2022 2nd Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS) | 978-1-6654-0034-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/ACCTCS53867.2022.00023

Fast Intra Mode Decision Algorithm of HEVC Based on Convolutional Neural Network

Yi Yingmin
Faculty of Automation and Information Engineering
Xi'an University of Technology
Communication University of China
Xi'an, China
yiym@xaut.edu.cn

Zheng Zhaoyang
Faculty of Automation and Information Engineering
Xi'an University of Technology
Communication University of China
Xi'an, China
z5mail@163.com

Yuan Yiwei
Faculty of Automation and Information Engineering
Xi'an University of Technology
Communication University of China
Xi'an, China
yyw@xaut.edu.cn

Xue Xianghong
Faculty of Automation and Information Engineering
Xi'an University of Technology
Communication University of China
Xi'an, China
xhxue@xaut.edu.cn

Li Yuxing
Faculty of Automation and Information Engineering
Xi'an University of Technology
Communication University of China
Xi'an, China
liyuxing@xaut.edu.cn

***Abstract*—Aiming at the high complexity and large amount of computation of HEVC intra mode decision algorithm, a fast intra mode decision algorithm based on convolutional neural network is proposed. Firstly, different size of prediction units are scaled to the same size through bilinear interpolation, and then sent to convolution neural network to perform convolution pooling and other operations. Finally, the mode decision results are selected by softmax layer. Five standard test sequences with different resolutions are used to form data sets, and the convolutional neural network model is trained and tested. Experimental results show that the proposed algorithm is equivalent to the standard test software HM16.0 in image quality and video bit rate, but the total coding time is improved by about 20.14%.**

Keywords—HEVC, convolutional neural network, intra coding, mode decision

I. INTRODUCTION

HEVC (High Efficiency Video Coding) is a new generation of video codec standard proposed by the ITU-T in 2013, which has about 50% performance improvement compared with the previous generation standard[1]. In recent years, deep learning algorithm has shown good application results in image processing and speech, including the field of video compression coding. Generally, in the intra coding process of HEVC, the rate distortion optimization algorithm was applied to select the intra mode, that is, the rate distortion calculation was carried out for each mode.So the whole optimization coding process was quite complex, which has affected the overall efficiency of video coding [2].

The intra prediction algorithm of HEVC has high complexity, mainly including the following aspects: first, the number of prediction units is multiple; second, there are 35 prediction models; and third, the calculation of rate distortion optimization cost function is complex. For the problem of many types of the second mode, relevant scholars have put forward many solutions. Some scholars have proposed a rough mode decision algorithm[3], that is, rough grouping calculation was carried out among 35 modes, and then accurate calculation and selection was implemented within the group, so as to select result from the final coding mode. Other scholars have proposed the most likely prediction mode decision algorithm[4], whose idea was to apply it to the current prediction unit directly. With this method, the mode results can be selected by adjacent coding units. Some scholars have proposed to introduce Sobel operator[5] and gradient information to determine the optimal mode calculation direction, the way which has reduced the range of mode decision and has avoided unnecessary calculation. However, the improvement of the above algorithms still have some shortcomings, such as slow speed or poor encoding effect, which can not meet the high requirements with encoding speed and encoding quality[6]. With the improvement of computer's computing power(computing power of computer), convolutional neural network(CNN) and other algorithms have achieved rapid development in image, audio and other fields by means of hardware acceleration. The application of deep learning method has achieved many effects which could not achieve by traditional methods[7].

Based on previous studies, this paper introduces the convolutional neural network algorithm in deep learning and carries it out in the HEVC video coding framework. This method replaces the rate distortion optimization process by the deep learning model, trains the convolutional neural network model with a large number of video sequence data, and outputs the mode decision results from the convolutional neural network, so as to reduce the coding time and improve the coding efficiency. This paper will focus on the main application of deep learning in HEVC intra coding, and combines depth model network with intra video coding. Section 1 briefly describes the algorithm flow of intra mode decision module algorithm and convolutional neural network in HEVC. Section 2 introduces the detailed steps of HEVC intra mode decision algorithm based on convolutional neural network. In Section 3, the effectiveness of the proposed model is verified by experiments. Section 4 summarizes the full paper.

978-1-6654-0034-3/22/$31.00 ©2022 IEEE
DOI 10.1109/ACCTCS53867.2022.00023

II. HEVC INTRA PREDICTION AND CONVOLUTIONAL NEURAL NETWORK

Generally, the collected original video has three aspects of data redundancy: human visual redundancy, inter frame temporal redundancy and intra frame spatial redundancy[8]. The elimination of human visual redundancy is generally realized by removing the insensitive high-frequency signals of human eyes by the DCT transformation; The time redundancy among frames can be eliminated by motion estimation and motion compensation; Intra prediction is the process of encoding the current pixel block according to the encoded pixel block in the current image, which can eliminate the spatial redundancy in the frame[9]. The elimination of redundant data in these three aspects is a research hotspot.

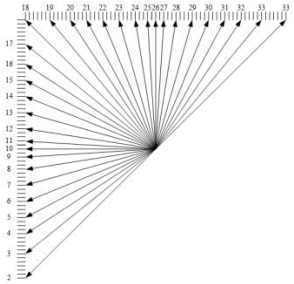

Fig.1 Angle modes for intra prediction.

Fig. 1 is a schematic diagram of 33 angle prediction modes. In the HEVC standards, there are 35 modes for the prediction of intra brightness component. Each mode represents the calculation method of predicting the pixels to be encoded from the encoded pixels, including mode 0 plane mode, mode 1 DC mode, and 33 angle prediction modes.

Convolutional neural network is the basic component of various deep learning models currently. Its structure includes several different types of layers, and improves learning efficiency by sparse connection and weight sharing[10].

III. MODE DECISION ALGORITHM OF INTRA ANGLE PREDICTION BASED ON CNN

Convolutional neural network shows better performance in image processing. Compared with the way of calculating the rate distortion cost one by one, it can get the best prediction mode directly.

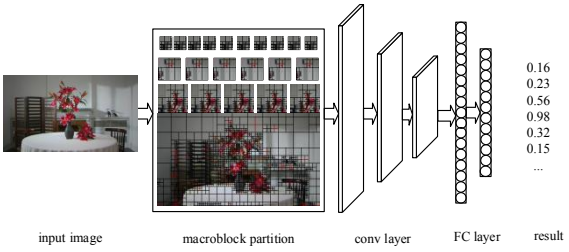

Fig. 2 Generating the best mode based on CNN

Fig. 2 is a schematic diagram of generating the best mode based on convolutional neural network, This part of the network structure is composed of a data processing layer, three convolution layers, two full connection layers and a softmax output layer. The data of network model training is the prediction unit in mode decision of intra angle prediction. The prediction unit has five different sizes, including 64 × 64, 32 × 32, 16 × 16, 8 × 8 and 4 × 4. Because the input image size of convolutional neural network is fixed, it is necessary to insert bilinear interpolation on prediction units of different sizes and adjust their size to 32 × 32. Then the size of the input layer image data becomes 32 × 32, and the convolution operation is performed. There are 32 convolution cores in the first convolution layer, and the size of convolution core is 5 × 5. The size of convolution step is 1 and the frame is filled with zero. In the initial layer of convolution neural network, a large convolution kernel is often used to obtain a large local receptive field, and then extract as many image features as possible. After one layer of convolution, the dimension of the output characteristic graph is 32 × 32 × 32.The function of activation layer is ReLU, which overcomes the gradient disappearance problem of sigmoid in the training process. The size of sliding window of the pool layer is 2 × 2. If the step size is 2, the dimension of the characteristic graph becomes 16 × 16 × 32 after passing through the pool. After that, the second layer of convolution operation is performed, with 32 convolution kernels in 3 x 3 size. The size of convolution step is 1, the frame is filled with zero, and size of the sliding window of the pool layer is 3 × 3. After the second layer of convolution operation, the size of the output feature map is 8 × 8 × 32. There are 64 convolution kernels in the third layer convolution, with a size of 3 × 3. The size of convolution step is 1, the frame is filled with zero, and size of the sliding window of the pool layer is 2 × 2. If the size of step is 2, the data dimension of the characteristic graph is 4 × 4 × 64 after the calculation of convolution layer. Then, the results of the convolution layer are output to two full connection layers. The number of neuron nodes in the first full connection layer is 1024 and the number of neuron nodes in the second full connection layer is 512. ReLU is used as the activation function and a random deactivation operator is added to reduce the over fitting further. Finally, the number of output neurons of the softmax layer is 35. For modes of 35 intra prediction angle, class 0 is planar mode, class 1 is DC mode, and the other classes correspond to modes of 33 angle prediction. The specific parameters of the network model are shown in Table I:

TABLE I CNN MODEL PARAMETERS

structure	Parameter
input	32×32
conv 1	kernel: 5×5, stride:1 pad:2, ReLU
pooling 1	kernel: 2×2, stride:2 Max
conv 2	kernel: 3×3, stride:1 pad:1, ReLU
pooling 2	kernel: 2×2, stride:2 Max
conv 3	kernel: 3×3, stride:1 pad:1, ReLU
pooling 3	kernel: 2×2, stride:2 Max
FC 1	1024, ReLU, dropout:0.5
FC 2	512, ReLU, dropout:0.5
softmax	35

The cross entropy function is used as the loss function, which can prevent convergence to the local optimum effectively. The expression of the loss function is as follows:

$$Loss = -\sum_{i=1}^{n} y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \qquad (1)$$

In equation 1, y_i represents the expected probability of mode i, $\hat{y}_i$ represents the actual output probability of mode i, and *Loss* represents the value of loss function.

The flow chart of the combination between mode decision algorithm of intra angle prediction and HEVC standards test software HM16.0 based on convolutional neural network is shown in Fig. 3.The convolutional neural network replaces the original rough mode decision or the most probable mode decision algorithms, which is gathered as candidate set by the rate distortion optimization mode . The algorithm flow of this part is: for the input video stream, intercept it into a single frame image firstly, and then input it into convolutional neural network after filtering. In the data processing layer, the picture will be divided into macroblocks, and the different size of macroblocks will be scaled into fixed size of 32 × 32 set of prediction units . Then input the convolutional neural network to obtain the results of angle prediction mode decision of all prediction units, and then continue the next coding operation.

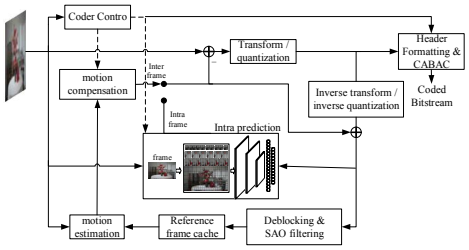

Fig.3 The flow chart of CNN intra mode decision and HM16.0 combination algorithm

IV. EXPERIMENT

Select 5 standard video sequences from 416 × 240 to 2560 × 1600 resolution to test, and all frames are selected for each video sequence. Under different quantization parameters (22, 27, 32 and 37), the mode decision decision algorithm for predicting angles within frames based on deep learning is compared with the standard reference software HM16.0 on the aspects of code rate, peak signal-to-noise ratio and encoding time. Three parameters are defined as follows:

$$\Delta B(\%) = \frac{Bitrate_{\text{improved}} - Bitrate_{\text{HM}}}{Bitrate_{\text{HM}}} \times 100\% \quad (2)$$

$$\Delta T(\%) = \frac{Time_{\text{improved}} - Time_{\text{HM}}}{Time_{\text{HM}}} \times 100\% \quad (3)$$

$$\Delta P(\text{dB}) = PSNR_{\text{improved}} - PSNR_{\text{HM}} \quad (4)$$

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \quad (5)$$

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \quad (6)$$

In equation 2, bitrate represents the video bitrate, that is, the number of bits transmitted in unit time, and its unit is kbps, ΔB represents the change of video bit rate before and after the introduction of convolutional neural network; In equation 3, time represents the total coding time, ΔT represents the change of the total coding time before and after the introduction of convolutional neural network; In equation 4, *PSNR* represents the peak signal-to-noise ratio, which is often used as an index to evaluate the quality of image reconstruction in image compression. The calculation method is shown in equation 5. MAX_I represents the bit width of each pixel, generally 8 bits. *MSE* is the mean square error of the image before and after encoding. The calculation method is shown in equation 6. *I* and *K* represent the image data before and after coding respectively, ΔP can represent the change of peak signal-to-noise ratio before and after the introduction of convolutional neural network.

TABLE II RESULTS OF VIDEO SEQUENCE TEST

video sequence	*QP*	B_{HM}	B_{imp}	Δ *B*/%	P_{HM}	P_{imp}	Δ *P*/dB	T_{HM}	T_{imp}	Δ *T*/dB
BlowingBubbles 416×240	22	2132.61	2144.81	0.57	39.32	38.98	-0.34	267.33	200.31	-25.07
	27	1653.22	1649.24	-0.24	34.58	35.34	0.76	203.44	184.21	-9.45
	32	932.57	950.33	1.90	30.22	29.48	-0.74	183.14	152.89	-16.52
	37	720.33	713.92	-0.89	28.43	27.98	-0.45	154.32	121.32	-21.38
BasketballDrill 832×480	22	3948.48	3894.81	-1.36	41.42	40.99	-0.43	1182.98	828.32	-29.98
	27	1932.24	2003.24	3.67	37.82	38.12	0.3	822.35	692.78	-15.76
	32	1032.32	989.53	-4.15	30.28	29.45	-0.83	683.21	512.34	-25.01
	37	523.28	531.34	1.53	30.12	29.49	-0.63	422.89	329.2	-22.15
KristenAndSara 1280×720	22	2284.82	2301.23	0.72	39.21	38.32	-0.89	203.42	167.19	-17.81
	27	1783.69	1803.94	1.13	36.18	35.96	-0.22	159.32	130.78	-17.91
	32	1327.54	1314.22	-1.00	32.32	33.01	0.69	110.13	98.43	-10.62
	37	932.45	953.21	2.23	29.31	30.43	1.12	78.32	74.58	-4.78
Kimono1 1920×1080	22	6892.32	6832.81	-0.86	41.56	40.31	-1.25	6365.4	4893.1	-23.13
	27	3643.88	3627.42	-0.45	40.18	39.71	-0.47	5425.2	4013.7	-26.02
	32	1244.7	1236.51	-0.66	35.09	34.98	-0.11	4835.7	3378.4	-30.14
	37	653.32	633.31	-3.06	32.21	32.10	-0.11	3593.7	1873.4	-47.87
Traffic 2560×1600	22	14893.41	14801.31	-0.62	41.89	41.12	-0.77	9061.31	6942.53	-23.38
	27	7832.43	7743.54	-1.13	38.32	37.94	-0.38	7323.66	6124.34	-16.38
	32	3421.54	3413.28	-0.24	35.43	34.89	-0.54	6072.41	5743.65	-5.41
	37	1289.73	1279.39	-0.80	32.58	31.88	-0.70	5909.62	5083.68	-13.98
mean	-	2953.74	2940.85	-0.19	35.32	35.02	-0.30	2652.89	2077.26	-20.14

According to the test results of video sequence in Table II, the HEVC standards software HM16.0 is as the benchmark, the mode decision algorithm for intra prediction angle using convolutional neural network is compared with the standards

software HM16.0. On the five resolutions of video streams, the average bit rate is reduced by about 0.19%, the average peak signal-to-noise ratio is reduced by about 0.3%, and the average total coding time is reduced by about 20.14%.

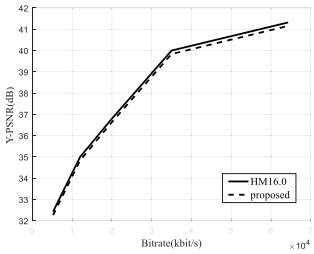

Fig. 4 Comparison of rate distortion curve of kimono1 video sequence

According to the comparison of rate distortion curves of kimono1 video sequence in Figure 4, taking kimono1 video sequence as an example, the rate distortion curves coincide basically before and after the improvement of intra encoding, that is, there is little difference between the image quality and code rate of the two algorithms, and the encoding results of convolutional neural network are acceptable. The coding performance of the algorithm in this paper is consistent with HEVC standards test software HM16.0. However, after introducing convolutional neural network for intra mode decision, the coding time is less than HM16.0 standard software, which proves the feasibility of the algorithm in this paper.

V. Conclusion

In this paper, the mode decision with high complexity in HEVC is improved partially by using the deep learning method, and a fast mode decision algorithm in HEVC based on convolutional neural network is proposed, that is, the rate distortion encoding module with high complexity in HEVC is replaced by convolutional neural network. By using kimono and other standard video sequences with different resolutions, it is found that the algorithm in this paper is similar to HEVC standard software HM16.0 in video bit rate and peak signal-to-noise ratio, with a improvement of 20.14% in the total encoding time. Therefore, by using convolutional neural network instead of rate distortion optimization module, combining with GPU hardware to accelerate calculation, this algorithm overcomes the problems of long coding time and high algorithm complexity of traditional methods to a certain increase. The disadvantage is that other coding modules are not fully considered. In the future, related deep learning algorithms can be used to improve other modules in HEVC.

Acknowledgement

The work is supported by the Key R & D plan of Shaanxi Province(2020ZDLGY06-01) and Science & Technology Innovation Guidance Project of Xi'an, China (21JY033). Science & Technology Innovation Guidance Project of Xi'an, China (No. 2020KJRC0087).

References

[1] Cui Pengtao, Zhang Qian, Liu Jinghuai, Zhou Chao, Wang Bin, Si Wen Fast intra prediction mode decision algorithm for depth image based on FSCD-CNN [J] Journal of Applied Science, 2021, 39(03): 433-442.

[2] Liu R, Li S, Hou C, et al. Multiple Residual Learning Network for Single Image Super-Resolution[C]. IEEE Visual Communications and Image Processing, 2018: 1-4.

[3] Li D, Liu Y, Wang Z. Video Super-Resolution Using Non-Simultaneous Fully Recurrent Convolutional Network[J]. IEEE Transactions on Image Processing, 2019, 28(3): 1342-1355.

[4] Li Wei, fan Caixia H. 266 / VVC intra prediction mode fast decision method [J] Computer engineering, 2021, 47(10): 221-225.

[5] Merkle P, Müller K. Depth Intra Coding for 3D Video Based on Geometric Primitives[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(3): 570-582.

[6] Si Xiaohua, Zhongguo Kingdom, Zhao Haiwu, Li Guoping, Teng Guowei Fast adaptive intra prediction mode decision algorithm for depth map [J] Journal of Shanghai University (NATURAL SCIENCE EDITION), 2015, 21(02): 197-205.

[7] Liu Yong, Xu Dawen. HEVC Information-Hiding Algorithm Based on Intra-Prediction and Matrix Coding[J]. International Journal of Digital Crime and Forensics (IJDCF), 2021, 13(6): 1792-1800.

[8] High cypress HEVC chroma intra prediction mode search optimization [J] Modern computer, 2021(17): 60-65.

[9] Younesi, Reza,Rastegar Fatemi, Mohammad Jalal, Rastgarpour, Maryam. Area-efficient HEVC core transform using multi-sized and reusable DCT architectures[J]. Multimedia Tools and Applications, 2021(prepublish): 957-968.

[10] Jia Kebin, Cui tenghe, Liu Pengyu, Liu Chang Fast intra prediction algorithm for efficient video coding based on deep feature learning [J] Journal of electronics and information, 2021, 43(07): 2023-2031.

[11] Paulraj Ranjith Kumar,M. Vimala,P. Govindamoorthi. An optimal weighted HEVC coding for video compression[J]. Multimedia Tools and Applications, 2021(prepublish): 899-921.