Academic Journal of Engineering and Technology Science
ISSN 2616-5667 Vol.3, Issue 2: 73-82, DOI: 10.25236/AJETS.2020.030210

Fast Intra Mode Coding Based on Convolutional Neural Network

Chengsi Lin*, Qingming Yi

School of information science and technology, Jinan University, Guangzhou 510632
**Corresponding author e-mail: linchengsi123@stu2017.jnu.edu.cn*

ABSTRACT. *In order to better adapt to the different texture features of video images, the number of intra coding modes in the new generation of video coding standard h.265/HEVC (high efficiency video coding) has increased to 35, which not only achieves better coding performance but also increases the computational complexity. In order to reduce the complexity of intra coding, a fast intra prediction method based on convolutional neural network (CNN) is proposed. For 4x4 or 8x8 PU, this paper gets the list of candidate modes by CNN, skipping the rough mode decision (RMD) process of prediction unit (PU). In this paper, the algorithm is embedded in HEVC coding framework, which effectively reduces the redundant intra prediction process in all intra configuration. The experimental results show that compared with HEVC official test model (HM16.12), the coding time of the algorithm proposed in this paper is reduced by 28.08% on average, while that of BD_BR and BD_PSNR is only 1.14% and - 0.055db.*

KEYWORDS: *High Efficiency Video Coding (HEVC), Intra Prediction, Deep Learning, CNN*

1. Introduction

H. 265 / HEVC (high efficiency video coding) is a new generation of high efficiency video coding and compression standard after h.264/AVC jointly issued by ITU-T and ISO / IE in 2013 [1]. HEVC standard aims to reduce the bit rate by 50% compared with h.264/AVC on the premise of ensuring the video coding quality [1], so HEVC introduces a series of new technologies. But at the same time, the coding complexity has increased dramatically, which makes HEVC have a higher threshold for the computing power of devices in multimedia applications. Therefore, HEVC needs to reduce the coding complexity through algorithm optimization.

In the intra coding mode, PU mode decision occupies a large part of the coding time consumption. In the process of intra mode prediction, the amount of computation is mainly distributed in the rough PU mode decision (RMD) process and the total rate distortion optimization (RMD) process. In order to fit different texture information of different video images, HEVC applies five different sizes of

ISSN 2616-5667 Vol.3, Issue 2: 73-82, DOI: 10.25236/AJETS.2020.030210

PU (prediction unit), each of which has 35 intra prediction modes. During encoding, it traverses all these modes to obtain the best encoding effect. In the intra prediction mode decision-making, the first step is the rough mode decision-making (RMD) stage, which tests mode 0 (DC mode), mode 1 (planar mode) and 33 different angle modes to obtain a list of candidate modes with the sum of the minimum absolute difference (SATD). Then, by adding the most likely mode (MPM), the possible mode of the current PU is inferred according to the prediction information of the adjacent coding units. Finally, through the time-consuming rate distortion optimization (RDO) stage, the rate distortion cost (RDC) of the candidate modes in the list of candidate modes is calculated to finally select the optimal intra prediction angle mode.

For RMD stage, some methods use edge strength extraction to obtain candidate PU list. For example, in reference [5], before selecting intra prediction mode, Sobel operator is used to calculate the horizontal and vertical gradient of current PU pixel points, calculate the edge vector angle of pixel points, and generate histogram by calculating the pixel angle and amplitude, and select the possible prediction mode from the generated histogram distribution In order to reduce the computation complexity of RMD and RDO process, the best model is obtained by entering RMD and RDO process. In reference [6], a two-step algorithm for quick mode selection is proposed. Firstly, the first mode to enter RMD is constructed by using PU size along four main directions. The larger PU is, the fewer modes are. After RMD, the second RMD calculation is performed by adding adjacent modes according to the first several candidate modes. In reference [7], a candidate pattern selection algorithm is proposed by analyzing the texture of the source image block. Considering the difference between adjacent prediction directions, an edge detector based on fixed-point algorithm is designed, and the concept of kernel density estimation is further introduced into histogram calculation. In reference [8], this algorithm not only considers the modes (prediction units) of adjacent PU, but also uses the edge information of current PU to select the best simplified direction set of prediction mode (direction). In reference [10], a two-stage angle pattern elimination method is proposed to reduce the current PU candidate list based on texture information. In reference [11], a method based on convolutional neural network is adopted to reduce the number of prediction modes of 4x4 and 8x8 size PU, thus reducing the computational complexity. In reference [12], a method of intra prediction mode based on decision tree is proposed to calculate the variance of all PU's upper, left and reference pixels, and then the PU prediction mode is divided into three groups by using software Weka to train the decision tree, so as to reduce the calculation amount of the mode.

The disadvantage of the above methods is that the selection of intra prediction mode is not only determined by a single edge direction and texture information, but also includes edge curvature, texture topology and quantization parameters. It is not accurate to use a single index to predict the intra prediction mode of PU. In this paper, for the smaller PU that is difficult to predict, we use CNN (convolutional neural network) to classify the patterns to get the candidate patterns, thus skipping the RMD and MPM process of PU intra prediction.

ISSN 2616-5667 Vol.3, Issue 2: 73-82, DOI: 10.25236/AJETS.2020.030210

The structure of this paper is as follows: in Section 2, we describe the CNN architecture used in this paper. Then, in Section 3, we explain the experimental results, and the last section 4 is the summary of this paper.

2. Fast algorithm of intra prediction mode based on CNN

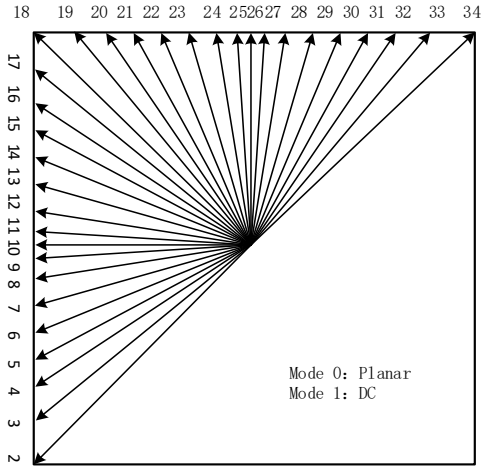

Figure. 1 Luma modes of intra prediction in HEVC

Figure 1 shows 35 different patterns of PU division. In order to see the statistical characteristics of different PU more clearly, we have made statistics on the best pattern distribution of different video blocks. Figure 2 shows the best intra prediction pattern distribution of all PU blocks less than 64x64 in a random frame of parkscene sequence.

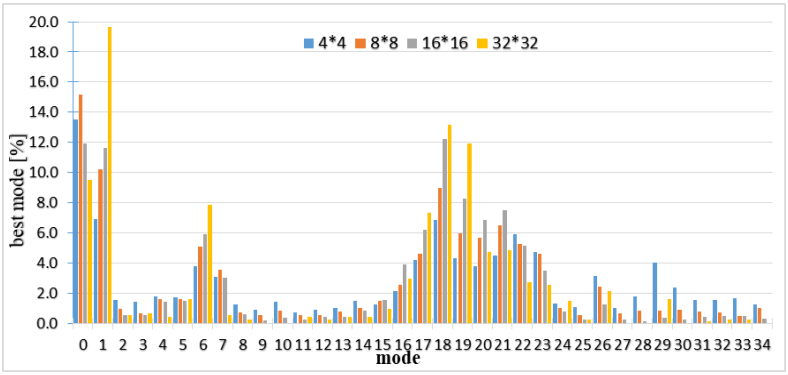

Figure. 2 Overall distribution of the best patterns of PU with different size in basketball drill

ISSN 2616-5667 Vol.3, Issue 2: 73-82, DOI: 10.25236/AJETS.2020.030210

From Figure 2, we can see that DC mode and planar mode account for a large part of the whole. It is predicted that PU blocks have more peaks and more average distribution under 4x4 and 8x8 distribution, while most PU blocks are distributed in the three peaks of 0, 6 and 18 under 16x16 and 32x32 sizes. In the video sequence with complex texture, the smaller the size of the predicted PU in the frame, the less obvious the distribution law. On the contrary, the larger the size, the stronger the distribution law. Therefore, this paper uses convolutional neural network (CNN) to deal with small PU of 4x4 or 8x8, infers the optimal partition mode, and reduces the computational complexity of small PU.

In the original RMD process, PU traverses 35 patterns to obtain several patterns that map the minimum SATD cost to a candidate list. When the PU size is 4x4 or 8x8, its candidate list has 8 patterns, while the larger PU has only 3 candidate lists. Therefore, in this algorithm, for smaller PU, we use CNN to deal with the mode decision-making problem of 4x4 and 8x8pu, and reduce the number of final modes, replacing the original RMD process. The overall process is shown in Figure 3.

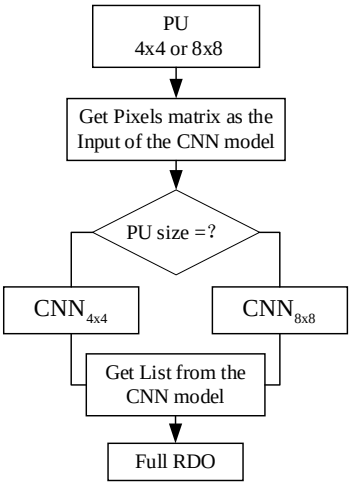

Figure. 3 Flow chart of fast algorithm of intra mode decision based on CNN

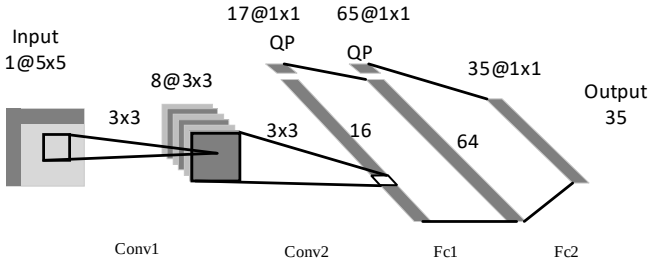

Figure. 4 Model prediction neural network for 4x4 PU

ISSN 2616-5667 Vol.3, Issue 2: 73-82, DOI: 10.25236/AJETS.2020.030210

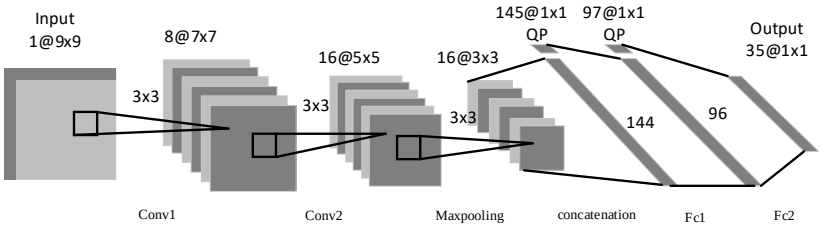

Figure. 5 Model prediction neural network for 8x8 PU

2.1 Prediction model of PU mode based on CNN

In this paper, neural network. In order to adapt to two different sizes of PU, we use two different models, one for 4x4 size PU and the other for 8x8 size PU. Both of them are designed with reference to the classic model of deep learning, lenet-5, but they are simplified to some extent. The overall structure is shown in Figure 4 and Figure 5, which are composed of input layer, two convolution layers and two full connection layers. The functions of these different layers are described as follows:

2.1.1 Input layer

In the process of intra prediction, the selection of PU partition mode is not only related to the current PU pixel value, but also closely related to the reference pixels on the top and on the left. For 4x4 and 8x8 PU blocks, in order to better include all relevant factors in the input, we take the current PU and the adjacent pixels in its top row and left column as the input. For a PU of size 4x4, enter a pixel matrix of 5x5; for a PU of size 8x8, enter a pixel matrix of 9x9. The input is then fed into the two roll up layers conv1 and conv2.

2.1.2 Convolution layer Conv1, Conv2

After two convolution layers, texture changes in PU pixels are extracted. Eight and 16 3 x 3 convolution kernels are used to convolute the size of 3 x 3 region and extract local features, which is convenient for the later full connection layer to reassemble the previous local features into complete features through the weight matrix.

2.1.3 Max-Pooling, Concatenation layer (only on 8x8 networks)

In the 8 x 8 PU model prediction neural network, after two layers, the size of the feature map is 5 x 5 and the number of channels is 16. In order to further simplify the features, keep the main features and reduce the parameters, a max pooling process is carried out. Here, the configuration padding = 'Save' is used, the size of

ISSN 2616-5667 Vol.3, Issue 2: 73-82, DOI: 10.25236/AJETS.2020.030210

the pooling core is 3 x 3, and the size of the feature map after the final pooling is 3 x 3. In order to facilitate the operation of the later full connection layer, the feature map is straightened into one dimension through the convergence layer.

2.1.4 Full connection layer FC1, FC2

Next, the two fully connected layers reassemble the previously extracted features into a complete feature map through the weight matrix, and further obtain the final output features. In addition, because the quantization parameters also have some influence on the selection of modes, the quantization parameter QP (quantification parameter) is also added to the two-layer full connection layer as an external input here.

2.1.5 Output layer

Output of output layer has 35 output values, corresponding to 35 prediction modes.

2.1.6 Training objective, loss function, activation function and training data

For the selection of training target, the method of reference [11] in this paper uses the target value based on the rate distortion cost (RDC), rather than selecting the mode as the classification as the traditional method. The reason is that it is almost impossible for neural network to get the meaning of PU mode by training with prediction mode as label, and RDC is the determination standard of PU mode decision-making. It is a better choice to use RDC related values as training label. Therefore, the text models the prediction problem as a regression problem rather than a classification problem. The goal is to predict the RDC correlation value of each pattern.

But at the same time, because the actual RDC value of each mode of HEVC is too large, which may lead to network saturation, we define the target value from the following steps:

First, get the RDC of 35 modes of the current PU. In this step, we traverse all patterns through RDO process instead of RMD process to obtain RDC of each pattern, a total of 35, which is represented by RDC_{m}, where M represents pattern.

Then calculate the average value of RDC:

$$\overline{\mathrm{C}}=\frac{1}{35}\sum_{\mathrm{m}=0}^{34}RDC_{\mathrm{m}} \tag{1}$$

Define the target value y_m as follows:

ISSN 2616-5667 Vol.3, Issue 2: 73-82, DOI: 10.25236/AJETS.2020.030210

$$y_m = \begin{cases} \ln(-RDC_{\mathrm{m}} + \overline{\mathrm{C}} + 1), RDC_{\mathrm{m}} < \overline{\mathrm{C}} \\ -\ln(RDC_{\mathrm{m}} - \overline{\mathrm{C}} + 1), RDC_{\mathrm{m}} \geq \overline{\mathrm{C}} \end{cases} \tag{2}$$

After the treatment of equation (2), when RDC_{m} is larger, the value of y_m is smaller; when RDC_{m} is smaller, the value of y_m is larger; and the closer RDC_{m} is to the average value $\overline{\mathrm{C}}$, the closer y_m is to 0. Therefore, the training goal of this paper is y_m. the larger the value of y_m is, the smaller the RDC is, the more likely the prediction model will become the best one. In the actual application of the model, the predicted values $\hat{y}_m$ are sorted, and the maximum four modes are selected as the final list of prediction modes, which are sent to the RDO process.

In the selection of loss function, this paper uses the loss function MSE commonly used in regression problems, which is defined as follows:

$$Loss = \frac{1}{35}\sum_{m=0}^{34}\left(y_m - \hat{y}_m\right)^2 \tag{3}$$

Where y_m is the true value and $\hat{y}_m$ is the predicted value.

The activation function uses the ReLU function, which converges faster than the sigmoid function, and there is no problem of gradient disappearing. The random gradient descent method is used to update the weights.

The training data in this paper comes from 12 video sequences, including PeopleOnStreet. RDO is carried out on all PU of 4x4 and 8x8 sizes. The real target values are collected as sample tags and encoded under four quantization parameters, QP = {22, 27, 32, 37}.

3. Analysis of experimental results

In order to verify the effectiveness of the proposed deep learning model based on the perception module, this section statistics and analyzes the experimental results of the proposed algorithm. The algorithm is embedded in HEVC reference software HM16.12 version [3] for experiments. The hardware configuration of the experimental platform is an intel core i7-8750h CPU with a main frequency of 3.20ghz, a Windows 10 64 bit operating system with a running memory of 8G, and a tensorflow1.14.0 deep learning framework. In this paper, jctvc-l1100 [13] test conditions are used to select the first 50 frames of 10 coding standard test sequences of five categories A to E for coding test. The QP value of the quantization parameter is set to 22, 27, 32, 37 respectively. The encoding configuration uses the all intra main (AI main).

ISSN 2616-5667 Vol.3, Issue 2: 73-82, DOI: 10.25236/AJETS.2020.030210

Statistical analysis of coding time and coding rate, test results and literature [5], literature [7], literature [8] and literature [11] proposed algorithm experimental results are compared, the evaluation method is BD-BR method proposed in the proposal VCEG-M33 by Bjøntegarrd [14]. Bd-BR represents the difference between coding rates under the same PSNR. The time saving of coding is calculated by Δ T, and the calculation formula is as follows:

$$\Delta T = \frac{1}{4}\sum \frac{\mathrm{Time}_{HM16.12}(QP) - Time_{prop}(QP)}{Time_{HM16.12}(QP)} \times 100\%$$

Time_{prop} represents the encoding time of this algorithm, $\mathrm{Time}_{HM16.12}$ represents the original encoding time of HM16.12, QP = {22, 27, 32, 37}.

Experimental results in Table 1 show that: compared with HEVC standard algorithm, the coding time of this algorithm is saved by 28.08% on average, while BD-BR is only increased by 1.14%. Under the condition that the increase of bit rate can be ignored, the coding time is greatly reduced. Compared with the algorithms in literature [5] and [7], the algorithm in this paper is obviously superior to the algorithm in literature [5], because it can save more time when the coding performance is improved and BD-BR is reduced. In addition, compared with algorithms [8] and [11], when BD-BR has little difference, the coding time is reduced more, which is suitable for real-time transmission with more stringent requirements on coding time.

Table 1 The performance comparison between the algorithm in this paper and the original encoder

Classes	Sequences	BD-BR/%	BD-PSNR/dB	ΔT/%
A	Traffic	0.79	-0.026	31.43
2560×1600	PeopleOnStreet	0.81	-0.029	28.35
B	BasketballDrive	0.67	-0.041	24.97
1920×1080	ParkScene	0.74	-0.064	28.45
C	BasketballDrill	1.25	-0.045	31.67
832×480	BQMall	1.06	-0.063	30.25
D	BasketballPass	1.54	-0.072	25.34
416×240	RaceHorses	1.48	-0.072	28.9
E	FourPeople	1.45	-0.059	26.16
1280×720	Johnny	1.57	-0.075	25.32
Average		1.14	-0.055	28.08

Table 2 Performance comparisons between the proposed algorithm and the existing algorithms

Algorithm	BD-BR/%	BD-PSNR/dB	ΔT/%
Ref [5]	1.42	-0.040	20.5
Ref [7]	1.46	-0.061	25.74
Ref [8]	1.37	-0.045	18.89
Ref [11]	1.15	-0.067	27.92
Proposed	1.14	-0.055	28.08

ISSN 2616-5667 Vol.3, Issue 2: 73-82, DOI: 10.25236/AJETS.2020.030210

4. Conclusion and discussion

In this paper, convolutional neural network (CNN) method is used to predict the process of intra prediction with high complexity in HEVC. For the size of 4x4 or 8x8 PU, CNN based deep learning model is used to predict the mode of entering RDO. Experimental results show that compared with the official HEVC test platform HM16.12, the algorithm in this paper can significantly improve the coding speed with little change in the coding performance. The disadvantage is that the inter frame coding can not be considered. In the future, the inter prediction processing can be increased to reduce the complexity of inter frame coding.

References

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

[2] Ohm J R, Sullivan G J, Schwarz H, et al. Comparison of the coding efficiency of video coding standards-including high efficiency video coding (HEVC) [J].IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22 (12): 1669-1684.

[3] JCT-VC, “HM Software”. [CP/OL]. Available: https:// hevc. hhi.fraunhofer. de/ svn/ svn_HE-VCSoftware/tags/HM-16.12/.

[4] Lainema J, Bosssen F, Han W J, et al. Intra Coding of the HEVC standard [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22 (12): 1792-1801.

[5] Jiang W, Ma H J, Chen Y W. Gradient based fast mode decision algorithm for intra prediction in HEVC [C]// 2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), Yichang, 2012: 1836-1840.

[6] Liao W H, Yang D Q, Chen Z Z. A Fast Mode Decision Algorithm for HEVC Intra Prediction [C]// 2016 Visual Communications and Image Processing (VCIP), Chengdu, 2016: 1-4.

[7] G. Chen, Z. Liu, T. Ikenaga and D. Wang, "Fast HEVC intra mode decision using matching edge detector and kernel density estimation alike histogram generation,"2013 IEEE International Symposium on Circuits and Systems (ISCAS), Beijing, 2013, pp. 53-56.

[8] T. L. da Silva, L. V. Agostini and L. A. da Silva Cruz, "Fast HEVC intra prediction mode decision based on EDGE direction information," 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO), Bucharest, 2012, pp. 1214-1218.

[9] Yi Q, Xie Z, Shi M. A fast decision-making algorithm for HEVC intra coding [J]. Journal of Chinese Computer Systems, 2019, 40 (01): 199-204. http:// kns.cnki.net/kcms/detail/detail.aspx?FileName=XXWX201901039&DbName=CJFQ2019

ISSN 2616-5667 Vol.3, Issue 2: 73-82, DOI: 10.25236/AJETS.2020.030210

[10] A. Heindel, C. Pylinski and A. Kaup, "Two-stage exclusion of angular intra prediction modes for fast mode decision in HEVC," 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, 2016, pp. 529-533.
[11] N. Song, Z. Liu, X. Ji and D. Wang, "CNN oriented fast PU mode decision for HEVC hardwired intra encoder," 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Montreal, QC, 2017, pp. 239-243.
[12] Zhu S P, Zhang C Y. A Fast Algorithm of Intra Prediction Modes Pruning for HEVC Based on Decision Trees and A New Three-Step Search [J]. Multimedia Tools and Applications, Springer, 2017, 76 (20): 21707-21728.
[13] Bossen F, Common test conditions and software reference configurations, JCTVC-L1100, in: 12TH JCT-VC meeting, Geneva, CH, January 2013: 1-4.
[14] Bjøntegaard G. Calculation of Average PSNR Differences between R-D curves [J]. ITU-T VCEG, 13th Meeting, 2001.