

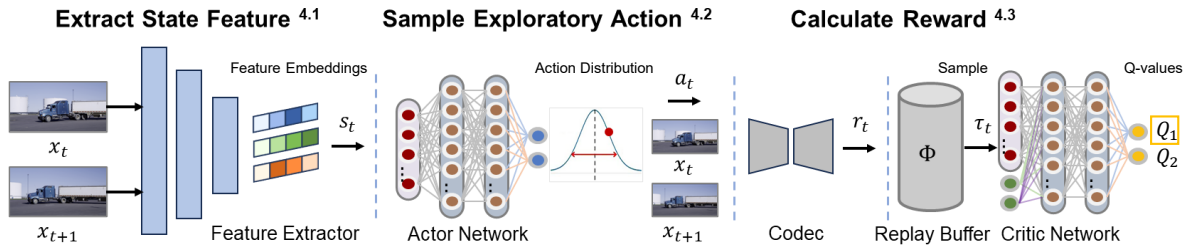


Figure 3: Training Pipeline of the Proposed Actor-Critic Network. The entire process consists of three main components, from state to action and then to reward, corresponding to Sec. 4.1, 4.2 and 4.3, in sequence.

bitrate and select coding parameters that minimize distortion based on a static R–D assumption—corresponding to the 6th point on the blue curve. However, due to the reference QP mismatch, the actual operating point drifts to the 6th orange point (denoted as “Zhang et al.”), leading to suboptimal performance and potential bitrate overshoot. In contrast, our proposed framework more accurately models the true R–D behavior, enabling more effective bitrate allocation. Consequently, the selected coding parameters better align with the target average bitrate while achieving lower distortion—corresponding to the 4th point (denoted as “Ours”).

These findings reveal a critical limitation in existing NVC rate control methods, echoing the observations in Sec. 3.1: both frame-level R–D characteristics and inter-frame rate and distortion dependencies must be considered. To this end, we propose an RL-based rate control framework that explicitly models frame content and reference-induced dependencies, and directly determines per-frame coding parameters by maximizing the expected cumulative reward. This approach inherently aligns with the formulation in Eq. (3), with further technical details provided in Sec. 4.

## 4 Reinforced Rate Control Framework

As shown in Eq. (6), the core challenge lies in accurately modeling frame-wise states while exploring each action’s long-term impact on future frames. To tackle this, we design an enhanced Actor–Critic framework (illustrated in Fig. 3), with the following key innovations.

### 4.1 State Modeling

In RL, an informative and compact state representation is crucial for both effective policy learning and generalization (Li, Walsh, and Littman 2006; Mnih et al. 2015). According to Eq. (6), the state in rate control should encapsulate both historical encoded references and current frames. Unlike prior schemes that rely on handcrafted features (Chen, Hu, and Peng 2018; Zhou et al. 2020), we learn this representation end-to-end using a neural embedding network conditioned on current frame, references, and auxiliary information.

Specifically, the current frame  $x_t$  and reference frame  $x_{t-1}$  are concatenated and passed through a cascaded residual network for spatial-temporal feature extraction. Additionally, intermediate features of  $x_{t-1}$ -extracted by the codec at multiple resolutions—are fused to enhance the temporal context. The resulting embeddings are further refined using several convolutional layers and average pooling operations.

To supplement visual context, auxiliary information such as the target bitrate and previously selected coding parameters is normalized, expanded, and embedded through fully connected layers. The combined visual and auxiliary embeddings form a comprehensive, learnable state representation.

Compared to prior methods, our learned state embedding flexibly incorporates dynamic references and frames’ characteristics, enabling more accurate and adaptive modeling of R–D behavior for frame-wise decision-making. Detailed architecture and setups are included in the Appendix.

### 4.2 Action Decision

The action in our RL framework is defined as a pair of continuous coding parameters  $\{\lambda_t, m_t\}$ , where  $\lambda_t \in [\lambda_{min}, \lambda_{max}]$  denotes the Lagrange multiplier controlling the R–D behavior, and  $m_t \in [0.5, 1.0]$  is down-sampling factor that adjusts the spatial resolution of both the current and reference frames.

Given the continuous and high-dimensional nature of the action space, we model the policy  $\pi_\phi$  as a Gaussian distribution, where both the mean and variance are predicted by the Actor network  $\phi$ . This probabilistic formulation allows for exploration beyond deterministic actions used in prior rate control strategies and improves adaptability across diverse states. To further encourage exploration, we incorporate policy entropy regularization (Haarnoja et al. 2018) and optimize the Actor using the policy gradient:

$$J_\pi(\phi) = \mathbb{E}_{s_t \sim \mathcal{S}, a_t \sim \pi_\phi} [\epsilon \log \pi_\phi(a_t | s_t) - Q_\theta(s_t, a_t)], \quad (7)$$

During inference, we adopt a greedy strategy by selecting the action with the highest likelihood. If resolution scaling ( $m_t < 1.0$ ) occurs, we resample the reference frame to match the current resolution for consistent inter-frame prediction, then bicubically upsample the output to the original resolution. This joint  $\lambda$ – $m$  policy not only provides precise rate control but also reduces computational cost by enabling lower-resolution processing when appropriate.

### 4.3 Reward Shaping

Rewards serve as the learning signal to evaluate the quality of selected actions. However, in rate control tasks, meaningful metrics such as total distortion or bitrate deviation are only available after coding an entire sequence, making rewards inherently sparse. While previous methods attempt to define intermediate rewards using off-the-shelf tools (Zhou et al. 2020; Ho et al. 2021a), heuristics (Mandhane et al. 2022), or fixed allocation rules, yet no general solution exists for NVC.

GOP	Codec	Method	Dataset						
			UVG	MCL-JCV	HEVC B	HEVC C	HEVC D	HEVC E	Avg.
32	DVC	Zhang et al.	—	—	—	—	—	—	—
		Chen et al.	1.64 / -14.57	2.11 / -12.34	1.69 / -15.01	1.45 / -14.72	1.38 / -13.89	1.65 / -20.23	1.65 / -15.13
		Ours	<b>1.32 / -16.97</b>	<b>1.82 / -17.11</b>	<b>1.23 / -16.64</b>	<b>1.12 / -16.28</b>	<b>1.08 / -15.97</b>	<b>1.23 / -21.43</b>	<b>1.30 / -17.40</b>
	DCVC	Zhang et al.	—	—	—	—	—	—	—
		Chen et al.	1.85 / -14.28	2.03 / -10.88	1.96 / -12.23	1.73 / -9.47	1.81 / -10.98	1.18 / -15.25	1.76 / -12.18
		Ours	<b>1.80 / -18.24</b>	<b>1.79 / -17.11</b>	<b>1.15 / -14.83</b>	<b>1.65 / -14.98</b>	<b>1.51 / -15.03</b>	<b>0.99 / -18.76</b>	<b>1.48 / -16.49</b>
	DCVC-DC	Zhang et al.	—	—	—	—	—	—	—
		Chen et al.	1.66 / -10.33	1.83 / -9.92	1.71 / -11.02	1.55 / -9.67	1.60 / -9.84	1.28 / -13.00	1.61 / -10.63
		Ours	<b>1.45 / -13.84</b>	<b>1.57 / -12.51</b>	<b>0.98 / -14.82</b>	<b>0.97 / -13.03</b>	<b>0.93 / -12.98</b>	<b>0.85 / -16.70</b>	<b>1.13 / -13.98</b>
	DCVC-RT	Zhang et al.	—	—	—	—	—	—	—
		Chen et al.	1.49 / -5.12	1.71 / -5.33	1.44 / -5.26	1.35 / -4.08	1.22 / -4.10	1.50 / -4.98	1.45 / -4.81
		Ours	<b>1.18 / -5.84</b>	<b>1.27 / -5.31</b>	<b>1.16 / -6.00</b>	<b>1.09 / -4.79</b>	<b>1.23 / -4.86</b>	<b>0.96 / -6.17</b>	<b>1.15 / -5.50</b>
100	DVC	Zhang et al.	2.82 / -11.61	2.79 / -8.78	1.35 / -10.99	1.18 / -10.63	1.91 / -12.17	<b>1.11 / -18.28</b>	1.86 / -12.08
		Chen et al.	1.73 / -16.11	2.16 / -13.08	1.68 / -15.33	1.66 / -16.02	1.50 / -14.54	1.79 / -20.15	1.75 / -15.87
		Ours	<b>1.30 / -17.04</b>	<b>1.88 / -17.28</b>	<b>1.20 / -16.63</b>	<b>1.01 / -16.55</b>	<b>1.05 / -16.14</b>	<b>1.21 / -20.05</b>	<b>1.28 / -17.28</b>
	DCVC	Zhang et al.	2.80 / -7.34	2.95 / -5.68	2.32 / -5.88	1.94 / -4.42	2.11 / -3.80	1.33 / -9.24	2.24 / -6.06
		Chen et al.	1.81 / -14.33	2.02 / -11.02	2.01 / -12.60	1.64 / -9.93	1.80 / -10.89	1.22 / -15.31	1.75 / -12.35
		Ours	<b>1.77 / -18.19</b>	<b>1.79 / -17.30</b>	<b>1.18 / -14.88</b>	<b>1.58 / -14.91</b>	<b>1.47 / -15.22</b>	<b>1.02 / -18.77</b>	<b>1.47 / -16.55</b>
	DCVC-DC	Zhang et al.	2.25 / -6.50	2.08 / -5.24	1.74 / -4.33	1.62 / -4.20	2.03 / -3.99	1.37 / -8.17	1.85 / -5.41
		Chen et al.	1.69 / -9.99	1.81 / -9.88	1.72 / -11.13	1.54 / -9.41	1.66 / -9.89	1.35 / -12.19	1.62 / -10.42
		Ours	<b>1.40 / -13.64</b>	<b>1.52 / -12.55</b>	<b>0.95 / -14.93</b>	<b>0.92 / -12.88</b>	<b>0.91 / -13.05</b>	<b>0.83 / -16.52</b>	<b>1.09 / -13.93</b>
	DCVC-RT	Zhang et al.	—	—	—	—	—	—	—
		Chen et al.	1.33 / -5.67	1.50 / -5.11	1.43 / -6.31	1.25 / -4.40	1.36 / -4.76	1.02 / -6.06	1.32 / -5.39
		Ours	<b>1.02 / -6.37</b>	<b>1.25 / -6.15</b>	<b>0.91 / -6.27</b>	<b>0.87 / -5.13</b>	<b>0.96 / -5.09</b>	<b>0.85 / -7.14</b>	<b>0.98 / -6.03</b>

Table 1: Performance comparison across datasets ( $\Delta R \downarrow$  / BD-Rate (%) $\downarrow$ ) with average performance. Bold values indicate the best performance. (Due to the lack of results at the GOP size of 32 in Zhang et al. (2023) and with DCVC-RT, where we have provisionally excluded comparisons with Zhang et al. (2023) to prevent potential discrepancies.)

This poses a fundamental exploration and exploitation dilemma: overly dense, per-frame rewards may accelerate convergence but discourage discovery of better global strategies; overly sparse rewards leave instant frames without guidance (Bellemare et al. 2016; Saunders et al. 2017). To strike a balance, we reshape the reward as a weighted inner product of distortion and rate deviation terms:

$$r_t = -\mathbf{w}_t^\top \mathbf{f}_t, \quad \mathbf{f}_t = \begin{pmatrix} D_t \\ |R_{rem}| \\ R_{tar} \end{pmatrix}, \quad \mathbf{w}_t = \begin{pmatrix} \delta_t \\ \eta_t \end{pmatrix}. \quad (8)$$

where  $R_{rem}$  is the remaining bitrate budget. The weight vector  $\mathbf{w}_t = (\delta, \eta)^\top$  balances distortion and rate accuracy, and is periodically updated every  $\mathcal{K}$  training steps using validation feedback. For the final frame, a large  $\eta_t$  is applied to enforce strict rate control. To prevent overspending the bitrate budget, over-allocation is penalized accordingly.

Furthermore, We adopt a twin-Critic architecture where two independent Q-values are estimated, and their minimum is used to mitigate overestimation bias (Hasselt 2010). In addition, we model the full return distribution instead of a scalar expected value, enhancing robustness in reward estimation (Bellemare, Dabney, and Munos 2017).

Unlike previous schemes that rely on fixed allocation rules or handcrafted heuristics, our proposed RL-based framework adaptively shapes rewards, improving the learned policy’s generalization and effectiveness. Implementation details and related hyperparameters are available in the Appendix.

## 5 Experiments

### 5.1 Experimental Setup

**Base Codecs:** We perform evaluation on four representative NVCs: DVC (Lu et al. 2019), DCVC (Li, Li, and Lu

2021), DCVC-DC (Li, Li, and Lu 2023) and the latest DCVC-RT (Jia et al. 2025). Since DVC and DCVC are fixed-rate, we extend their highest bitrate pretrained models to support variable-rate coding according to Duan et al. (2023), with their original training setups. For DCVC-DC and DCVC-RT, we directly adopt their pre-trained variable-rate model. To achieve flexible rate control, we set  $m_t \in [0.5, 1.0]$ ,  $\lambda_t \in [256, 2048]$  for DVC and DCVC,  $\lambda_t \in [85, 840]$  for DCVC-DC, and  $\lambda_t \in [1, 768]$  for DCVC-RT according to their original setup. Details can be found in the Appendix.

**Datasets:** For training our rate control module (parameters of codecs remain fixed), we construct a mixed dataset combining BVI-DVC (Ma, Zhang, and Bull 2021) and Vimeo sequences (Xue et al. 2019). For validation, we use the USTC-TD dataset (Li et al. 2024). For evaluation, we follow standard benchmarks, selecting UVG dataset (Mercat, Viitanen, and Vanne 2020), MCL-JCV dataset (Wang et al. 2016), and HEVC Class B to E sequences (Bossen et al. 2013).

**Implementation Details:** All experiments are implemented using the PyTorch framework on an NVIDIA RTX 3090 GPU. To enhance sample efficiency during RL training, we adopt a replay buffer of length 200 for offline updates, sampling 32 trajectories per iteration. Initially, all networks are pretrained for 50 epochs using 4-frame sequences. The feature extractor is then fixed, and training continues for an additional 250 epochs with 32-frame sequences. Further training strategies and details are provided in the Appendix.

**RC Benchmarks:** We compare our approach with two state-of-the-art methods: (i) Chen et al. (2023), which models the R- $\lambda$ -m and D- $\lambda$ -m relationships using a hyperbolic function with iterative updates; and (ii) Zhang et al. (2023), which employs a neural network to predict rate allocation and the R- $\lambda$  mapping. For a fair comparison, we adopt the

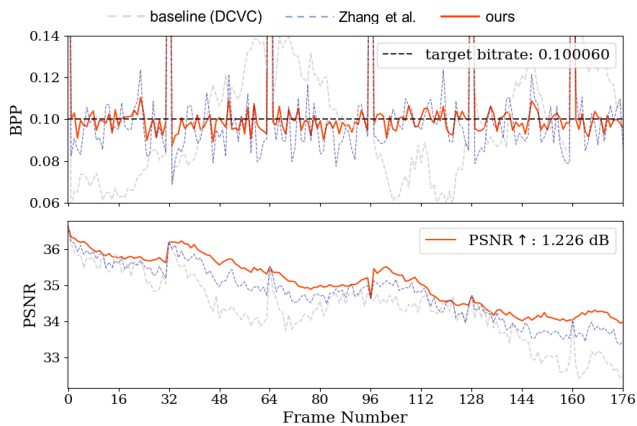


Figure 4: Frame-Level Rate Control. Evaluated with a unified target bitrate  $0.10006$  BPP on *BasketballDrive* sequence.

same test conditions across all methods—for example, evaluating all frames and using a GOP size of 32 or 100 under an LDP configuration. Notably, our approach is intended to explore a general, plug-and-play rate-control method for NVC; schemes designed for traditional codecs, or those tailored to specific architectures or other objectives, are not within the scope of comparisons. Details can be found in the Appendix.

## 5.2 Experimental Results

**Performance Analysis:** As shown in Table 1, our RL-based method demonstrates superior performance in both rate control accuracy and  $R$ - $D$  performance across various GOP sizes. For a GOP size of 32, it achieves lower rate errors across almost all evaluated codecs and datasets, along with higher BD-Rate gains. Even when applied to NVC frameworks with complex inter-frame dependencies, such as DCVC-DC, our method achieves a rate error of just 1.13% and a bitrate saving of 13.98%, verifying the effectiveness of the proposed RL-based scheme. In DCVC-RT-whose native models are trained with hierarchical quality on long sequences and thus inherently perform implicit rate allocation to enhance  $R$ - $D$  performance—the integration of an additional rate control module brings only marginal improvements. Nevertheless, our method maintains a low rate error of just 1.15%, indicating strong robustness. For a longer GOP size of 100, our approach continues to demonstrate clear advantages. Notably, as the underlying codec improves (e.g., from DVC to DCVC-RT), the BD-Rate gains achieved by the method of Zhang et al. (2023) diminish significantly—shrinking to just 5.41% on DCVC-DC—whereas our approach maintains a stable 13.93% gain, highlighting its superior generalization capability.

To further evaluate the effectiveness of our RL-based scheme, we visualize per-frame performance and compare it with the method of Zhang et al. (2023) on DCVC. Specifically, (i) *Fixed Target Bitrate*: We encode the entire sequence under a constant bitrate constraint to simulate stable network conditions. As shown in Fig. 4, our method exhibits significantly lower rate fluctuation and reduced quality degradation across frames. (ii) *Dynamic Network Bandwidth*: In real-world streaming scenarios, network bandwidth varies over

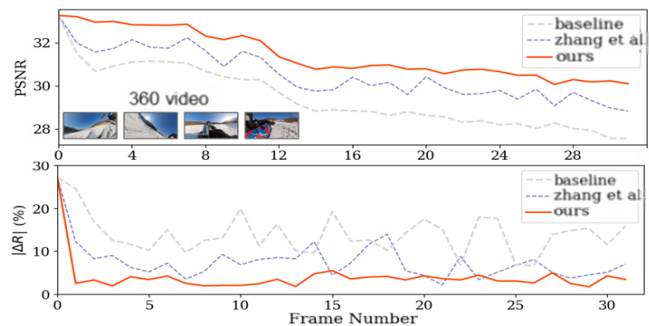


Figure 5: Performance Comparison over 360° Video. The test sequence is downloaded from <https://www.youtube.com/watch?v=4T8yFnHaJtc>, with the results of quality degradation above and rate fluctuation below.

short time intervals, necessitating adaptive rate allocation. To simulate this, we use real-life bandwidth traces from Federal Communications Commission (FCC) (2023) to set target bitrate. As illustrated in Fig. 1(b), our method adapts more effectively to bandwidth fluctuations, achieving smoother bitrate transitions and more stable quality. In contrast, other methods struggle to cope with such variations. This finding also supports our discussion in Sec.3.2—the necessity to consider both frames contents characteristics and inter-frame rate and distortion dependencies in rate control of NVC.

**Generalization Analysis:** An effective rate control scheme must generalize well to diverse video contents and motion patterns. To assess this, we evaluate our method on an unseen 360-degree video sequence, which exhibits substantially greater content variability and motion dynamics than those seen during training. As shown in Fig. 5, the naive baseline suffers from significant quality decline and pronounced rate fluctuations. While Zhang et al. (2023) partially alleviates the rate fluctuation, it still exhibits a PSNR drop exceeding 4.4 dB and a rate bias of approximately 7.6%. In contrast, our method demonstrates a smaller PSNR degradation and maintains a lower rate bias of 3.9%. These results underscore the superior generalization capability of our approach. This improvement is attributed to our balanced exploration-exploitation strategy—enabled by random action sampling and stable reward feedback—which facilitates a more adaptive and robust policy when encountering volatile content distributions. Additional generalization comparisons with other methods are provided in the Appendix.

**Complexity Analysis:** We compare the computational complexity of our RL-based scheme with existing methods (Li et al. 2022; Zhang et al. 2023; Chen et al. 2023), using the real-time NVC method DCVC-RT as the baseline. The evaluation considers multiple metrics, including network parameter count ( $M$ ), computational cost measured in KMACs per pixel, memory usage (GB), and encoding/decoding throughput (FPS, Frames Per Second). Benefited from the lightweight network design, our proposed method introduces minimal computational overhead—only an additional 0.57 M parameters, 1.60 KMACs per pixel, and 0.33 GB of memory compared to the baseline. Furthermore, by incorporating a down-sampling operation, our method can even improve

Method	Params.	Complexity		Throughput (FPS)	
		KMACs/pxl	Mem.	Enc.	Dec.
Baseline	66.33	421.31	2.27	102	95
Zhang et al.	+2.12	+6.40	+1.22	68	-
Chen et al.	-	-	-	54	108
Ours	+0.57	+1.60	+0.33	<b>111</b>	<b>109</b>

“—” indicates no change compared to the baseline.

Table 2: Complexity Comparison over DCVC-RT

encoding and decoding throughput. In contrast, other methods tend to compromise real-time performance due to either heavy network architectures (Zhang et al. 2023) or additional pre-coding steps (Chen et al. 2023). Although Chen et al. (2023) also adopts a down-sampling strategy, their method requires pre-coding equidistant frames with all candidate parameters to initialize the model, which incurs substantial extra encoding time. These results highlight the efficiency and practicality of our approach, demonstrating its potential for real-world deployment without sacrificing performance.

### 5.3 Ablation Studies

To further understand the effectiveness of our scheme, we conduct a series of in-depth evaluations. Unless otherwise specified, DCVC is used as the baseline, and comparisons are made with the learned method (Zhang et al. 2023).

**Training Frame Numbers:** To evaluate our scheme’s ability to capture inter-frame dependencies, we train the model with varying frame numbers. As shown in Table 3, increasing training frame numbers steadily improves both R-D performance and rate control accuracy. In contrast, Zhang et al. (2023) reports no clear benefits from longer training sequences in their own ablation study. A potential reason is that the deviation caused by not considering inter-frame rate dependencies gradually as frame numbers increase.

By contrast, our method models inter-frame dependencies in a frame-wise manner, which accurately models per-frame rate and distortion dependencies, naturally scaling to longer sequences. Furthermore, as shown in Fig. 6, our approach maintains linear training complexity with respect to sequence length, ensuring both computational efficiency and feasibility.

	4	8	16	32	64
BD-Rate (%)	-8.84	-11.15	-15.03	-16.49	-16.90
$\Delta R$ (%)	2.48	1.82	1.67	1.48	1.43

Table 3: Results with Different Training Frame Numbers

#### Performance Under the Setup of Zhang et al. (2023):

In Zhang et al. (2023), the mini-GOP size is fixed at 4, and only the Lagrange multiplier  $\lambda_t$  is used for rate control. To ensure a fair comparison, we replicate this setup. As shown in Table 4, our method consistently outperforms Zhang et al. (2023) across multiple datasets, demonstrating its superiority brought about by considering inter-frame rate and distortion dependencies, even under identical setups.

**Impact of Down-Sampling Factors  $m_t$ :** We further investigate the contribution of jointly deciding both  $\lambda_t$  and the down-sampling factor  $m_t$ . As shown in Table 5, while both

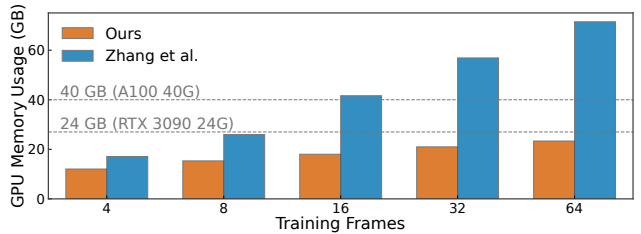


Figure 6: Comparison over the GPU Memory Usage with Increasing Training Frames. The memories of RTX 3090 and A100 are marked with dashed lines.

Metric	UVG	MCL-JCV	Cls.B	Cls.C	Cls.D	Cls.E
BD-Rate (Zhang et al.)	-7.34	-5.68	-5.88	-4.42	-3.80	-9.24
BD-Rate (Ours)	-7.74	-6.60	-7.32	-5.22	-4.16	-11.05
$\Delta R$ (Zhang et al.)	2.80	2.95	2.32	1.94	2.11	1.33
$\Delta R$ (Ours)	2.58	2.44	1.98	1.91	1.80	1.07

Table 4: Comparison with Zhang et al. (2023) with Only  $\lambda_t$

strategies achieve accurate rate control, using only  $\lambda_t$  yields limited PSNR gains, particularly at high bitrates. In contrast, jointly optimizing  $\lambda_t$  and  $m_t$  enables a richer R–D trade-off space, leading to significantly improved overall performance.

$\lambda_t$	Baseline		w/ $\lambda_t$		w/ $m_t$ and $\lambda_t$	
	BPP	PSNR	BPP	PSNR	BPP	PSNR
256	0.0251	33.502	0.0249	34.138	0.0250	34.306
512	0.0489	35.329	0.0489	35.582	0.0488	35.827
1024	0.0710	36.248	0.0709	36.344	0.0711	36.560
2048	0.1006	37.028	0.1005	37.052	0.1001	37.245

Table 5: Ablation Results Regarding The Impact of  $m_t$

## 6 Conclusion

In this paper, we revisited the rate control problem in NVC and highlighted the importance of jointly modeling inter-frame rate and distortion dependencies. To this end, we proposed a reinforced rate-control framework that accurately models these environmental conditions as states, learns a dynamic policy to directly map them to per-frame coding parameters, and optimizes via expected cumulative discounted return. Extensive experiments—covering multiple codecs and diverse settings—show that explicitly incorporating inter-frame rate and distortion dependencies significantly reduces rate error, enhances R-D performance, improves generalization, and maintains low computational complexity. These advancements position our method as a practical solution for real-world NVC deployments, particularly in bandwidth-constrained or dynamic network environments. Future work will integrate network transmission conditions (e.g., packet loss and congestion) to develop network-aware rate control schemes for NVC that jointly optimize performance and network adaptability.

## Acknowledgments

This work was supported in part by Natural Science Foundation of Jiangsu Province (Grant No. BK20241226, BK20243038) and Natural Science Foundation of China (Grant No. 62401251, 62431011, 62471215, 62231002). The authors would like to express their sincere gratitude to the Interdisciplinary Research Center for Future Intelligent Chips (Chip-X) and Yachen Foundation for their invaluable support.

## References

- Alexandre, D.; Hang, H.-M.; and Peng, W.-H. 2022. Two-layer learning-based p-frame coding with super-resolution and content-adaptive conditional anf. In *Proceedings of the 4th ACM International Conference on Multimedia in Asia*, 1–7.
- Altman, E. 2021. *Constrained Markov decision processes*. Routledge.
- Bellemare, M.; Srinivasan, S.; Ostrovski, G.; Schaul, T.; Sutton, D.; and Munos, R. 2016. Unifying count-based exploration and intrinsic motivation. *Advances in neural information processing systems*, 29.
- Bellemare, M. G.; Dabney, W.; and Munos, R. 2017. A distributional perspective on reinforcement learning. In *International conference on machine learning*, 449–458. PMLR.
- Bossen, F.; et al. 2013. Common test conditions and software reference configurations. *JCTVC-L1100*, 12(7): 1.
- Bross, B.; Wang, Y.-K.; Ye, Y.; Liu, S.; Chen, J.; Sullivan, G. J.; and Ohm, J.-R. 2021. Overview of the versatile video coding (VVC) standard and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(10): 3736–3764.
- Chen, J.; Wang, M.; Zhang, P.; Wang, S.; and Wang, S. 2023. Sparse-to-Dense: High Efficiency Rate Control for End-to-end Scale-Adaptive Video Coding. *IEEE Transactions on Circuits and Systems for Video Technology*.
- Chen, L.-C.; Hu, J.-H.; and Peng, W.-H. 2018. Reinforcement learning for HEVC/H. 265 frame-level bit allocation. In *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)*, 1–5. IEEE.
- Chen, Z.; Zhou, L.; Hu, Z.; and Xu, D. 2024. Group-aware parameter-efficient updating for content-adaptive neural video compression. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 11022–11031.
- Duan, Z.; Lu, M.; Ma, J.; Huang, Y.; Ma, Z.; and Zhu, F. 2023. Qarv: Quantization-aware resnet vae for lossy image compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Federal Communications Commission (FCC). 2023. Measuring Broadband America: Raw Data Releases. <https://www.fcc.gov/oet/mba/raw-data-releases>.
- Gadot, U.; Shocher, A.; Mannor, S.; Chechik, G.; and Hallak, A. 2025. RL-RC-DoT: A Block-level RL agent for Task-Aware Video Compression. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 12533–12542.
- Gu, B.; Chen, H.; Lu, M.; Yao, J.; and Ma, Z. 2024. Adaptive Rate Control for Deep Video Compression with Rate-Distortion Prediction. arXiv:2412.18834.
- Haarnoja, T.; Zhou, A.; Abbeel, P.; and Levine, S. 2018. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, 1861–1870. PMLR.
- Haarnoja, T.; Zhou, A.; Hartikainen, K.; Tucker, G.; Ha, S.; Tan, J.; Kumar, V.; Zhu, H.; Gupta, A.; Abbeel, P.; and Levine, S. 2019. Soft Actor-Critic Algorithms and Applications. arXiv:1812.05905.
- Hasselt, H. 2010. Double Q-learning. *Advances in neural information processing systems*, 23.
- Ho, Y.-H.; Jin, G.-L.; Liang, Y.; Peng, W.-H.; and Li, X. 2021a. A dual-critic reinforcement learning framework for frame-level bit allocation in HEVC/H. 265. In *2021 Data compression conference (DCC)*, 13–22. IEEE.
- Ho, Y.-H.; Jin, G.-L.; Liang, Y.; Peng, W.-H.; and Li, X. 2021b. A Dual-Critic Reinforcement Learning Framework for Frame-Level Bit Allocation in HEVC/H.265. In *2021 Data Compression Conference (DCC)*, 13–22.
- Hu, S.; Wang, H.; Kwong, S.; Zhao, T.; and Kuo, C.-C. J. 2011. Rate control optimization for temporal-layer scalable video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(8): 1152–1162.
- Hu, Z.; Lu, G.; and Xu, D. 2021. FVC: A new framework towards deep video compression in feature space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1502–1511.
- Jia, Z.; Li, B.; Li, J.; Xie, W.; Qi, L.; Li, H.; and Lu, Y. 2025. Towards practical real-time neural video compression. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 12543–12552.
- Konda, V.; and Tsitsiklis, J. 1999. Actor-critic algorithms. *Advances in neural information processing systems*, 12.
- Li, B.; Li, H.; Li, L.; and Zhang, J. 2014.  $\lambda$  domain rate control algorithm for High Efficiency Video Coding. *IEEE Transactions on Image Processing*, 23(9): 3841–3854.
- Li, J.; Li, B.; and Lu, Y. 2021. Deep Contextual Video Compression. *Advances in Neural Information Processing Systems*, 34.
- Li, J.; Li, B.; and Lu, Y. 2023. Neural Video Compression with Diverse Contexts. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, Canada, June 18-22, 2023*.
- Li, J.; Li, B.; and Lu, Y. 2024. Neural Video Compression with Feature Modulation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 17-21, 2024*.
- Li, L.; Walsh, T. J.; and Littman, M. L. 2006. Towards a unified theory of state abstraction for MDPs. *AI&M*, 1(2): 3.
- Li, Y.; Chen, X.; Li, J.; Wen, J.; Han, Y.; Liu, S.; and Xu, X. 2022. Rate control for learned video compression. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2829–2833. IEEE.

- Li, Y.; Liu, Z.; Chen, Z.; and Liu, S. 2020. Rate control for versatile video coding. In *2020 IEEE International Conference on Image Processing (ICIP)*, 1176–1180. IEEE.
- Li, Z.; Liao, J.; Tang, C.; Zhang, H.; Li, Y.; Bian, Y.; Sheng, X.; Feng, X.; Li, Y.; Gao, C.; et al. 2024. USTC-TD: A Test Dataset and Benchmark for Image and Video Coding in 2020s. *arXiv preprint arXiv:2409.08481*.
- Liang, X.; Wang, Q.; Zhou, Y.; Luo, B.; and Men, A. 2013. A novel RQ model based rate control scheme in HEVC. In *2013 Visual Communications and Image Processing (VCIP)*, 1–6. IEEE.
- Liu, H.; Lu, M.; Chen, Z.; Cao, X.; Ma, Z.; and Wang, Y. 2022. End-to-end neural video coding using a compound spatiotemporal representation. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(8): 5650–5662.
- Liu, H.; Lu, M.; Ma, Z.; Wang, F.; Xie, Z.; Cao, X.; and Wang, Y. 2020. Neural video coding using multiscale motion compensation and spatiotemporal context model. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(8): 3182–3196.
- Lu, G.; Cai, C.; Zhang, X.; Chen, L.; Ouyang, W.; Xu, D.; and Gao, Z. 2020. Content adaptive and error propagation aware deep video compression. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, 456–472. Springer.
- Lu, G.; Ouyang, W.; Xu, D.; Zhang, X.; Cai, C.; and Gao, Z. 2019. Dvc: An end-to-end deep video compression framework. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11006–11015.
- Lu, M.; Duan, Z.; Zhu, F.; and Ma, Z. 2024. Deep Hierarchical Video Compression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 8859–8867.
- Ma, D.; Zhang, F.; and Bull, D. R. 2021. BVI-DVC: A training database for deep video compression. *IEEE Transactions on Multimedia*, 24: 3847–3858.
- Mandhane, A.; Zhernov, A.; Rauh, M.; Gu, C.; Wang, M.; Xue, F.; Shang, W.; Pang, D.; Claus, R.; Chiang, C.-H.; et al. 2022. Muzero with self-competition for rate control in vp9 video compression. *arXiv preprint arXiv:2202.06626*.
- Mao, H.; Gu, C.; Wang, M.; Chen, A.; Lazic, N.; Levine, N.; Pang, D.; Claus, R.; Hechtman, M.; Chiang, C.-H.; et al. 2020. Neural rate control for video encoding using imitation learning. *arXiv preprint arXiv:2012.05339*.
- Mercat, A.; Viitanen, M.; and Vanne, J. 2020. UVG dataset: 50/120fps 4K sequences for video codec analysis and development. In *Proceedings of the 11th ACM Multimedia Systems Conference*, 297–302.
- Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A. A.; Veness, J.; Bellemare, M. G.; Graves, A.; Riedmiller, M.; Fidjeland, A. K.; Ostrovski, G.; et al. 2015. Human-level control through deep reinforcement learning. *nature*, 518(7540): 529–533.
- Saunders, W.; Sastry, G.; Stuhlmüller, A.; and Evans, O. 2017. Trial without error: Towards safe reinforcement learning via human intervention. *arXiv preprint arXiv:1707.05173*.
- Sheng, X.; Li, L.; Liu, D.; and Li, H. 2024. Prediction and Reference Quality Adaptation for Learned Video Compression. *arXiv:2406.14118*.
- Sullivan, G. 1998. Rate-Distortion Optimization for Video Compression. *IEEE Signal Processing Magazine*, 84.
- Sullivan, G. J.; Ohm, J.-R.; Han, W.-J.; and Wiegand, T. 2012. Overview of the high efficiency video coding (HEVC) standard. *IEEE Transactions on circuits and systems for video technology*, 22(12): 1649–1668.
- Tang, C.; Sheng, X.; Li, Z.; Zhang, H.; Li, L.; and Liu, D. 2024. Offline and online optical flow enhancement for deep video compression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 5118–5126.
- Wang, H.; Gan, W.; Hu, S.; Lin, J. Y.; Jin, L.; Song, L.; Wang, P.; Katsavounidis, I.; Aaron, A.; and Kuo, C.-C. J. 2016. MCL-JCV: a JND-based H. 264/AVC video quality assessment dataset. In *2016 IEEE international conference on image processing (ICIP)*, 1509–1513. IEEE.
- Wang, S.; Ma, S.; Wang, S.; Zhao, D.; and Gao, W. 2013a. Quadratic  $\rho$ -domain based rate control algorithm for HEVC. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1695–1699. IEEE.
- Wang, S.; Ma, S.; Wang, S.; Zhao, D.; and Gao, W. 2013b. Rate-GOP based rate control for high efficiency video coding. *IEEE Journal of selected topics in signal processing*, 7(6): 1101–1111.
- Wiegand, T.; Sullivan, G. J.; Bjontegaard, G.; and Luthra, A. 2003. Overview of the H. 264/AVC video coding standard. *IEEE Transactions on circuits and systems for video technology*, 13(7): 560–576.
- Xue, T.; Chen, B.; Wu, J.; Wei, D.; and Freeman, W. T. 2019. Video Enhancement with Task-Oriented Flow. *International Journal of Computer Vision (IJCV)*, 127(8): 1106–1125.
- Yang, J.; Liu, M.; Hou, P.; Xu, Y.; and Sun, J. 2025. Neural Adaptive Contextual Video Streaming. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.
- Zhang, C.; and Gao, W. 2024. Learned Rate Control for Frame-Level Adaptive Neural Video Compression via Dynamic Neural Network. In *European Conference on Computer Vision*. Springer.
- Zhang, Y.; Lu, G.; Chen, Y.; Wang, S.; Shi, Y.; Wang, J.; and Song, L. 2023. Neural Rate Control for Learned Video Compression. In *The Twelfth International Conference on Learning Representations*.
- Zhou, M.; Wei, X.; Kwong, S.; Jia, W.; and Fang, B. 2020. Rate control method based on deep reinforcement learning for dynamic video sequences in HEVC. *IEEE Transactions on Multimedia*, 23: 1106–1121.
- Zhou, M.; Wei, X.; Kwong, S.; Jia, W.; and Fang, B. 2021. Rate Control Method Based on Deep Reinforcement Learning for Dynamic Video Sequences in HEVC. *IEEE Transactions on Multimedia*, 23: 1106–1121.