applied sciences

MDPI

Article

Region-of-Interest Based Coding Scheme for Live Videos

Xiuxin Dou [1,2,*], Xixin Cao [1] and Xianguo Zhang [2]

[1] School of Software and Microelectronics, Peking University, Beijing 100871, China; cxx@ss.pku.edu.cn
[2] Shannon Lab, Cloud Architecture and Platform Department, Technology and Engineering Group, Tencent, Shenzhen 518054, China
* Correspondence: douxiuxin@pku.edu.cn

Abstract: In this paper, we introduce a novel rate control scheme specifically tailored for live broadcasting scenarios. Notably, in high-definition live transmissions of sports events and video game competitions that typically exceed 1080 p resolution and run at frame rates of 60 fps or higher, the transcoding speed of encoders often becomes a limiting factor, leading to streams with substantial bitrates but unsatisfactory quality metrics. To enhance the overall Quality of Service (QoS) without increasing the bitrate, it is essential to improve the quality of Regions of Interest (ROI).Our proposed solution presents an ROI-based rate reservoir model that ingeniously leverages Convolutional Neural Networks (CNNs) to predict rate control parameters. This approach aims to optimize the bitrate allocation within high bitrate live broadcasts, thus enhancing the image quality within ROIs. Experimental outcomes demonstrate that this algorithm manages to increase the bitrate by no more than 5%, effectively redistributing the reduced bitrate across the entire Group of Pictures (GOP). As a result, it ensures a gradual decrease in the quality of Regions of Uninterest (ROU), thereby maintaining a balanced quality experience throughout the broadcasted content.

Keywords: video coding; rate control; ROI; live broadcast; high bit rate

Citation: Dou, X.; Cao, X.; Zhang, X. Region-of-Interest Based Coding Scheme for Live Videos. *Appl. Sci.* **2024**, *14*, 3823. https://doi.org/10.3390/app14093823

Academic Editor: Chilukuri K. Mohan

Received: 1 February 2024
Revised: 22 March 2024
Accepted: 25 March 2024
Published: 30 April 2024

1. Introduction

With the burgeoning growth of the live video industry, there is an escalating demand for efficient video encoding that ensures quality of service (QoS). This need is particularly pronounced in high-resolution and high-frame-rate broadcasts of sports events and electronic games.It is a well-established fact that human viewers do not uniformly distribute their attention across the entire visual scene but rather focus on specific areas that are critical to their perception tasks. These areas, known as Regions of Interest (ROI), play a key role in influencing perceptual video quality and subsequent viewer analysis, as evidenced by studies such as [1]. To enhance the perceived video quality at constant bitrates, aligning with the inherent heuristics of the Human Visual System (HVS), our approach advocates allocating more bits to Coding Units (CU) within the ROI compared to those in the Region of Uninterest (ROU). By doing so, we can optimize the coding process to better serve the viewer's perceptual priorities while maintaining overall bitrate efficiency [2–6].

There have been several methods proposed for ROI-based video coding. In the work of [7], an adjustable quality ROI-centric rate control scheme is introduced, which adopts the same quadratic model implemented in H.264/AVC for calculation purposes. This algorithm assigns a quantitative parameter (QP) to each area based on user-selected quality levels and subsequently calculates the QP for every macroblock using the quadratic model as seen in [8]. The QP values are then adjusted according to an input ROI map and the bit allocation for each region. Doulamis et al. [9] employed neural networks to detect ROIs and allocate a higher number of bits to these areas. The rate control mechanism in [10] utilizes the linear Rate-Quantization (R-Q) model to determine the stream's bit allocation. It employs the Viola Jones face detector to identify ROIs and assigns different QPs to both ROI and Region of Uninterest (ROU), maintaining a constant QP difference between them.

The research in [11] presents an ROI-based rate control method specifically designed for High Efficiency Video Coding (HEVC). This approach involves independent processing at the Coding Unit (CU) level for the two regions and applies a larger QP clipping range while adhering to the global bitrate constraint. In Ping-Hao Wu et al.'s study [12], a dual encoding system is utilized where a "basic encoder" encodes low-resolution full views, and a separate "ROI encoder" focuses on high-resolution regions of interest. The bit allocation for the ROI encoder is determined based on the distortion metrics obtained from the corresponding region in the basic encoder. The paper [13] proposes an unequal error protection scheme for compressed HEVC video bitstreams that prioritizes ROI data. Meanwhile, ref. [14] optimizes the quantization parameters for both ROI and ROU source videos by minimizing an objective function, which are then used during the encoding process of the synthesized video. Ref. [15] discusses the ROI location coding tool adopted in the surveillance profile of the AVS2 video coding standard, illustrating three distinct coding schemes: direct-coding, temporal differential-coding, and reconstructed-coding. Ref. [16] realizes a more efficacious distribution of bitrate within machine learning-based image encoders, attenuating the quantization factors in the background regions of the images, thus diminishing the overall level of quantization and consequentially reducing the bitrate assigned to such areas, while concurrently preserving the quality of the ROI sections that are critical for machine learning operations. Ref. [17] proposes a Transformer-based variable-rate image compression system that can achieve variable rate compression with a single model while supporting ROI functionality, introducing a Prompt Generation Network to condition the Transformer Autoencoder's compression process. Lastly, ref. [18] constructs a framework based on FFmpeg and the X.264 codec, integrating a Yolov7 detection model for ROI extraction from video frames, which subsequently generates corresponding masks. This setup prioritizes bitrate allocation to ROI, sacrificing background bitrate allocation in order to maintain overall bitrate stability.

The primary aim of rate control in video compression is to attain the lowest possible quality distortion within a predefined bit rate constraint. In this context, rate control commonly integrates Rate-Distortion Optimization (RDO) as a key strategy. To arrive at the most optimized coding mode decision for each Coding Unit (CU), a Lagrange multiplier λ is employed as per Equation (1). This multiplier links the distortion metric D (which is directly related to the Quantization Parameter, QP) with the number of bits R (also associated with QP). By doing so, it facilitates the evaluation of all feasible coding modes and selects the one that minimizes the overall cost function. Thus, rate control fundamentally ensures an effective trade-off between visual quality and bitrate efficiency.

$$cost = D + \lambda \times R \tag{1}$$

All the previously developed rate control (RC) algorithms in HEVC have not adequately considered the varying importance of different regions within a video frame. In response, we propose a novel rate control scheme tailored for high bit rate live streaming systems that separately addresses Regions of Interest (ROI) and Regions of Uninterest (ROU). Given that the reference software HM's encoding speed is insufficient to meet the real-time requirements of live streaming transcoding, our proposed algorithm builds upon the open-source encoding software x265 . The algorithm introduces three distinctive features: Firstly, it employs the open-source mm-detection [19] tool to automatically detect ROI in each video frame and generate an ROI map accordingly. And mm-detection is an object detection toolbox that contains a rich set of object detection, instance segmentation, and panoptic segmentation methods as well as related components and modules. Secondly, we introduce a reservoir model to predict and supervise the overall bitrate increase, enabling real-time adjustment of rate control allocation bits for both ROI and ROU regions. Lastly, we allocate target bitrates to each ROI and ROU region with a focus on maintaining the stability of the live stream size; recognizing the multitude of factors affecting rate stability, we propose a Convolutional Neural Network (CNN) model to predict joint rate control parameters for ROI and ROU.

The paper is structured as follows: Section 1 presents an overview of the general rate control problem and existing HEVC solutions. Subsequently, Section 2 elucidates the details of our proposed rate control approach. The experimental evaluation and results are discussed in Section 3. Finally, Section 4 offers a thorough discussion of the findings, while Section 5 concludes the paper with summarizing remarks and future directions.

2. Related Works

The objective of video encoding rate control is to closely approximate a predefined constant target bit rate while minimizing quality degradation. Central to this task, the rate control algorithm seeks to determine the most suitable Quantization Parameter (QP) for each video segment under the condition that the encoded bitrate $R_v(QP)$ does not exceed the maximum allowed limit, denoted by R_{max}. This constraint is crucial because quantization inherently compresses video signals at reduced bitrates, with R_{max} being the fixed upper bound on the total number of bits, and $R_v(QP)$ representing the actual number of encoded bits in the live stream.

In the context of video coding, rate control typically integrates Rate-Distortion Optimization (RDO), as described in Equation (1). Given the QP assigned by the rate control process, RDO strives to minimize the cost function within each Coding Unit (CU), thereby reducing the overall cost of the encoded video.

To address these challenges, an explicit Rate-Distortion model is required to correlate the average bitrate with the QP value. A variety of studies have been conducted to develop such models that associate perceived video quality with bitrate. Various rate models have emerged, ranging from simple linear expressions to more intricate mathematical formulations. For instance, the reference software for HEVC encoding in [20] employs a linear model for bit rate estimation. In contrast, the quadratic model is often represented as follows:

$$R = C_1 \times MAD/QP + C_1 \times MAD/QP^2 \quad (2)$$

where C_1 and C_2 are the model parameters, has been adopted in VM8 for MPEG4 [21], H.264/AVC [22] and also for HEVC [23].

These models play a pivotal role in achieving the delicate balance between video quality and bitrate constraints in video encoding applications.

As previously mentioned, each rate control model is tailored for video encoding systems operating under specific conditions. However, the core objective of all these methods is to allocate an optimal number of bits and determine the quantization parameters for every coding unit. The encoder's rate controller operates across three principal levels: Group of Pictures (GOP), frame, and Coding Unit (CU) [12]:

i. GOP Level: At this level, input parameters include the target bit rate, sequence frame rate, GOP size, and virtual buffer occupancy. The rate control algorithm calculates the average number of bits per GOP.
ii. Frame Level: This stage involves considering the average bit allocation per frame, where a fixed target bit amount is set for the current frame. Bit allocation takes into account the hierarchical structure of the frames, followed by the application of the $R - \lambda$ model to compute the frame-level QP.
iii. CU Level: The process unfolds in three stages. Firstly, the required bit allocation for each CU is calculated based on the frame budget and the Mean Absolute Difference (MAD) of that CU. Secondly, using the established frame budget, the $R - \lambda$ model is employed to calculate a λ value within a fixed QP range for each CU. Finally, Rate-Distortion Optimization (RDO) is performed to find the optimal mode decision, which refers back to the given QP.

In summary, the rate control mechanism at these different levels ensures a systematic and adaptive approach to bitrate management while maintaining optimal video quality throughout the encoding process.

3. Proposed Approach

Our proposed method is grounded on the R-λ_{cu} model tailored for HEVC. Building upon the GOP rate control strategy, our scheme introduces a joint optimization and control mechanism at both frame level and Coding Unit (CU) level, thereby enhancing the subjective quality of Regions of Interest (ROI) while maintaining overall bitrate stability. The relationship between R and λ described earlier is leveraged to compute the Quantization Parameter (QP) for each frame and individual CU within an image. This approach has demonstrated superior performance compared to traditional quadratic models [7,8]. In this section, we delve into the details of our proposed methodology, with a particular focus on the two steps in the ROI rate control process: Firstly, utilizing known information coupled with a Convolutional Neural Network (CNN) model to infer the λ_{cu} values for both ROI and Region of Uninterest (ROU); Secondly, discussing methods for gathering training data to effectively train the CNN model.

3.1. Region λ Infer

Before implementing ROI-based rate control, we establish a dedicated ROI rate reservoir R, which serves to track the bitrate consumption in both the ROI and ROU regions. As depicted in Figure 1, this ROI reservoir operates independently of the reservoir used by the original codec's rate control mechanism. In our algorithm, as more bits are allocated to the ROI area, causing an increase in bit rate ΔR_i, the water level in the ROI reservoir rises correspondingly. Conversely, when fewer bits ($-\Delta R_u$) are consumed in the ROU region, the water level in the ROI reservoir decreases.

The video encoding process commences with the first step of utilizing mm-detection for object detection to identify and gather information on the ROI regions. The second step involves allocating GOP bits according to conventional rate control methods, ensuring consistency with prior processes. The third stage entails acquiring encoded QP and RDO information for each frame and using the subsequent model to calculate the rate control parameters for both ROI and ROU regions.

The initial λ_{cu} values for each block within a frame can be derived from the r-λ rate control framework outlined above. However, to refine the rate allocation specifically for ROI and ROU, we need to adjust these base values and derive the final adjustment parameters λ_{roi} and λ_{rou} for the respective regions. This tailored approach ensures that the bitrate is optimally distributed between areas of interest and non-interest, enhancing overall perceptual quality while maintaining bitrate stability.

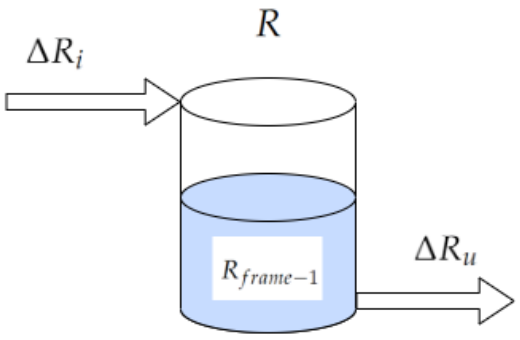


Figure 1. ROI Reservoir Model.

When the reservoir has not yet reached its overflow threshold R_{over}, adjustments to λ_{rou} alone are sufficient to ensure that the overall bitrate growth remains within an acceptable limit. When the reservoir's water level is low, the increment of λ_{rou} should be appropriately reduced to guarantee a steady bitrate consumption from the reservoir; conversely, when the water level rises high, the increment in λ_{rou} should increase proportionally to accelerate the continuous outflow of the coding rate from the reservoir. If the reservoir surpasses the overflow value R_{over}, the adjustment to λ_{roi} should decrease,

thereby reducing the inflow of code rate into the reservoir and helping it maintain a level below the overflow threshold.

The intensity of lambda adjustment for ROI (λ_{roi}) and ROU (λ_{rou}) is contingent upon the state of the ROI reservoir. A smaller λ value indicates a higher likelihood of dividing the CU into smaller blocks, which, in turn, leads to a smaller calculated quantization parameter and thus increased bitrate consumption in the corresponding region.

The target quality enhancement factor for ROI is denoted as K_{roi}, which is user-defined. For a given frame, the coding complexity of the ROI area is represented by C_{roi}, computed as the sum of absolute transform coefficients (SATD) following motion compensation during the lookahead process.

We conduct downsampling filtering on the current frame's image, resulting in a reduced-resolution image at half the original dimensions. Upon this lower-resolution image, we execute predictions with half-pixel accuracy and subsequently subtract the outcome from the original full-resolution image to yield the residual signal. Thereafter, we subject residual to the Hadamard transformation and compute the aggregate sum of absolute values, thereby obtaining the SATD. The sum of SATDs for all pixels within the ROI and the ROU region is C_{roi} and C_{rou}, respectively.

$$C_{roi} = \sum SATD_{roi} \tag{3}$$

$$C_{rou} = \sum SATD_{rou} \tag{4}$$

The size of the ROI is measured by the number of pixels in the region, P_{roi}. Correspondingly, the coding complexity of the ROU is C_{rou}, and its size is represented by P_{rou}. Additionally, the base quantization parameter for the entire frame is QP_{frame}.

By utilizing λ_{roi} and λ_{rou} with the pre-configured coding table, we can determine the quantization parameters QP_{roi} and QP_{rou} for the ROI and ROU regions, respectively. To estimate the increased bitrate allocation for the ROI area, the calculation method proceeds as follows:

$$\Delta R_i = R_i - R_i' = \alpha \times C_{roi}/QP_{roi} - \alpha \times C_{roi}/QP_{frame} \tag{5}$$

$$\Delta R_u = R_u - R_u' = \alpha \times C_{rou}/QP_{rou} - \alpha \times C_{rou}/QP_{frame} \tag{6}$$

where α is the rate Proportional coefficient. The terms ΔR_i and ΔR_u signify intermediate quantities that capture the incremental changes in bitrate attributed specifically to the ROI and ROU, respectively.

As depicted in Figure 2, we employ a shallow 4-layer neural network, which takes into account the real-time constraints in live streaming scenarios where each frame has specific encoding speed requirements. This network is configured with 8 input variables: $R_{frame-1}$, R_{over}, C_{roi}, C_{rou}, P_{roi}, P_{rou}, K_{roi}, and QP of the previous frame $QP_{frame-1}$. The architecture includes a first hidden layer with 52 intermediate feature nodes, followed by a second hidden layer consisting of 36 intermediate feature nodes, culminating in two output nodes that represent the predicted values for λ_{roi} and λ_{rou} for the upcoming frame.

The complete encoding and rate control flowchart is shown in Figure 3. In the preparatory analysis and encoding stages, systematically acquire data on the positional information of the ROI and ROU, their associated encoding complexities, the employed quantization parameters across encoded frames, along with the actual bit count per frame. Feed these empirical measurements into the reservoir model to dynamically update its bit rate and buffer occupancy status, serving as inputs to the CNN model. Ultimately, the CNN model predicts and outputs the rate control parameters λ_{roi} and λ_{rou} that correspond to and differentiate between ROI and ROU. By utilizing this neural network model, we calculate the adjusted λ parameters for the respective regions and proceed to apply these values in the coding process. This approach ensures that our rate control algorithm efficiently balances bitrate allocation between ROI and ROU while meeting the stringent latency and performance demands inherent in live video streaming environments.

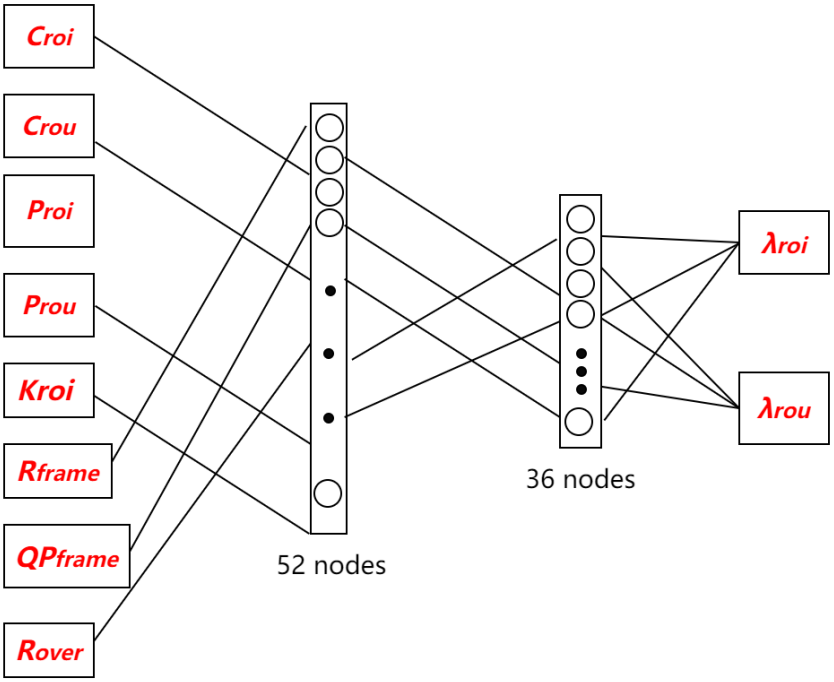


Figure 2. Fully Connected Inference Networks.

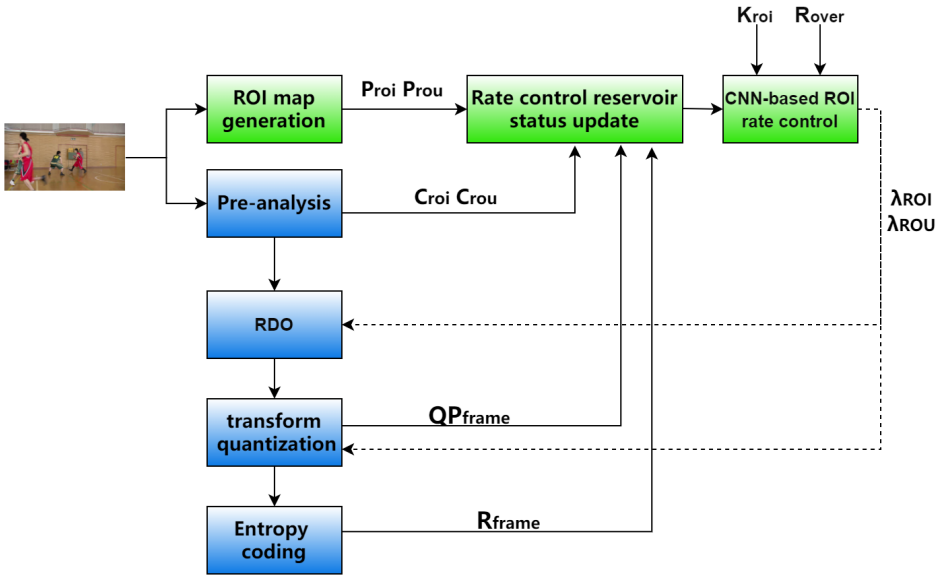


Figure 3. CNN-based ROI rate control.

After one frame encoding is completed, the reservoir R updating formula is

$$R_{frame} = R_{frame-1} + \Delta R_i + \Delta R_u \tag{7}$$

$$R_{frame} = R_{frame-1} + \alpha \times C_{roi}/QP_{roi} + \alpha \times C_{rou}/QP_{rou} - \alpha \times (C_{roi} + C_{rou})/QP_{frame} \tag{8}$$

3.2. Model Training Data Generation

To train the proposed neural network, a meticulous dataset collection process is executed in several stages, as depicted in Figure 4:

Step 1 Firstly, we establish a Quantization Parameter (QP) range of $[QP_{min}, QP_{max}]$, which encompasses the QP values suitable for all videos undergoing live streaming en-

coding. In the case of HEVC, when subjective quality is not a primary concern, this range can be directly set to [0,51].

Step 2 Next, utilizing mm-detection, we pre-detect and store ROI maps from the training video dataset. The detection includes identifying objects of interest such as faces, human bodies, and common scene elements.

Step 3 For each individual QP_i within the set range, we initially encode the video to acquire λ_{roi} data across all ROI regions. Following this, we fix the λ_{roi} values in the ROI areas and re-encode the video with varying QPs (QP_j) ranging from QP_{min} to QP_{max}. This allows us to collect λ_{rou} data for ROU regions. Concurrently, we gather all relevant input variables for the CNN, including $R_{frame-1}$, R_{over}, average λ_{cu} in ROI and ROU, C_{roi}, C_{rou}, P_{roi}, P_{rou}, and the previous frame's QP_{frame}. The output information from the CNN consists of λ_{roi} and λ_{rou}. Here, K_{roi} is defined as the difference between the initial encoding QP (QP_i) and the subsequent loop QP (QP_j).

Step 4 We repeat Step 3 for every possible QP value within the entire range for every video in our dataset to generate an exhaustive set of training data.

Data preparation and cleaning play a pivotal role. To ensure data integrity, we segment the training videos into scenes containing only one intra frame per video; multiple intra frames are strictly avoided as they could potentially contaminate the data. Furthermore, any collected training videos where the QP values for ROI or ROU exceed their respective limits are discarded.

Ultimately, after rigorous preprocessing, 42,000 valid videos with diverse content—such as video games, sports events, shows, news broadcasts, movies, and TV dramas—are utilized as training samples. Our model achieved a commendable test set accuracy of 92%. We refined the model structure by incorporating Huber Loss as a regression constraint during the input phase. Additionally, we applied pair-wise ranking loss to further constrain the λ_{roi} and λ_{rou} values with respect to the QP upper and lower bounds.

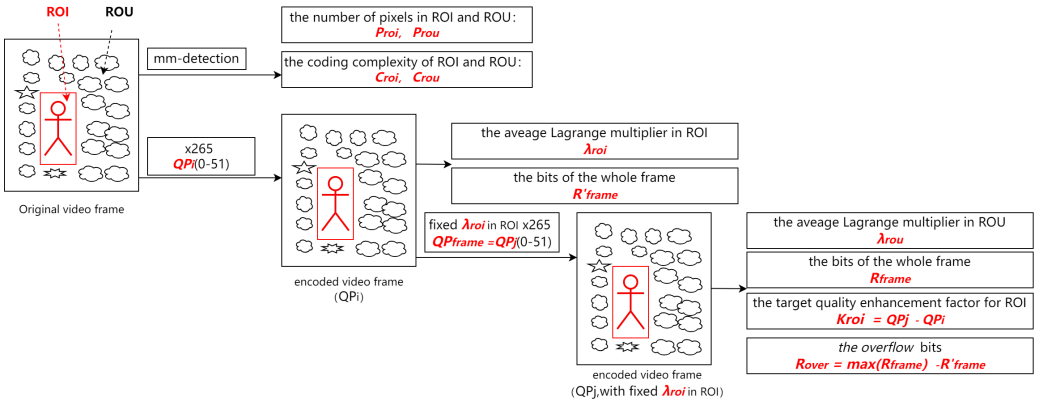


Figure 4. Model Training Data Generationg.

We also experimented with integrating features like video resolution and frame rate as inputs, but empirical results indicated that these had no significant impact on prediction accuracy.

3.3. Fast Object Detection

To meet the real-time constraints of object detection in live streaming and ensure each frame is encoded within the time limit, we leverage reference frame relationships to propagate and track detected objects.

In our approach, as depicted in Figure 5, a target area is initially detected using mm-detection in a preceding frame. If there exists a corresponding area with modifications in the subsequent frame, these Coding Units (CUs) are concatenated to produce a new region, which serves as the targeted detection outcome for that particular frame. Concurrently, to

accommodate potential new targets entering the scene in later time frames, a fresh round of detection is executed on the entire screen every n frames.

Given the low latency requirements of live streaming scenarios where videos are typically encoded as a Group of Pictures (GOP) per encoding cycle, empirical validation has shown that setting n equal to 8 achieves the optimal balance between detection quality and speed, effectively boosting the algorithm's efficiency by a factor of 8. This strategy ensures both timely detections and seamless video encoding without compromising visual fidelity or stream stability.

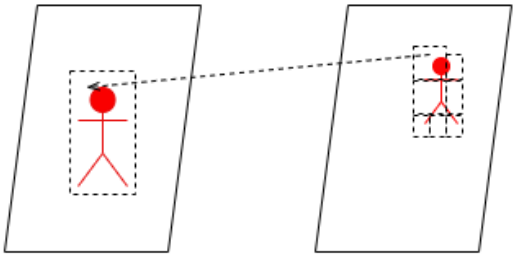


Figure 5. Reference Frame Relationship for Target Tracking.

4. Results

In the encoding of live videos, we have integrated our proposed rate control scheme into the open-source commercial HEVC encoder x265 [24], employing the low-delay-B configuration. For ROI detection, the coordinates are obtained using the mm-detection toolkit for face and common object detection [19]. We evaluate our method on test sequences recommended by the HEVC standard committee, encompassing CLASS-B, CLASS-C, CLASS-D, and CLASS-E sets.

Our ROI reservoir rate control algorithm is compared with a conventional algorithm that lacks ROI coding, where λ is directly modified in the ROI region.

For a detailed comparison, consider the 1080 p video stream from the BasketballDrive sequence in Class-B at a frame rate of 50 fps with a bitrate of approximately 1500 kbps. In Table 1, the first row serves as an anchor, showcasing the performance of a rate control algorithm based on R-lambda. The second row represents the commonly employed method of directly adjusting QPoffset for the ROI region, setting QPoffset to -6.

Our results demonstrate that while this approach significantly improves PSNR and SSIM metrics in the ROI area, it leads to a 60% increase in bit rate. This substantial addition of bitrate during live streaming can cause increased stutter rates during transmission and playback, negatively impacting the viewer's experience. The subsequent three rows of the table show the outcomes when progressively enhancing the parameter K_{roi} for the ROI region. As K_{roi} increases, objective quality metrics within the ROI region improve correspondingly.

Concurrently, the ROI reservoir algorithm reduces bits allocated to the ROU region to balance the overall bitrate, which consequently leads to a decrease in the objective metrics of the ROU region. Notably, throughout the process of tuning the ROI enhancement coefficient, the overall file bitrate increases by less than 5%, thereby satisfying the bitrate fluctuation requirements for video coding in live broadcast scenarios.

Table 1. BasketballDrive@1920x1080@50 fps, 1.5 Mbps, ROI metrics.

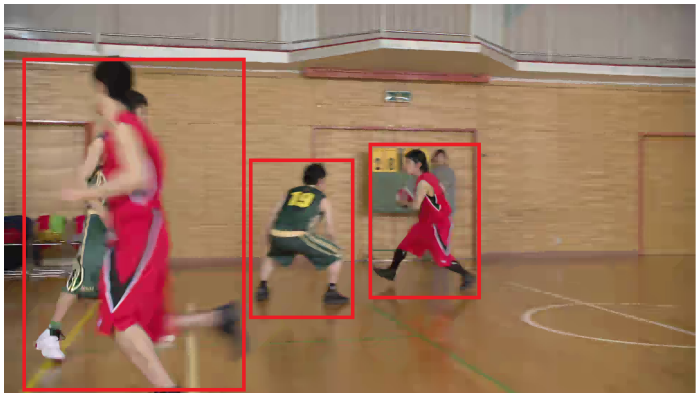
Algorithm	BitRate (kbps)	ROI-PSNR (dB)	ROI-SSIM	ROU-PSNR (dB)	ROU-SSIM
R-λ [25]	1526.26	35.01	0.881	34.88	0.881
ROI constant QPoffset [10]	2472.54 (+62.0%)	36.91 (+1.90)	0.907 (+0.026)	34.80 ($-$0.08)	0.881($-$0.000)
$K_{roi} = 2$	1544.58 (+1.2%)	35.71 (+0.70)	0.901 (+0.020)	34.77 ($-$0.11)	0.875 ($-$0.006)
$K_{roi} = 4$	1552.21 (+1.7%)	37.23 (+2.22)	0.908 (+0.027)	34.48 ($-$0.40)	0.874 ($-$0.007)
$K_{roi} = 6$	1556.79 (+2.0%)	38.47 (+3.46)	0.926 (+0.045)	34.37 ($-$0.51)	0.870 ($-$0.011)

Experimental outcomes demonstrate the superiority of our proposed method in both objective and subjective evaluations for ROI, as evidenced in Figure 6, and the red box in

(a) is the ROI area. Our scheme effectively enhances the discernibility of details within ROI regions, particularly human bodies detected by mm-detection, while maintaining an acceptable level of visual quality in the ROU areas without any significant degradation.

Comparatively, against the original video, while there is a general uniform increase in PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) metrics across all configurations, a subtle improvement in subjective quality is observed with $K_{roi} = 2$. However, a more substantial leap in image subjective quality becomes apparent when setting K_{roi} to 4. Although the PSNR and SSIM values for $K_{roi} = 6$ surpass those of $K_{roi} = 4$, the subjective quality difference between the ROI regions of these two settings is not appreciably distinguishable.Simulation results further validate that our custom encoder can adeptly adjust the subjective visual quality of ROI in relatively incremental steps. The visual fidelity of the ROI is consistently improved, while ensuring that the overall frame's visual quality remains satisfactory and acceptable.

As evidenced in Tables 2 and 3, by employing $K_{roi} = 4$ to analyze the variations in bit rate and performance metrics across different datasets and resolutions. Compared to the ROI constant QPoffset method, our CNN-based ROI rate control exhibits marginally superior PSNR and SSIM quality improvements within the ROI regions while incurring smaller PSNR and SSIM quality degradation within those same ROIs. Most critically, this CNN-based approach achieves quality gains that would typically require an average 41% bitrate consumption when using ROI constant QPoffset, but does so with only a 2.8% bitrate overhead. It is evident that our method can effectively manage bitrate growth while maintaining stable metric returns for ROI regions. The controlled loss within ROU areas is evenly distributed throughout each frame.



(**a**) 1.5 Mbps original R-λ live stream

(**b**) 1.5 Mbps live stream, $K_{roi} = 2$

Figure 6. *Cont.*

(c) 1.5 Mbps live stream, $K_{roi} = 4$

(d) 1.5 Mbps live stream, $K_{roi} = 6$

Figure 6. BasketballDrive_1920 x 1080@50 fps 1.5 Mbps live stream.

As evidenced in Table 4, in a study involving 100 participants rating video quality on a 1-to-5 Likert scale—with 1 representing the lowest perceived quality and 5 the highest—the use of ROI constant QPoffset strategy as a baseline scenario yields an average enhancement of 2.5% in the mean subjective score (MOS). The methodology presented herein attains a superior uplift of 9.1% in the mean subjective scores through the application of R-λ.

The results underscore the efficacy of this scheme in delivering enhanced ROI quality and consistency. By optimizing the QP (Quantization Parameter) in the ROI area to maintain a lower value in the proposed solution, we achieve superior perceived quality. Furthermore, the stability of QP ensures a consistent level of quality within ROI. Notably, this solution can be successfully implemented even under tight bandwidth constraints in live stream encoding scenarios.

Most importantly, the integration of a CNN-based ROI rate control model sets our proposed scheme apart, enabling it to perform superior rate control on the majority of test sequences. This approach thus represents an innovative and effective means of managing bitrate allocation for improved visual experience, especially in the regions of interest.

Table 2. ClassB–ClassF ΔBitRate and Δmetrics, ROI constant QPoffset vs. R-λ.

Video Sets	ΔBitRate (kbps)	ΔROI-PSNR (dB)	ΔROI-SSIM	ΔROU-PSNR (dB)	ΔROU-SSIM
Class B	47.5%	1.57	0.0471	−0.96	−0.004
Class C	42.9%	2.47	0.0108	−0.59	−0.009
Class D	35.1%	2.26	0.0192	−0.81	−0.012
Class E	38.0%	2.49	0.0508	−0.54	−0.006
Class F	43.9%	2.07	0.0377	−0.52	−0.011
Average	41.5%	2.17	0.0331	−0.68	−0.008

Table 3. ClassB–ClassF ΔBitRate and Δmetrics, K_{roi} = 4 vs. R-λ.

Video Sets	ΔBitRate (kbps)	ΔROI-PSNR (dB)	ΔROI-SSIM	ΔROU-PSNR (dB)	ΔROU-SSIM
Class B	4.2%	1.43	0.017	−0.32	−0.004
Class C	2.8%	1.52	0.028	−0.45	−0.007
Class D	1.5%	3.25	0.038	−0.71	−0.011
Class E	2.2%	2.92	0.043	−0.53	−0.009
Class F	3.5%	1.87	0.025	−0.78	−0.006
Average	2.8%	2.20	0.030	−0.56	−0.007

Table 4. ClassB–ClassF MOS.

Video Sets	R-λ	ROI Constant QPoffset	ROI Constant QPoffset vs. R–λ(%)	K_{roi} = 4	K_{roi} = 4 vs. R-λ(%)
Class B	3.60	3.79	5.4%	4.01	11.4%
Class C	2.85	2.90	1.5%	3.09	8.4%
Class D	3.61	3.71	2.7%	3.93	9.0%
Class E	4.22	4.23	0.4%	4.45	5.5%
Class F	3.18	3.26	2.5%	3.54	11.0%
Average	3.49	3.58	2.5%	3.80	9.1%

5. Discussions

The experimental findings reveal that when the ROI enhancement parameter K_{roi} surpasses a certain threshold, an excessive reduction in bitrate can lead to severe degradation of metrics within the ROU area. This may result in subjective quality issues that cannot be overlooked, such as pronounced blocking artifacts or texture loss. To address this issue, it is crucial to set a minimum λ_{min} value for the ROI region and a maximum λ_{max} value for the ROU region, thereby establishing upper and lower bounds on QP values in the corresponding areas and ensuring that subjective image quality remains uncompromised. These parameters, λ_{min} and λ_{max}, can be derived by consulting a predefined λ-QP lookup table using the corresponding QP_{min} and QP_{max}.

In contrast with conventional linear or nonlinear models manually designed, leveraging Convolutional Neural Networks (CNNs) for model parameter prediction allows for a more comprehensive consideration of diverse influencing factors. Traditional manual empirical modeling techniques are prone to overlook high-order features and their intricate interactions, which can inadvertently introduce inaccuracies into the system. On the other hand, employing CNN models in encoding's mode decision-making processes, especially those involving multiple inputs and outputs, offers a superior alternative. The intrinsic capability of CNNs to discern complex relationships among these variables makes them a more reliable choice, effectively reducing potential errors associated with less advanced methodologies.

In our CNN-based ROI rate control model, the accuracy in predicting the encoding complexity C_{roi} and C_{rou} for both ROI and ROU constitutes a paramount parameter. Currently, our prediction relies on estimating this complexity using Sum of Absolute Difference Transform (SADT) within the current frame. However, future refinements could involve enhancing the estimation of encoding complexity for ROI and ROU by incorporating inter-frame prediction using multiple reference frames.

6. Conclusions

In this paper, we propose an ROI-centric rate control algorithm for High Efficiency Video Coding (HEVC). The innovative algorithm operates akin to a reservoir management system and synergistically combines it with a Convolutional Neural Network (CNN)-based parameter prediction network. This integration aims to enhance the quality metric of Regions of Interest (ROI) while maintaining the overall stability of the video bitrate.This proposed scheme proves highly effective in scenarios where high-resolution, high-frame-rate, and high-bitrate live broadcasts are essential. It is versatile across various content types and delivers superior results. Not only does it ensure improved quality and uniformity within the ROI regions but also achieves a significantly lower bitrate compared to direct Quantization Parameter (QP) adjustments in CU as implemented by x265. Thus, it offers better Quality of Service (QoS) in demanding live broadcast environments.

In future work, we will enhance the prediction accuracy by employing CNN methods and conducting temporal analysis through the use of multiple reference frames, aiming to improve the precision of encoding complexity prediction. Moreover, the selection of coding modes in the current frame significantly impacts the accuracy of bitrate control; hence, in subsequent stages, we plan to model and quantify this influence, incorporating it into our CNN-based bitrate prediction model.

Author Contributions: Conceptualization, X.D.; Methodology, X.D.; Software, X.D.; Validation, X.D.; Writing—review & editing, X.D.; Supervision, X.C. and X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: Author Xianguo Zhang was employed by the company Tencent. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. *Intelligent Multimedia Technologies for Networking Applications: Techniques and Tools: Techniques and Tools*; IGI Global: Hershey, PA, USA, 2013.
2. Liu, Y.; Li, Z.G.; Soh, Y.C. Region-of-interest based resource allocation for conversational video communication of H. 264/AVC. *IEEE Trans. Circuits Syst. Video Technol.* **2008**, *18*, 134–139. [CrossRef]
3. Sun, Y.; Ahmad, I.; Li, D.; Zhang, Y.Q. Region-based rate control and bit allocation for wireless video transmission. *IEEE Trans. Multimed.* **2006**, *8*, 1–10. [CrossRef]
4. Chi, M.C.; Chen, M.J.; Yeh, C.H.; Jhu, J.A. Region-of-interest video coding based on rate and distortion variations for H. 263+. *Signal Process. Image Commun.* **2008**, *23*, 127–142. [CrossRef]
5. Song, H.; Kuo, C.C.J. A region-based H. 263+ codec and its rate control for low VBR video. *IEEE Trans. Multimed.* **2004**, *6*, 489–500. [CrossRef]
6. Meuel, H.; Kluger, F.; Ostermann, J. Region of interest (roi) coding for aerial surveillance video using avc & hevc. *arXiv* **2018**, arXiv:1801.06442.
7. Yang, L.; Zhang, L.; Ma, S.; Zhao, D. A ROI quality adjustable rate control scheme for low bitrate video coding. In Proceedings of the 2009 Picture Coding Symposium, Chicago, IL, USA, 6–8 May 2009; pp. 1–4.
8. Chiang, J.C.; Hsieh, C.S.; Chang, G.; Jou, F.D.; Lie, W.N. Region-of-interest based rate control scheme with flexible quality on demand. In Proceedings of the 2010 IEEE International Conference on Multimedia and Expo, Singapore, 19–23 July 2010; pp. 238–242.
9. Doulamis, N.; Doulamis, A.; Kalogeras, D.; Kollias, S. Low bit-rate coding of image sequences using adaptive regions of interest. *IEEE Trans. Circuits Syst. Video Technol.* **1998**, *8*, 928–934. [CrossRef]
10. Wu, C.Y.; Su, P.C. A region of interest rate-control scheme for encoding traffic surveillance videos. In Proceedings of the 2009 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Kyoto, Japan, 12–14 September 2009; pp. 194–197.
11. Meddeb, M.; Cagnazzo, M.; Pesquet-Popescu, B. Region-of-interest-based rate control scheme for high-efficiency video coding. *APSIPA Trans. Signal Inf. Process.* **2014**, *3*, e16. [CrossRef]
12. Wu, P.H.; Chen, H.H. Frame-layer constant-quality rate control of regions of interest for multiple encoders with single video source. *IEEE Trans. Circuits Syst. Video Technol.* **2007**, *17*, 857–867.
13. Paudel, B.; Vafi, S.; Bhattarai, P. An adaptive ROI based UEP scheme for HEVC compressed video bitstreams. In Proceedings of the 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 15–18 October 2019; pp. 373–374.
14. Zhao, W.; Fu, J.; Lu, Y.; Li, S.; Zhao, D. Region-of-interest based coding scheme for synthesized video. In Proceedings of the 2015 Visual Communications and Image Processing (VCIP), Singapore, 13–16 December 2015; pp. 1–4.
15. Chen, M.; Lin, W.; Zheng, X. An efficient coding method for coding region-of-interest locations in avs2. In Proceedings of the 2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), Chengdu, China, 14–18 July 2014; pp. 1–5.
16. Ahonen, J.I.; Le N.; Zhang, H.; Cricri, F.; Rahtu, E. Region of Interest Enabled Learned Image Coding for Machines. In Proceedings of the 2023 IEEE 25th International Workshop on Multimedia Signal Processing (MMSP), Poitiers, France, 27–29 September 2023; pp. 1–6. [CrossRef]

17. Kao, C.H.; Weng, Y.C.; Chen, Y.H.; Chiu, W.C.; Peng, W.H. Transformer-Based Variable-Rate Image Compression with Region-of-Interest Control. In Proceedings of the 2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, 8–11 October 2023; pp. 2960–2964. [CrossRef]
18. Lin, P. Video Bitrate Allocation Algorithm Based on Regions of Interest. In Proceedings of the 2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 21–23 April 2023; pp. 1458–1461. [CrossRef]
19. Chen, K.; Wang, J.; Pang, J.; Cao, Y.; Xiong, Y.; Li, X.; Sun, S.; Feng, W.; Liu, Z.; Xu, J. mm-detection: Open mmlab detection toolbox and benchmark. *arXiv* **2019**, arXiv:1906.07155.
20. Ma, S.; Si, J.; Wang, S. A study on the rate distortion modeling for high efficiency video coding. In Proceedings of the 2012 19th IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September–2 October 2012; pp. 181–184.
21. Yu, S.; Ahmad, I. New rate control algorithm for MPEG-4 video coding. In Proceedings of the Visual Communications and Image Processing 2002, SPIE, San Jose, CA, USA, 19 January 2002; Volume 4671, pp. 698–709.
22. Li, Z.G.; Gao, W.; Pan, F.; Ma, S.W.; Lim, K.P.; Feng, G.N.; Lin, X.; Rahardja, S.; Lu, H.Q.; Lu, Y. Adaptive rate control for H. 264. *J. Vis. Commun. Image Represent.* **2006**, *17*, 376–406. [CrossRef]
23. Naccari, M.; Pereira, F. Quadratic modeling rate control in the emerging HEVC standard. In Proceedings of the 2012 Picture Coding Symposium, Krakow, Poland, 7–9 May 2012; pp. 401–404.
24. MulticoreWare Inc. x265 HEVC Encoder/H.265 Video Codec. Available online: http://x265.org (accessed on 1 January 2020).
25. Li, B.; Li, H.; Li, L.; Zhang, J. λ Domain Rate Control Algorithm for High Efficiency Video Coding. *IEEE Trans. Image Process.* **2014**, *23*, 3841–3854. [CrossRef]