ResNet Oriented Fast Mode Decision Algorithm for HEVC Intra Coding

Da Ai, Yang Gao, Nam Ling and Ying Liu

November 16, 2018

ResNet Oriented Fast Mode Decision Algorithm for HEVC Intra Coding

Da Ai[1], Yang Gao[1*], Nam Ling[2], Ying Liu[1]

[1]The Key Laboratory of Electronic Information Application Technology of Site-survey of The Ministry of Public Security,

Xi'an University of Posts and Telecommunications, Xi'an, 710121, China

aida@xupt.edu.cn, gy_690278129@163.com, ly_yolanda@sina.com

[2]Department of Computer Engineering, Santa Clara University, Santa Clara, CA 95053, USA

nling@scu.edu

Abstract. The High Efficiency Video Coding (HEVC) standard improves video coding efficiency remarkably at the cost of the highly increased computational complexity. A fast algorithm of coding unit (CU) depth prediction and intra-mode decision was proposed to simplify the CU partitioning and optimal mode decision procedure of intra coding. Firstly, a residual network (ResNet) was trained with the results of HEVC CU depth prediction and intra-prediction mode and the corresponding video sequence. The CU partition process of HEVC is replaced by this trained ResNet that can reduce the coding time effectively. Meanwhile, the process of rough mode selection (RMD) is simplified by the texture direction matching on the current prediction unit (PU) with lower computational complexity. The experimental results show that the proposed algorithm could reduce around 71.2% encoding time with BDBR increased by 1.75% and BDPSNR decreased by 0.093dB on average in all I frame encode scheme compared with HEVC testing model HM-16.7.

Keywords: High Efficiency Video Coding (HEVC), intra prediction, ResNet, coding unit (CU) depth prediction, intra-mode decision.

1 Introduction

The high efficiency video coding (HEVC) standard is developed by the Joint Collaborative Team on Video Coding [1]. Compared with h.264 /AVC, HEVC can better realize the coding operation of high resolution video, and the coding bit rate is reduced by nearly 50% under the same video quality perception [2, 3]. Within HEVC, intra coding adopts a more flexible structure and introduces coding unit (CU), prediction unit (PU) and transform unit (TU) [4]. In addition, the number of intra-prediction modes has increased from 9 to 35 [5], which improve the coding performance and increase the coding complexity greatly. In order to simplify rate-distortion optimization (RDO) intra-mode selection, HEVC reference model HM adopts rough mode decision (RMD) [6].

There have many optimal algorithms been proposed can achieve good experimental results of CU partition judgment and intra-mode decision. Pan et al. [7] used the correlation between adjacent CUs, combined with different weights to predict the depth traversal range of the current CU, so as to reduce the coding time. Liu et al. [8] proposed a fast coding algorithm based on texture complexity, and selectively skipped or terminated depth partition by calculating whether the complexity coefficient of CU block over the preset threshold. Although this method can save the coding time by 35% compared with the HM test model, the threshold selection is empirically depend on the statistical analysis of the video content.

For the intra-mode decision algorithm, Liu et al. [9] divided 33 prediction modes into 8 groups by analyzing the texture direction of PU pixel blocks with size of 4×4, and obtained the corresponding mode list instead of the RMD. Chen et al. [10] used pixel gradient to represent texture information, and calculated the Sobel processing results of 8×8 pixel blocks, and the depth range and the mode list of the current block were determined according to the prediction angle of pixel points. Compared with the HM test model, the coding time is reduced by 29.5% on average. The performance of the algorithm mentioned above is affected by the threshold selection and additional computational procedure, and their improvement is limited.

By introduce the machine learning method, the optimal algorithm based on image content classification has more advantages, and the coding time can be further reduced. Liu et al. [11] divided the CU texture structure into three categories through SVM, and obtained the judgment result of whether the corresponding depth was split, which decreased the coding complexity to a certain extent and reduced the time by 53%; Hu et al. [12] used logistic regression to establish a two-class model, which was optimized by various feature indicators and improved evaluation functions to achieve the different size of CU classification of the accurate judgment, and reduced the coding time by 55.51%; Li et al. [13] took CU of different sizes as training data, and obtained three kinds of binary prediction results through CNN, which respectively represent the three-order division judgment

* Corresponding Author

of depth 0-3; Kim et al. [14] proposed a fast CU depth-decision algorithm based on CNN, which inferenced whether to split at each depth from depth 0 to 2. By comparing the optimization results, it can be seen that machine learning can better reduce the complexity of intra coding. However, the limitations of the CNN model hinder the improvement of the prediction performance, and the prediction results need further analysis before they can be used for coding operation.

In this paper, fast CU depth prediction and intra-mode decision are proposed to decrease the computational complexity of HEVC intra coding. The network model adopts the residual network (ResNet) [15], combined with some optimization analysis to improve the classification accuracy, and finally obtains the depth prediction values of the 16×16 CUs, to meet the requirement of reducing complexity. At the same time, texture analysis and depth prediction results are combined to simplify the decision process of intra-mode list and further reduce the intra coding complexity.

The rest paper is organized as follows. Section 2 provides a CU depth prediction model based on ResNet. The intra-mode decision optimization method based on texture direction and the depth been predicted is described in Section 3. The experiment results are discussed in Section 4. Finally, the conclusions are given in Section 5.

2 CU Depth Prediction Model Based on ResNet

2.1 ResNet Structure

ResNet model has achieved better application effect in the field of image recognition, which solves the gradient disappearing phenomenon of ordinary CNN model when the amount of data is large or the number of convolution layers is high. According to the existing theoretical research, in the deep CNN structure, it is necessary to apply the identity mapping to the added network to construct the network model, so as to improve the performance of the shallow network model. ResNet introduces the residual fitting method, and when the network depth is large enough, the identification accuracy can still increase stably. If the actual mapping value is represented by pre(x), the convolution result F(x) is expressed as F(x) = pre(x) - x, which can simplify the optimization problem to solve the minimum value of F(x). In this way, the identity mapping can be approximated, the accuracy degradation problem in ordinary networks can be solved, and the prediction accuracy can be improved. The original residual basic structure (Res_block) is shown in Fig. 1.

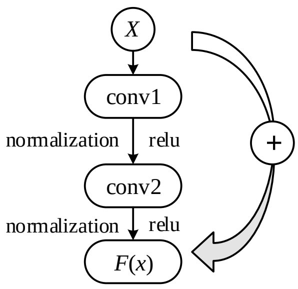


Fig. 1. Basic structure of Res_block

Relu is selected as the activation function, which can effectively solve the problem of gradient disappearance and explosion in the residual network. In a residual unit, the extracted characteristic relationship is:

$$f_{i+1}(x) = \left[max\left(0, f_i(x) \cdot w_0^i + b_0^i\right) \cdot w_1^i + b_1^i\right] + f_i(x) \,. \tag{1}$$

where the $f_i(x)$ and $f_{i+1}(x)$ are the input and output matrices of the *i-th* a residual unit, w and b represent the weight matrix and bias, $max\left(0, f_i(x) \cdot w_0^i + b_0^i\right)$ is activation function.

2.2 Training

The maximum depth of the intra prediction corresponds to a CU size of 8×8, and the depth within the CU of 16×16 is the same, which can be used as the matrix dimension of the data set. The 16×16 pixel blocks are adopted as the input data which can extract image features effectively and avoid the dimension reduction of network. The diversity of training data is an important factor to construct a robust network. The content, texture, moving properties are different in each video sequence, even the coding results of the same CTU for different Quantization Parameters (QP) in one video clip are also different. Hence, the images in every 50 frames in different video sequence with different resolutions are selected to form the data set. The coding depth of the corresponding different QP values {22, 27, 32 and 37} is selected as the label set. The specific data set is shown in Table 1.

Table 1. Experimental data set composition

Classes	Sequence	Number of 16x16 pixel blocks			
		Depth= 0	Depth= 1	Depth=2	Depth=3
Class A (2560×1600)	PeopleOnStreet	2500	1500	2000	2000
	Traffic	2500	1500	2000	2000
Class B (1920×1080)	BasketballDrive	2200	1500	1500	1600
	BQTerrace	2000	1500	1400	1500
	Cactus	2500	1500	1500	1500
	ParkScene	2000	2000	1500	1500
	Kimono	2500	2000	1500	1500
Class C (832×480)	BasketballDrill	1500	1200	1500	1500
	BQMall	500	1500	2000	1500
	PartyScene	0	1000	1500	1500
	RaceHorses	700	1500	1500	1500
Class D (416×240)	BasketballPass	350	950	800	800
	BQSquare	0	1200	1000	1500
	BlowingBubbles	50	350	1500	1500
	Flowervase	1000	800	600	400
Class E (1280×720)	FourPeople	2000	1500	1200	1200
	Johnny	1500	1500	1000	1000
	KristenAndSara	1200	2000	1000	1000
	Total	25000	25000	25000	25000

The ResNet model used in this paper, as shown in Fig. 2, which includes one input layer, nine residual units and two fully connected layers. The learning platform is the GPU version of tensorflow [16], which realizes the classification of the input pixel blocks and introduces the QP value into the image feature judgment. In the system of network predictive value evaluation and optimization, the loss function is based on the cross entropy calculation method, and the optimizer is MomentumOptimizer, which evaluates the network through top-1 error.

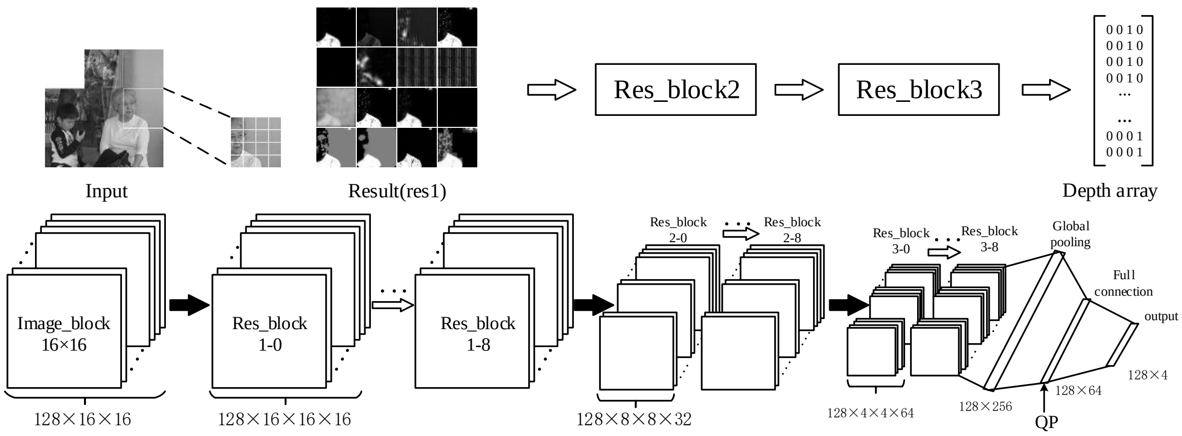


Fig. 2. Structure of ResNet model

Input layer. Extract 128 image blocks randomly and normalize them.

Residual unit. The network consists of nine residual units containing 56 convolution layers in total. A single residual unit includes three Res_blocks, one data processing layer and one global pooling layer, where there is also pooling operation between Res_blocks. At the same time, the output results of each convolution layer are regularized and the excitation function is processed to further solve the data over-fitting.

Fully connected layer. After extracting the features of the residual units, the fitting result of the image blocks array are obtained through the full connection layer, and the predicted results are finally obtained by softmax. The global pooling process changes the matrix into an image vector, which is passed into the full connection layer together with QP, and finally obtains the calculation results of image blocks in the output layer. In addition, dropout [17] is used in the network to improve the performance of the deep network by discarding half of the neurons randomly. The flow of intra depth optimization algorithm is based on ResNet, as shown in Fig. 3:

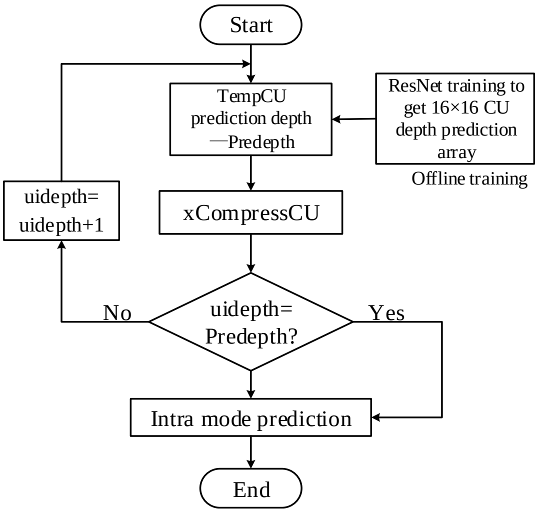
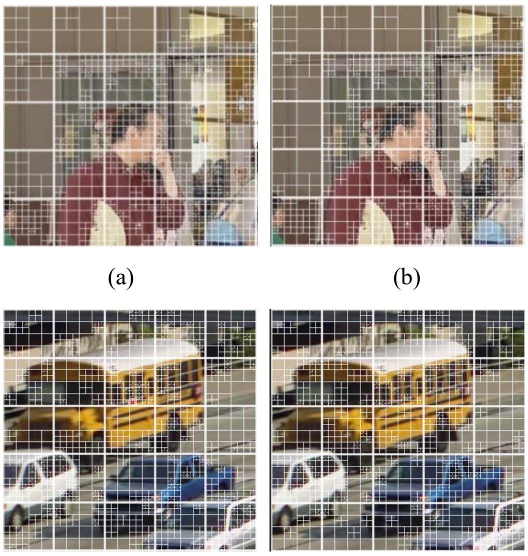


Fig. 3. Depth prediction optimization algorithm flow

The training recognition rate of the ResNet been used is up to 92.6%, while the prediction accuracy of the model is nearly 86%. When the training network model is basically stable, a 16×16 pixel matrix of one frame video image is sequentially passed through the model to obtain a depth prediction value, and the predicted result will be different with different QP value. In order to verify the accuracy of the algorithm, BQMall and Traffic are selected to obtain CU partitioning result of the original HM model and the Prop.depth (proposed depth prediction in abbreviation) model respectively in the case of QP = 32, as shown in Fig. 4:



(a) (b)

(c) (d)

Fig. 4. CU partitioning result of Prop.depth and HM model when QP=32. (a) BQMall, original HM. (b) BQMall, Prop.depth. (c) Traffic, original HM. (d) Traffic, Prop.depth

3 Intra-mode Decision Optimization Based on Texture Direction

In the intra coding process of HEVC, the optimal intra mode of PU has a certain correlation with its texture direction, the SATD value and RD-cost value [18, 19], all of them can be used as an analysis method of PU's own characteristics. During RMD calculations, the intra modes are not optimal when they have the same SATD values. In other words, the mode with the minimum SATD value is unique and the RMD can be optimized based on this analysis. Firstly, the matching angle mode can be obtained by comparing the different texture directions of PU. Then it forms a predictive mode list with its adjacent angle mode, adding mode 0 and mode 1. Secondly, a new candidate mode list can be formed by MPM and the mode select results, which are obtained by comparing the SATD values of each mode. Finally the optimal mode is obtained through RDO process. This algorithm simplifies RMD process and reduces the number of modes in the candidate mode list, thus reducing the coding time and coding bits.

3.1 Texture Direction of PU

Combined with the CU depth prediction results of ResNet model in the previous section, the texture direction of PU at current depth is further analyzed to calculate the optimal prediction direction of a certain size from 4×4 to 64×64 PU. As shown in Fig. 5, eight PU prediction modes is corresponding to eight different texture directions. Mean of the absolute difference (MAD) of image pixel block corresponding to 8 texture directions under the same size is calculated respectively, and the corresponding direction of the minimum value is the predictive result of the PU texture direction.

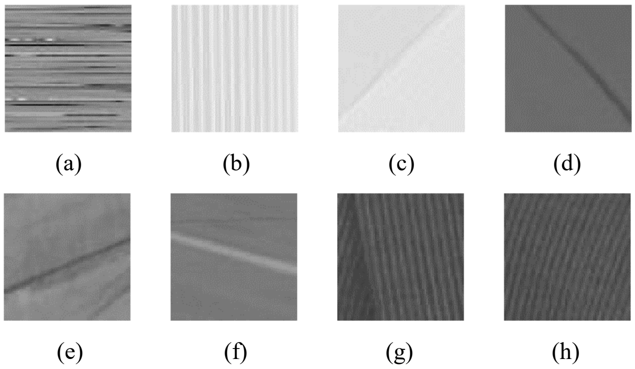


Fig. 5. Eight texture directions in PU corresponding to different prediction modes.
(a) Mode=10, Horizontal. (b) Mode=26, Vertical. (c) Mode=2, Diagonal Down. (d) Mode=18, Diagonal Up. (e) Mode=6, Horizontal Down. (f) Mode=14, Horizontal Up. (g) Mode=22, Vertical Left. (h) Mode=30, Vertical Right

1. The calculation method of vertical direction is similar to horizontal direction. For the MAD calculation result of vertical direction: TD_{ver}, can be obtained by the following calculation method:

$$TD_{Ver} = \frac{1}{N}\sum_{n=0}^{N-1}\left(\frac{1}{N}\sum_{i=0}^{N-1}|P(i,n) - M_{Ver}(n)|\right). \quad (2)$$

where: the size of PU is N×N (N=4,8,16,32,64), and $P(i,n)$ represents the pixel value of the *n-th* column in PU (i=0,1,2,…, N; n = 0,1,…,N), $M_{Ver}(n)$ represents the mean of pixel values of the *n-th* column:

$$M_{Ver}(n) = \frac{1}{N}\sum_{i=0}^{N} p(i,n). \quad (3)$$

2. Compared with 1, the calculation methods of Dig-D and Dig-U are quite different, including the number of sub-columns *n*, the number of pixel values and mean values of each sub-column. The MAD value of DU direction is calculated by the following method:

$$TD_{DU} = \frac{1}{2N-3}\sum_{n=1}^{2N-3}\left(\frac{1}{N-|n-N+1|}\sum_{i+j=n}|p(i,j) - M_{DU}(n)|\right). \quad (4)$$

$$M_{DU}(n) = \frac{1}{N-|n-N+1|}\sum_{i+j=n} p(i,j). \quad (5)$$

3. For the MAD value of Ver-L, Ver-R, Hor-U and Hor-D direction in the Fig. 5, in order to reduce the amount of calculation, the calculation results of 1 and 2 can be respectively expressed as:

$$\begin{cases} TD_{HD} = 1/2 \times (TD_{Hor} + TD_{DD}) \\ TD_{HU} = 1/2 \times (TD_{Hor} + TD_{DU}) \\ TD_{VL} = 1/2 \times (TD_{Ver} + TD_{DU}) \\ TD_{VR} = 1/2 \times (TD_{Ver} + TD_{DD}) \end{cases}. \quad (6)$$

3.2 Prediction Mode Optimization

An angle mode TD_{mode} matching the PU texture direction is obtained by the above calculation method before the RMD process of intra-mode prediction. At the same time, as shown in Table 2, several angle prediction modes adjacent to TD_{mode} need to be added into the group, while Planar mode and DC mode do not participate in texture direction prediction, so it is also necessary to add mode 0 and mode 1 into each mode list to form 8 mode lists group.

Table 2. Mode list corresponding to the texture direction

Texture Direction (TD)	Angle mode	Mode list
Diagonal Down (DD)	2	{0,1,4,3,2,34,33,32}
Horizontal Down (HD)	6	{0,1,4,5,6,7,8}
Horizontal (Hor)	10	{0,1,8,9,10,11,12}
Horizontal Up (HU)	14	{0,1,12,13,14,15,16}
Diagonal Up (DU)	18	{0,1,16,17,18,19,20}
Vertical Left (VL)	22	{0,1,20,21,22,23,24}
Vertical (Ver)	26	{0,1,24,25,26,27,28}
Vertical Right (VR)	30	{0,1,28,29,30,31,32}

Finally, by comparing the SATD values of different modes for rough selecting, the number of modes (*Num*) in the list is changed from {3, 3, 3, 8 and 8} to {2, 2, 2, 3 and 3}. The main mode selection algorithm is mainly described as follows:

Step1. The SATD values are obtained and sorted one by one from the mode list corresponding to the texture direction. If the SATD values of any two modes are the same, they will be deleted.

Step2. The *Num* modes with smaller SATD value are retained. If the number of remaining modes is 0, the mode with the smallest SATD value in TD_{mode} and mode 0, 1 is selected as the result. If the number of remaining modes is not 0 and less than *Num*, these modes are selected as the result.

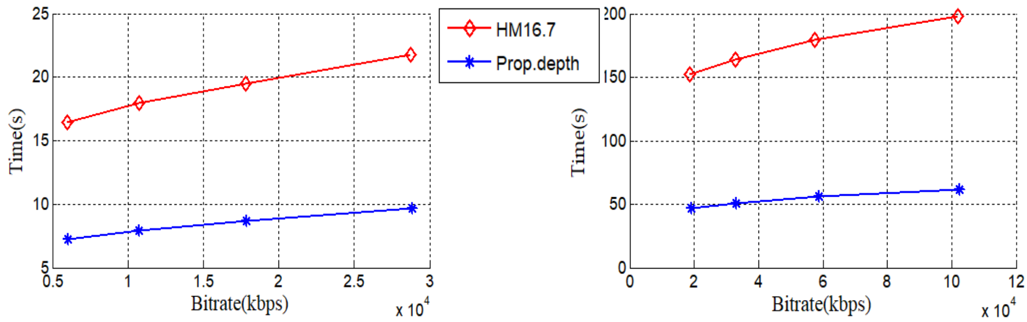
Step3: Add MPM to the Step2 mode list and calculate RD-cost to get the best mode.

4 Experimental Results

According to the above optimization scheme, experimental simulation is carried out based on the HEVC test model HM16.7 [20]. The experimental hardware platform is Intel(R) Core(TM) i7-6820hk CPU @2.70GHZ with 4 cores, 16.0GB RAM and Geforce GTX 1080M, and the operating system is Windows 10. The compilation and debugging software is provided for Visual Studio 2013 and Tensorflow learning system, the configuration parameter is the all I frame coding model, using 18 video sequences with the QP values 22, 27, 32 and 37. The experimental data uses video sequences with different resolutions and complexity to obtain Bitrate, PSNR and total Time corresponding to different QP values. Then, coding performance is represented by evaluation indexes BDBR and BDPSNR [21], and coding complexity is represented by $\Delta Time$, with calculation formula as follows:

$$\Delta Time = \frac{Time_{HM} - Time_{prop}}{Time_{HM}} \times 100\% \,. \tag{7}$$

where the $Time_{prop}$ is the coding Time of this paper, the $Time_{HM}$ is the coding Time of HM16.7. As shown in Fig. 6, the performance comparison results of the BQMall sequence and Traffic sequence are encoded respectively by the HM16.7 model and the Prop.depth optimization model.



(a) BQMall (b) Traffic

Fig. 6. Comparison of intra coding performance between HM16.7 and Prop.depth model

Table 3 shows encoding performance of the ResNet model (Prop.depth) in this paper, Li et al. [13] and k. Kim et al. [14] versus the HM16.7.

Table 3. Comparison of encoding performance of CNN model

	BDBR(%)	BDPSNR(dB)	ΔT(%)
Prop.-depth	1.42	-0.074	60.4
Li et al. [13]	2.15	-0.12	61.7
Kim K et al. [14]	4.07	-0.27	60.9

There are 18 video sequences with 4 different QP (22, 27, 32 and 37) have been tested in the experiment. In each sequence, the mean of the four QP using the depth optimization algorithm and the overall optimization algorithm (Prop.overall) can be seen in Table 4. The Prop.depth of this paper has significant advantages compared with the HM16.7 that the coding time reduction can be up to 75%, and the average reduction is 71%. Meanwhile, the increase of BDBR is 1.75%, and the average decrease of BDPSNR is 0.093db on average.

Table 4. Video sequence coding result of BDBR(%) , BDPSNR(dB) and ΔT(%)

Sequences	Prop.depth			Prop.overall		
	BDBR	BDPSNR	ΔT	BDBR	BDPSNR	ΔT
PeopleOnStreet	1.55	-0.069	60.1	1.55	-0.092	73.7
Traffic	1.32	-0.076	61.3	1.51	-0.086	71.7
BasketballDrive	1.79	-0.057	64.8	1.82	-0.068	72.3
BQTerrace	1.57	-0.073	59.0	1.54	-0.089	71.1
Cactus	1.22	-0.069	57.1	1.59	-0.070	70.8
ParkScene	0.88	-0.041	58.4	1.04	-0.056	71.3
Kimono	0.78	-0.045	69.2	0.90	-0.054	74.5
BasketballDrill	2.06	-0.107	56.2	2.15	-0.133	70.5
BQMall	1.53	-0.061	55.4	1.86	-0.079	69.6
PartyScene	1.15	-0.058	52.3	1.27	-0.084	67.7
RaceHorses	1.08	-0.060	56.7	1.23	-0.081	70.3
BasketballPass	1.66	-0.072	57.3	1.79	-0.074	71.8
BQSquare	1.11	-0.108	59.7	2.33	-0.118	66.2
BlowingBubbles	0.76	-0.096	48.1	2.46	-0.134	64.4
Flowervase	0.62	-0.053	68.9	1.06	-0.063	72.9
FourPeople	1.98	-0.085	61.3	2.14	-0.101	71.8
Johnny	2.26	-0.098	70.2	2.52	-0.139	75.6
KristenAndSara	2.34	-0.116	71.5	2.83	-0.143	75.0
Average	1.42	-0.074	60.4	1.75	-0.093	71.2

As shown in Table 5, the coding performance comparison of the proposed overall optimization algorithm, literature [22] and literature [23] (mean values under 4 QP conditions) is presented, all of them reduce the intra coding complexity from the perspective of depth prediction and the mode decision.

Table 5. Algorithm optimization performance comparison

Classes	J. Gu et al.[22]		T. Zhang et al. [23]		Prop.overall	
	BDBR	ΔT	BDBR	ΔT	BDBR	ΔT
Class A	1.23	64.9	0.92	61.5	1.53	72.7
Class B	0.99	63.5	0.71	58.6	1.37	72.0
Class C	1.27	61.7	0.80	48.7	1.62	69.5
Class D	1.22	61.5	1.38	47.5	1.91	68.8
Class E	1.97	66.8	1.04	65.2	2.49	74.1
Average	1.34	63.7	0.97	56.3	1.78	71.4

5 Conclusion

In this paper, a fast intra coding optimization algorithm based on ResNet is proposed, and constructs an intra coding depth prediction model through the ResNet, saving the coding time of non-predicted depth in CTU recursive partitioning. At the same time, combined with texture direction judgment and SATD value analysis, the RMD process is optimized to reduce the computational complexity of RDO further. Compared with other fast coding algorithms based on CNN, the network model of this paper has better coding performance. Experimental results show that, the algorithm has a significant reduction in coding time and a good optimization effect on video sequences with different resolutions comparing with HM16.7. Furthermore, the classification optimal neural network still has a great potential and better prospect for next generation video coding.

Acknowledgement

This work is supported by the Scientific Research Plan Projects of Shaanxi Provincial Department of Education(18JK0716).

References

[1] G.J. Sullivan, J.R. Ohm, W.J. Han, et al. Overview of the High Efficiency Video Coding (HEVC) Standard[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2012, 22(12): 1649-1668.

[2] T. Wiegand, G.J. Sullivan, G. Bjøntegaard, et al. Overview of the H.264/AVC video coding Standard[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2003, 13(7):560-576.

[3] J. Ohm, G.J. Sullivan, H. Schwarz, et al. Comparison of the Coding Efficiency of Video Coding Standards-Including High Efficiency Video Coding (HEVC)[J]. IEEE Transaction on Circuits & Systems for Video Technology, 2012, 22(12): 1669-1684.

[4] I.K. Kim, J. Min, T. Lee, et al. Block Partitioning Structure in the HEVC Standard[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2012, 22(12): 1697-1706.

[5] Bossen, Frank, et al. Intra Coding of the HEVC Standard[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2013, 22(12): 1792-1801.

[6] L.L. Wang, W.C. Siu. Novel Adaptive Algorithm for Intra Prediction With Compromised Modes Skipping and Signaling Processes in HEVC[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2013, 23(10): 1686-1694.

[7] Z. Pan, S. Kwong, Y. Zhang, et al. Fast Coding Tree Unit depth decision for high efficiency video coding[A]. IEEE International Conference on Image Processing[C]. IEEE, 2015: 3214-3218.

[8] Y. Liu, X. Liu and P. Wang. A Texture Complexity Based Fast Prediction Unit Size Selection Algorithm for HEVC Intra-coding[A]. 2014 IEEE 17th International Conference on Computational Science and Engineering[C]. Chengdu, China: IEEE, 2015: 1585-1588.

[9] X. Liu, Y. Liu, P. Wang, C. Lai and H. Chao. An Adaptive Mode Decision Algorithm Based on Video Texture Characteristics for HEVC Intra Prediction[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(8): 1737-1748.

[10] H. Chen, R. Xie, L. Zhang. Gradient based fast mode and depth decision for high efficiency intra frame video coding[A]. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting[C]. IEEE, 2015: 1-6.

[11] D. Liu, X. Liu, Y. Li. Fast CU Size Decisions for HEVC Intra Frame Coding Based on Support Vector Machines[A]. IEEE, Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress[C]. IEEE, 2016: 594-59.

[12] Q. Hu, Z. Shi, X. Zhang and Z. Gao. Fast HEVC intra mode decision based on logistic regression classification[A]. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting[C]. IEEE, 2016: 1-4.

[13] T. Li, M. Xu, X. Deng. A deep convolutional neural network approach for complexity reduction on intra-mode HEVC[A]. IEEE International Conference on Multimedia and Expo[C]. IEEE, 2017: 1255-1260.

[14] K. Kim and W.W. Ro. Fast CU Depth Decision for HEVC using Neural Networks[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, (99): 1-1.

[15] Kaiming He, X. Zhang, S. Ren, et al. Deep residual learning for image recognition[J]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[16] M. Abadi, A. Agarwal, P. Barham, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems[J]. 2016.

[17] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R.R. Salakhutdinov. Improving Neural Networks by Preventing Coadaptation of Feature Detectors[J]. Computer Science, 2012, 3(4): 212–223.

[18] Z. Liu, X. Yu, S. Chen, et al. CNN oriented fast HEVC intra CU mode decision[A]. IEEE International Symposium on Circuits and Systems[C]. IEEE, 2016.

[19] J. Gu, M. Tang, J. Wen and H. Zhang, A novel satd based fast intra prediction for HEVC[J], 2017 IEEE International Conference on Image Processing (ICIP), Beijing, 2017: 1667-1671.

[20] HEVC Test Model (HM reference software)[J]. 2011.

[21] G. Bjøntegaard, Calculation of Average PSNR Differences Between R-D Curves[J]. 2001.

[22] J. Gu, M. Tang, J. Wen and Y. Han. Adaptive Intra Candidate Selection with Early Depth Decision for Fast Intra Prediction in HEVC[J]. IEEE Signal Processing Letters, 2018, 25(2): 159-163.

[23] T. Zhang, M. Sun, D. Zhao and W. Gao. Fast Intra-Mode and CU Size Decision for HEVC[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 27(8): 1714-1726.