An Efficient Compression Coding Method for Multimedia Video Data Based on CNN

Xu Liu[1](✉) and Yanfeng Wu[2]

[1] School of Media and Communication, Changchun Humanities and Sciences College, Changchun 130117, China
liuxu517@163.com

[2] School of Physical Education, Changchun University of Finance and Economics, Changchun 130122, China

Abstract. Due to the phenomenon of object occlusion and inconsistent motion of different objects in video, high-efficiency compression and coding of video data will generate prediction residuals related to texture structure. To solve this problem, this study proposes an efficient compression coding method for multimedia video data based on CNN. First, the multimedia video coding unit is divided, and the coded frames are arranged in POC order for coding. Then, the coding structure adjustment parameters are calculated, and after coding, the determined reference frame and the bit consumption caused by using the reference frame can be obtained. Finally, an intra-prediction algorithm for video data is established based on CNN. CNN encoder uses a series of down-sampling convolution and ReLU nonlinear mapping to extract and fuse global information, and conducts analysis on areas with low human visual sensitivity on the HEVC transform domain. Frequency coefficient suppression. The experimental results show that the method has good coding performance for test sequences with different contents and different resolutions.

Keywords: CNN · Multimedia video · Video data · Compression coding · Video coding · Efficient compression

1 Introduction

Since entering the Internet era, multimedia technology and communication technology have developed rapidly, and high-definition 1080P and ultra-high-definition 4K, 8K resolution video images are gradually entering our work and life. Compared with text, picture and sound data, video signal has a larger amount of data, especially in the case of limited network bandwidth and storage resources, the research on video coding technology is very important.

HEVC is a new generation of video coding standards. HEVC is based on the traditional hybrid video coding framework, and adopts more technological innovations, including flexible block division, finer intra-frame prediction, newly added Merge mode,

W. Fu and L. Yun (Eds.): ADHIP 2022, LNICST 468, pp. 123–136, 2023.
https://doi.org/10.1007/978-3-031-28787-9_10

tile division, adaptive sample compensation, etc. The core problem of coding technology is to obtain the optimal rate-distortion performance under the condition of limited time, space and transmission bandwidth. Flexible block partitioning, which includes coding units (CU), prediction units (PU), and transform units (TU), improves coding performance the most. These technologies double the encoding performance of HEVC compared to H.264/AVC. Although the existing video coding standards can effectively compress video images, as the resolution of video images becomes larger and larger, the compressed video data is still very large, and video data still faces great challenges in transmission and storage Challenges [1].

At present, the various modules of HEVC are not optimal. For example, in the next generation of VVC, the performance indicators of intra-frame prediction are further improved, which indicates that there is still room for further improvement in the video codec standard of HEVC. At the same time, the existing video coding standards ignore the visual characteristics of the human eye during coding, and do not allocate bit resources according to the importance of the video content. For those areas that do not meet the visual characteristics of the human eye, a large amount of bit resources and computing resources are often consumed. Ensure the image quality of important areas in the video image and affect the subjective experience of the observer.

In recent years, with the vigorous development of the country's most artificial intelligence, people have applied artificial intelligence technology to many real-world scenarios, such as face recognition, machine translation, and speech recognition. The commercial implementation of these applications is an important manifestation of technological innovation in social production. Optimization of the video coding framework has been ongoing since the standard was established. Research workers and industrial designers at home and abroad have made a lot of effort and made a lot of effective work. Under the condition of limited network bandwidth, computing resources and storage resources, it is necessary to study the coding method that can make the video image quality more in line with the visual characteristics of the human eye [2]. In recent years, the development of traditional video codec technology has been hindered, and many performance improvements have to sacrifice a lot of resources and time. Therefore, deep learning technology is introduced into the existing video codec standard (HEVC), and the encoding performance can be further improved. The improvement is a challenging task [3].

Generally speaking, two important factors in video coding: computational complexity and rate-distortion performance, are opposed to each other, so these optimization efforts can be simply divided into two categories. One is to reduce the required computational complexity while ensuring that the coding performance does not drop too much; the other is to appropriately increase the computational complexity to improve the overall coding performance. Introducing deep learning technology into traditional video encoding and decoding standards can solve some problems that traditional algorithms cannot solve, and can improve the encoding and decoding performance of video encoding and decoding.

Based on the above analysis, this paper proposes a new efficient compression and coding method for multimedia video data based on CNN. The design idea is as follows:

(1) Divide the multimedia video coding unit and complete the coding according to the POC sequence.
(2) Calculate the adjustment parameters of the coding structure.
(3) An intra prediction algorithm of video data is established based on CNN, and a series of down sampling convolution and relu nonlinear mapping are used to extract and fuse video coding information.
(4) In the hevc transform domain, the frequency coefficients of the regions with low visual sensitivity are suppressed.

This method accelerates the decision-making process of intra prediction mode from a new perspective, so as to improve the coding quality of multimedia video images.

2 Method Design

2.1 Division of Multimedia Video Coding Units

In the compression and coding process of multimedia video data, after traversing all possible partition modes and optimizing rate distortion, the optimal CU/PU can be obtained, which means that each CU/PU will be encoded many times, which will greatly increase the computational complexity [4]. In this paper, we define the division depth of a 4×4 PU as 4, so the division problem of the entire CTU can be transformed into a combination of division problems at the four levels of depth 0–3. A piece of input multimedia video data has three channels of red, green and blue, and the pixel value of each channel ranges from 0 to 255. Considering that CNN has better performance on data in the range of 0 to 1, in the compressed autoencoder, the encoding end will first normalize each channel of the input video image in the normalization layer as follows: Unification operation:

$$B(x, y) = \frac{A(x, y)}{M} \tag{1}$$

In formula (1), (x, y) represents the pixel position; $A(x, y)$ is the original pixel value at position (x, y); $B(x, y)$ is the normalized pixel value; M represents the value of the pixel value of each channel. The three branches will output feature maps of the same size, and then we combine these three types of features in the depth dimension and pass them through two convolutional layers with small convolution kernels to effectively learn the relationship between these features. Correlation and diverse features can help CNN better understand the content information of the current CU.

The coded frames are coded in POC order, and there is no direct reference between frames. All coded frames are I-frames, which are independently coded using only intra-frame prediction techniques. In this case, the GOP length is 1, that is, each frame is managed as an independent GOP. Correspondingly, in the decoding part of the auto-encoder, the de-normalization layer will de-normalize the input image as follows to restore the original image.

Since there are too many possibilities for CTU division, it is not advisable to predict it directly. Therefore, the scheme we design is to make independent predictions at these four decision-making levels. On the encoder side, the connection layer is a convolutional

layer with a convolution kernel of 3 × 3 and 32 channels, and on the decoder side, the number of channels is set to 128. A 1-unit padding operation is also used in the connection layer to ensure that the size of the input image is not changed during the convolution process. The designed neural network extends the first convolutional layer to three different branches. The first branch is a traditional convolutional layer, using a regular square convolution kernel, while the remaining two branches use asymmetrical A convolution kernel designed to detect texture details in near-horizontal and near-vertical directions. This structure enables our network to extract features more efficiently, the reference chain is a single chain, and the frame POC in the chain is not repeated. B frame in the LDB structure uses a double-chain reference, and the frames in the two reference chains are exactly the same, which requires one more reference process than LDP [5]. Finally, the extracted features are passed through three fully connected layers to obtain the final prediction result.

2.2 Calculate Coding Structure Tuning Parameters

The inter prediction technology of video coding uses reference frames to eliminate the time-domain redundancy of video and reduce the information entropy. The bit consumption of inter prediction is related to the bit consumption of the video source reference frame. The stronger the video motion and the more complex the texture, the more vectors are required for inter prediction.

In traditional video coding frameworks, quantization parameters are used to achieve different tradeoffs between bit rate and video quality PSNR. Under different quantization parameters, the division of CU/PU is different. For videos with non-rigid and complex textures such as water surface ripples, the inter-frame temporal correlation is weak, and the inter-frame prediction residual is relatively significant. To ensure image quality and maintain a low level of distortion, a smaller quantization step size is required. Residual information is retained, which consumes significantly more bits [6].

Generally speaking, the smaller the quantization parameter is, the more bits are used, and the higher the requirement of video quality is. At this time, in order to obtain finer prediction results, the encoder will more likely use smaller blocks. In view of this phenomenon, it is necessary to adapt our acceleration framework to diverse quantization parameters. Each coded frame refers to the previous frame adjacent to its POC and the previous key frame at most 3 GOPs. Then the reference frame set can be expressed as the following form:

$$U = \begin{Bmatrix} \alpha(t-1) & 4\alpha\left\lfloor \frac{t}{4} - 1 \right\rfloor \\ 4\alpha\left\lfloor \frac{t}{4} - 2 \right\rfloor & 4\alpha\left\lfloor \frac{t}{4} - 3 \right\rfloor \end{Bmatrix} \tag{2}$$

In formula (2), U represents the reference frame set; α represents the coded frame; t represents the time. For each coding unit, the determined reference frame and the bit consumption caused by using the reference frame can be obtained after coding. The calculation formula of its coded bits or entropy is as follows:

$$\delta(\beta_t) = \min\{E(\beta_t, \chi)\} \tag{3}$$

In formula (3), δ represents coding bits; β_t represents the coding unit of the t frame of the video; E represents coding; χ represents the reference CU that the frame to be coded finally adopts in the set of reference images that can be used. The spatial neighbors of the current block and the spatial neighbors of the reference block can help to further improve the coding performance.

Except for the two reference blocks, this paper takes the spatial neighbor pixels of the current block and the spatial neighbor pixels of the reference block as the input of the convolutional network. For the current coding CU, the entropy generated by coding the CU consists of three parts, including the entropy of the recorded motion vector, the recorded residual, the entropy of the transform and quantization coefficient matrix, and the entropy related to the recording mode and control information. In order to fully exploit the temporal long-term correlation and short-term correlation of video sequences, the video coding standard adopts a long-term reference mechanism and a short-term reference mechanism. Therefore, there are two types of reference frames for the video coding process - long-term reference frames and short-term reference frames.

In the HEVC coding standard, motion vectors pointing to long-term reference frames and motion vectors pointing to short-term reference frames cannot be cross-predicted [7]. The longer the reference distance, the more times it is indirectly referenced, and the more obvious the attenuation of the indirect reference intensity. The dependency factor is used as the coding structure adjustment parameter to improve the dependency strength of the key frame. The formula for calculating the dependency factor is as follows:

$$\gamma(\alpha) = q^{|\alpha - u|} \tag{4}$$

In formula (4), γ represents the dependency factor; q represents the dependency strength of the reference frame; u represents the distance between the reference frames. The reference block image is expanded to the left and upward to obtain a new image block. In order to utilize the spatially adjacent pixels of the current block, the prediction block obtained by the uniform weighting of the two reference blocks is spliced with the spatially adjacent pixels to form a new image block. There are strict reference rules between each level. Frames at higher levels can only refer to frames at lower levels, and frames at the highest level are not referenced. Frames in the lower layers are referenced the most, which enables the encoder to output a stream with frame rate diversity. When the frame rate needs to be reduced, some low-level frame encoding can be selected, and high-level frames are not considered, thereby greatly reducing the bit rate consumption.

2.3 Building an Intra-frame Prediction Algorithm for Video Data Based on CNN

Intra-frame prediction technology plays an irreplaceable role in the current video coding system. It can effectively remove the spatial redundancy of video signals, prevent the spread of coding errors, and improve the random access performance of video streams. Due to the phenomenon of object occlusion and inconsistent motion of different objects in video, high-efficiency compression coding of multimedia video data will generate prediction residuals related to texture structure.

In this paper, an intra-frame prediction algorithm for video data is established based on CNN. The algorithm adopts cross-level direct connection, which can combine deep

and global semantic information with shallow and local representation information. Since the image denoising and super-resolution tasks confirm that residual learning can effectively improve the output image quality, in the inter prediction pixel correction technique, we also adopt the residual learning structure to improve the prediction accuracy. CNN-based intra prediction mode for video data which consists of a convolutional autoencoder and an auxiliary trained discriminator. The network structure draws on image inpainting technology for intra-frame prediction. It uses 3 reference blocks to predict the current block to be predicted. Uniformly weighted prediction blocks are used in the skip structure to implement residual learning. Due to the addition of spatially adjacent pixels in the network input, the size of the input image block is larger than the output prediction block, so the network needs to crop the image block and output the target prediction block.

CNN takes the image of the current block and its 3 neighboring blocks as the input image. Among them, CNN uses the intra-frame prediction result of HEVC as the initialization value. The output of the network is the optimized intra prediction result, which is used for prediction coding by HEVC. The convolutional network is trained by extracting the relevant information of all bidirectional prediction blocks according to the code stream information [8]. Uncompressed video sequences are limited, and we increase the number of training samples by downsampling and cropping operations. For intra prediction tasks, we require the convolutional auto-encoder to be able to accurately generate the texture information of the current block to minimize the prediction residual. The loss function of the convolutional autoencoder is as follows:

$$\eta = \frac{\sum_{w_1} \sum_{w_2} (z^* - z)^2}{s w_1 w_2} \tag{5}$$

In formula (5), η represents the loss function of the convolutional autoencoder; w_1 and w_2 represent the width and height of the feature map, respectively; z^* and z represent the MSE of the real map and the predicted map, respectively; s represents the number of prediction branches of the decoder. On sequences of different resolutions, the same block size contains different scales of information. Therefore, it is more universal to obtain training samples generated by sequences of multiple resolutions by downsampling for network training [9]. Following the CNN training mode, a competitive alternating training method is adopted. Then, CNN objective function includes the loss function of the above-mentioned convolutional autoencoder and the adversarial loss function of the generator. The generator's adversarial loss function is calculated as follows:

$$\varphi = \nu\eta + (1 - \nu)\big[-\log K(z)\big] \tag{6}$$

In formula (6), φ represents the adversarial loss function; ν represents the adjustment parameter, which is set to 0.999 in this paper; K represents the discriminator. The loss function of the convolutional autoencoder is not fed back to the encoder, because the decoder is responsible for image synthesis, and the encoder is mainly responsible for extracting and compressing image features. According to the size of the largest CU block, take the center point of the target area as the fixed point, adjust the coordinates of the edge point of the target area, that is, expand the detected target area to the nearest

64 times the pixel boundary. The obtained region is used as the region of interest for subsequent video coding, and the other regions are used as non-interested regions.

The CNN encoder employs a series of down-sampled convolutions and ReLU nonlinear maps to extract and fuse global information. We double the number of feature channels when performing each downsampling convolution, preventing a rapid reduction in the number of features. At the top of the CNN encoder, we adopt the Dropout technique with a dropout rate of 0.5 to eliminate and weaken the joint fitness between neurons and enhance the generalization ability of the network.

2.4 Building a High-Efficiency Compression Coding Model for Video Data

After a frame in the video is input to the encoder, the motion vector, division mode, and prediction information such as the residual of each encoded pixel block are obtained through the intra-frame and inter-frame prediction modules, and the residual is subjected to frequency domain transformation and quantization through the transformation/quantization module, and then through entropy coding, the information is synthesized as a binary bit stream.

At the encoding end, after prediction, transformation and quantization, filtering and other modules, a reconstructed image with certain distortion is formed for subsequent images to rely on. In the process of video encoding, it is necessary to control the encoding rate according to the actual situation. In order to make the actual bit rate after video encoding within the set target bit rate range, various encoding parameters of the encoder need to be adjusted during encoding set up. During video encoding, the encoding parameters can be adjusted to ensure that the actual bit rate of the encoding is close to the set target bit rate, so as to achieve the purpose of bit rate control. The principle of rate control is shown in Fig. 1.

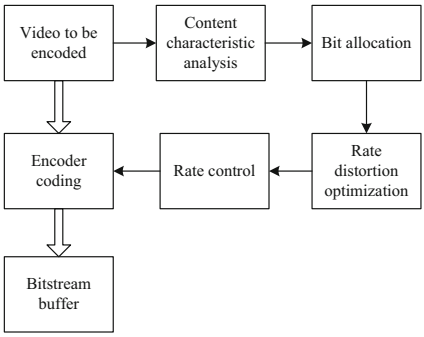

Fig. 1. The principle of rate control

When the code rate changes, the current CTU needs to be observed, and then the quantization parameters used are determined. The CTU will be sent to the encoder in the coding order, that is, from left to right and top to bottom in the current frame. After the prediction residual transformation, most of the energy is concentrated in the

low-frequency coefficients in the upper left corner of the matrix, and more detailed information in the image will be scattered in the high-frequency region [10].

Since the luminance component consumes the highest bit rate, we extract the luminance component of the CTU to represent its spatial information. Then, we merge the luminance component and importance map depthwise. Considering that the human eye is not very sensitive to the distortion of high-frequency signals, this section proposes a frequency coefficient suppression method. The areas with high visual sensitivity of the human eye are suppressed by the frequency coefficients of lower intensity.

The bit rate and distortion we use are relative values encoded with the standard quantization parameters, rather than absolute values. This design can better reflect the actual situation through the value of the reward. The frequency coefficient matrix can be expressed as:

$$P = \vartheta \cos\left[\frac{\vartheta \pi (2c+1)}{8}\right] R \tag{7}$$

In formula (7), P represents the transformed frequency coefficient matrix; ϑ represents the compensation coefficient; c represents the encoding transformation size; R represents the residual signal matrix to be transformed. HEVC uses floating-point coefficients to multiply larger values when transforming integer prediction residuals to make the transform results more accurate. Specifically, the floating-point coefficient can be positive or negative. If it takes a negative value, it means that the advantage of reducing the bit rate used is not enough to cover the negative impact of the resulting distortion; if it is a positive value, it may be used the current quantization parameter does not produce additional semantic distortion, or the semantic distortion value is small and can be ignored.

For each CTU to be encoded, according to its texture-aware weight value, calculate the QP parameter down-regulation value DQP. The larger the DQP and the larger the quantization step size, the more low-frequency signals are eliminated in the coefficient matrix, the less accurate the residual matrix of inverse quantization and inverse transformation, and the greater the distortion of the image. For the non-interesting area, each CTU selects a frequency coefficient suppression matrix according to its texture perception weight value to suppress its DCT frequency coefficients to different degrees. With quantization suppression, the residual matrix to be transmitted contains very few significant values and a large number of zeros. The candidate frequency coefficient suppression matrix group is calculated as follows:

$$f = \begin{cases} 1, \left(\omega + \sigma \le \frac{hc}{4} + b\right) \\ 0, \ otherwise \end{cases} \tag{8}$$

In formula (8), f represents the candidate frequency coefficient suppression matrix; ω and σ are the abscissa and vertical coordinates of the matrix elements respectively; h represents the index of the suppression matrix, which is 1, 2, and 4, and the suppression intensity increases in turn; b is the partial shift. High-intensity pressing for randomly textured areas, medium-intensity pressing for flat areas, and lower-intensity pressing for structured textured areas. High-frequency signals, DC component values are also mapped

into smaller intervals, so that the number of bits required to express the coefficient matrix is drastically reduced.

Combining the above processes, the design of an efficient compression and coding method for multimedia video data based on CNN is completed. The design steps are shown in Fig. 2.

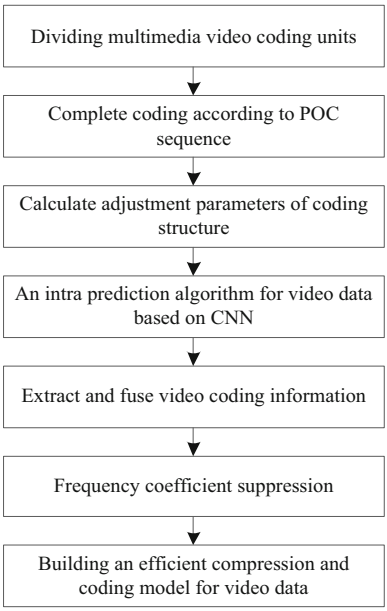

Fig. 2. Steps of method design

3 Experimental Study

3.1 Experimental Environment and Configuration

The following experiments are designed to verify the performance of the CNN-based efficient compression coding method for multimedia video data.

In this experiment, the HEVC encoding software is X265_1.8, the development environment is VisualStudio2012, and the processor of the test platform is Intel Core i5-2520 with a main frequency of 2.5 GHz.

The configuration of the X265 encoder is as follows: frame rate 30fps, IPP mode, I-frame interval is 100, and DCT coefficients are suppressed by odd-numbered frames. The specific configuration of LD is the LDP encoding method, and the quantization parameters are set to 22, 27, 32, 37 and 42 respectively, and the above parameters are used for compression encoding.

The experiments build a dataset containing 90 uncompressed video sequences from the derf and SJTU datasets. For the test set, 18 class A-E sequences with different resolutions proposed by the Joint Video Coding Collaborative Group are used in the

experiments. The first 100 frames of each sequence were selected for statistical analysis. The test sequence is shown in Table 1.

Table 1. Test sequence

Serial number	Sequence	Video
1	Class A	2560 × 1600
2	Class B	1080p
3	Class C	WVGA
4	Class D	WQVGA
5	Class E	720p

When training this encoding model, in each original video and its compressed video, we randomly select the original frame and its corresponding decoded target frame and adjacent frames to form training frame pairs. In order to verify the performance advantages of the CNN-based efficient compression coding method for multimedia video data, the analysis results are compared with the RNN-based and LSTM-based efficient compression coding methods for multimedia video data.

3.2 Results and Analysis

In order to measure the coding performance of this method, BD-rate index is used for measurement and analysis. A negative value of the BD-rate index indicates the saving degree of the code rate, that is, the coding compression rate. The experimental results of the A-E test sequences are shown in Tables 2, 3, 4, 5 and 6.

Table 2. BD-rate of Class A (%)

Testing frequency	Compression coding method based on CNN	Compression coding method based on RNN	Compression coding method based on LSTM
1	4.28	0.97	1.38
2	4.26	0.95	1.32
3	4.29	0.98	1.36
4	4.31	0.99	1.35
5	4.30	0.97	1.37
6	4.34	0.96	1.33
7	4.29	0.98	1.31
8	4.32	0.96	1.35
9	4.35	0.97	1.36
10	4.28	0.92	1.41

In the Class A sequence test, the maximum value of BD-rate index can reach 4.35% after applying the method in this paper, while the maximum value of BD-rate index is 0.99% and 1.41% after applying the video coding methods based on RNN and LSTM respectively. It can be seen that the compression effect of this method in Class A sequence is better.

Table 3. BD-rate of Class B (%)

Testing frequency	Compression coding method based on CNN	Compression coding method based on RNN	Compression coding method based on LSTM
1	3.05	1.42	0.92
2	3.08	1.15	0.94
3	3.09	1.48	0.96
4	3.07	1.46	0.94
5	3.05	1.43	0.95
6	3.11	1.36	0.98
7	3.09	1.38	0.97
8	3.10	1.42	0.99
9	3.15	1.37	1.02
10	3.03	1.29	0.98

In the Class B sequence test, the maximum value of BD-rate index can reach 3.15% after applying the method in this paper, while the maximum value of BD-rate index is 1.48% and 1.02% after applying the video coding methods based on RNN and LSTM respectively. It can be seen that the compression effect of this method in Class B sequence is better.

In the Class C sequence test, the maximum value of BD-rate index can reach 2.92% after applying the method in this paper, while the maximum value of BD-rate index is 1.41% and 1.32% after applying the video coding methods based on RNN and LSTM respectively. It can be seen that the compression effect of this method in Class C sequence is better.

In the Class D sequence test, the maximum value of BD-rate index can reach 1.52% after applying the method in this paper, while the maximum value of BD-rate index is 1.23% and 1.17% after applying the video coding methods based on RNN and LSTM respectively. It can be seen that the compression effect of this method in Class D sequence is better.

In the Class E sequence test, the maximum value of BD-rate index can reach 4.47% after applying the method in this paper, while the maximum value of BD-rate index is 1.26% and 1.20% after applying the video coding methods based on RNN and LSTM respectively. It can be seen that the compression effect of this method in Class E sequence is better.

Table 4. BD-rate of Class C (%)

Testing frequency	Compression coding method based on CNN	Compression coding method based on RNN	Compression coding method based on LSTM
1	2.84	1.28	0.96
2	2.86	1.26	0.98
3	2.89	1.23	1.02
4	2.88	1.25	1.04
5	2.85	1.27	0.99
6	2.91	1.31	1.03
7	2.92	1.32	0.99
8	2.87	1.29	1.02
9	2.67	1.41	1.32
10	2.54	1.35	1.32

Table 5. BD-rate of Class D (%)

Testing frequency	Compression coding method based on CNN	Compression coding method based on RNN	Compression coding method based on LSTM
1	2.47	1.19	0.95
2	2.48	1.18	1.04
3	2.49	1.21	0.96
4	2.46	1.22	0.98
5	2.48	1.19	1.03
6	2.51	1.23	1.05
7	2.52	1.22	1.04
8	2.49	1.18	0.99
9	2.46	1.17	1.03
10	2.45	1.21	1.17

The compression coding method designed in this paper significantly improves the coding performance on all test sequences. Compared with the RNN-based and LSTM-based coding methods, the CNN-based efficient compression coding method for multimedia video data has achieved obvious BD-rate gains. These experimental results can confirm that the efficient compression coding method of multimedia video data based on CNN has good performance for test sequences with different contents and different resolutions.

Table 6. BD-rate of Class E (%)

Testing frequency	Compression coding method based on CNN	Compression coding method based on RNN	Compression coding method based on LSTM
1	4.41	1.15	1.13
2	4.43	1.18	1.18
3	4.45	1.22	1.05
4	4.43	1.25	1.07
5	4.46	1.24	1.09
6	4.42	1.19	1.08
7	4.47	1.18	1.12
8	4.45	1.26	1.14
9	4.47	1.15	1.20
10	4.40	1.23	1.18

4 Conclusion

With the continuous development of information multimedia technology, high-definition video and ultra-clear video are gradually popularized in people's lives. In order to relieve the pressure of storage and network transmission, the importance of video coding technology is getting higher and higher.

This paper proposes a high-efficiency compression coding method for multimedia video data based on CNN. Intra-frame prediction is performed by means of a convolutional auto-encoder, which can effectively reduce the prediction residual and improve the coding rate-distortion performance. This paper only designs the intra-frame coding mode, if it can be extended to the inter-frame mode, it will be more valuable. Compared with the intra-frame mode, the inter-frame mode needs to additionally consider the influence of the reference frame, and the lower-quality reference frame will propagate the distortion, so a more reasonable bit allocation scheme needs to be studied.

Fund Project. Higher Education Teaching Reform Research Project of Jilin Province: Construction and Practice of "Online and Offline" Hybrid Teaching Mode for Film and Television Arts Majors in Universities under MOOC Environment (20213F2ENY4001J).

Education Science "Fourteen Five-Year" Project of Jilin Province: Research on the Construction of College of Modern Industry for Film and Television Media Major (ZD21088).

China Association of Private Education 2022 Annual Planning Project (School Development): Research on the Construction of College of Modern Industry for Film and Television Media Major (CANFZG22274).

References

1. Guo, J., Cao, L., Zhu, F.: Progressive compression method of real-time data under load balancing strategy. Comput. Simul. **38**(3), 365–368,429 (2021)

2. Wu, Y., Peng, Y., Lu, A.: Research on image compression coding technology based on deep learning CNN. Comput. Eng. Softw. **41**(12), 18–23 (v)
3. Wang, H., Ma, J., Qu, J.: Design and implementation of full HD video real-time compression coding and storage system. Chin. J. Electron Devices **44**(03), 513–518 (2021)
4. Wang, G., Jin, Y., Peng, H., et al.: Error correction of Lempel-Ziv-Welch compressed data. J. Electron. Inf. Technol. **42**(6), 1436–1443 (2020)
5. Pan, P., Yao, Y., Wang, H.: Detection of double compression for HEVC videos with the same coding parameters. J. Image Graph. **25**(5), 879–889 (2020)
6. Gu, H.: Data compression coding technologies for computer-generated holographic three-dimensional display. Infrared Laser Eng. **47**(06), 42–47 (2018)
7. Wang, T., He, X., Sun, W., et al.: Improved HEVC intra coding compression algorithm combined with convolutional neural network. J. Terahertz Sci. Electron. Inf. Technol. **18**(2), 291–297 (2020)
8. Jiang, W., Fu, Z., Peng, J., et al.: 4 Bit-based gradient compression method for distributed deep learning system. Comput. Sci. **47**(7), 220–226 (2020)
9. Yi, Y., Feng, G.: A video zero-watermarking algorithm against recompression coding for 3D-HEVC. J. Sig. Process. **36**(05), 778–786 (2020)
10. Li, F., Zhan, B., Xin, L., et al.: Target recognition technology based on a new joint sensing matrix for compressed learning. Acta Electron. Sin. **49**(11), 2108–2116 (2021)