

# An Effective CU Depth Decision Method for HEVC Using Machine Learning

Xuan Sun<sup>1,2,3</sup>, Pengyu Liu<sup>1,2,3,\*</sup>, Xiaowei Jia<sup>4</sup>, Kebin Jia<sup>1,2,3</sup>, Shanji Chen<sup>5</sup> and Yueying Wu<sup>1,2,3</sup>

<sup>1</sup>Beijing University of Technology, Beijing, 100124, China

<sup>2</sup>Beijing Laboratory of Advanced Information Networks, Beijing, 100124, China

<sup>3</sup>Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing, 100124, China

<sup>4</sup>Department of Computer Science, University of Pittsburgh, Pittsburgh, 15260, USA

<sup>5</sup>School of Physics and Electronic Information Engineering, Qinghai Nationalities University, 810007, Qinghai, China

\*Corresponding Author: Pengyu Liu. Email: liupengyu@bjut.edu.cn

Received: 12 November 2020; Accepted: 14 December 2020

**Abstract:** This paper presents an effective machine learning-based depth selection algorithm for CTU (Coding Tree Unit) in HEVC (High Efficiency Video Coding). Existing machine learning methods are limited in their ability in handling the initial depth decision of CU (Coding Unit) and selecting the proper set of input features for the depth selection model. In this paper, we first propose a new classification approach for the initial division depth prediction. In particular, we study the correlation of the texture complexity, QPs (quantization parameters) and the depth decision of the CUs to forecast the original partition depth of the current CUs. Secondly, we further aim to determine the input features of the classifier by analysing the correlation between depth decision of the CUs, picture distortion and the bit-rate. Using the found relationships, we also study a decision method for the end partition depth of the current CUs using bit-rate and picture distortion as input. Finally, we formulate the depth division of the CUs as a binary classification problem and use the nearest neighbor classifier to conduct classification. Our proposed method can significantly improve the efficiency of inter-frame coding by circumventing the traversing cost of the division depth. It shows that the mentioned method can reduce the time spent by 34.56% compared to HM-16.9 while keeping the partition depth of the CUs correct.

**Keywords:** HEVC; inter-frame coding; CU depth

## 1 Introduction

Inter-frame prediction is an important part in the encoder process of HEVC. By using the time correlation between continuous video frames [1], temporal redundancy can be available removed. The objective of inter-frame prediction is to get the image which is composed of the neighboring encode images. Through the comparison of the rate-distortion cost between the CU and the total lower four sub-CU, HEVC keeps checking whether the nowadays CU (Coding Unit) needs to be partitioned from the maximal CU ( $64 \times 64$ ) to the minimal CU ( $8 \times 8$ ).



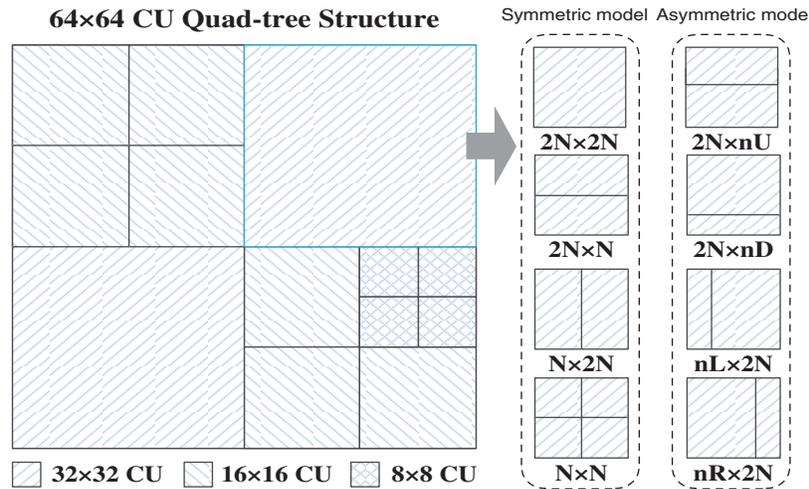
This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The mathematical equation of the rate-distortion value  $J_{mode}$  can be described as:

$$D = SSE_{luma} + W_{chroma} \cdot SSE_{chroma}, \quad (1)$$

$$J_{mode} = D + \lambda \cdot R_{mode}, \quad (2)$$

where  $SSE_{luma}$  refers to the sum of the brightness difference square between the initial frame and the coding frame; and  $SSE_{chroma}$  is the sum of the chroma difference square; the parameter  $W_{chroma}$  is the weighted factor of Chroma distortion; the parameter  $D$  represents the total distortion;  $\lambda$  is the Lagrange multiplier;  $R_{mode}$  corresponds to the coded bits of the mode. In addition, every CU can consist of many PUs (Prediction Units) while making predictions [2]. As shown in Fig. 1, HEVC defines eight different types of PU.



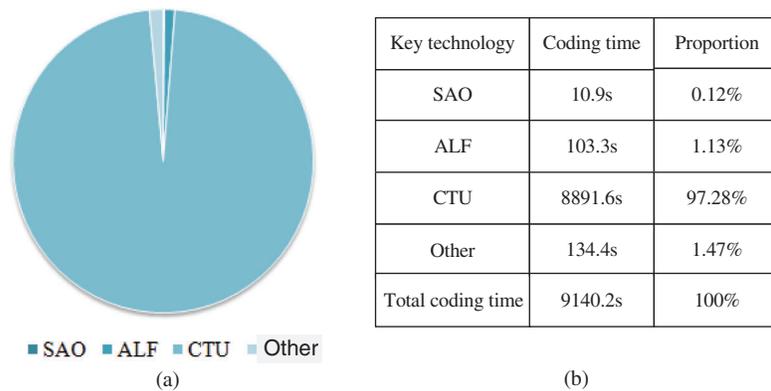
**Figure 1:** Correlation between PU and CU

To make the optimal PU decision with the lowest distortion rate, HEVC encoder needs to exhaustively check all the possible PU partitions. When the encoder completes the motion predictions for all the depths of CU, it finds the best PU partitioning and merges the CU motion vectors in a bottom-up order. The computational volume of the rate-distortion for each CTU is very great [3–6].

In addition, it shows that the BasketballDrill sequence (QP = 22) are encoded by the HM-16.9 test model in the LD (low delay) configuration. We show the coding time proportion of each part in HEVC in Fig. 2. It can be seen that the coding time proportion of CTU accounts is 97.28% of the total coding time, while the remaining operations only accounted for 2.72% of the calculation time. Complicated CU dividing is very beneficial to the efficiency of HEVC, but it also takes a lot of computation time which greatly increases the coding time of HEVC. Therefore, it is of great importance to optimize the depth partition of CUs so as to reduce the computational complexity of the HEVC.

## 2 Reason Analysis of Proposed Method

According to the different judgment methods, the effective depth decision methods of the inter-frame CU can be divided into two categories: threshold comparison-based fast depth selection algorithm and machine learning-based fast depth selection algorithm. Threshold comparison is the most common method used to determine whether the current coding unit needs to further divided. For example, [7] proposed an effective depth decision method for inter-frame CU by comparing the texture complexity and motion complexity with a threshold value. However, its effectiveness also depends on the choice of the threshold, and thus have robustness issues.

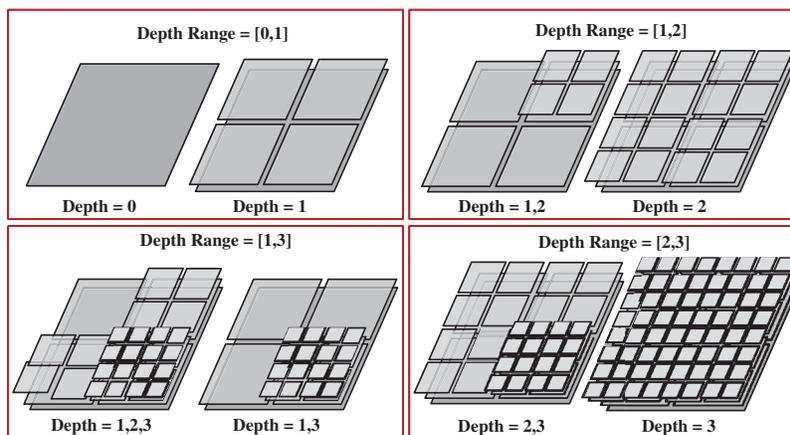


**Figure 2:** The coding time proportion of each part in HEVC

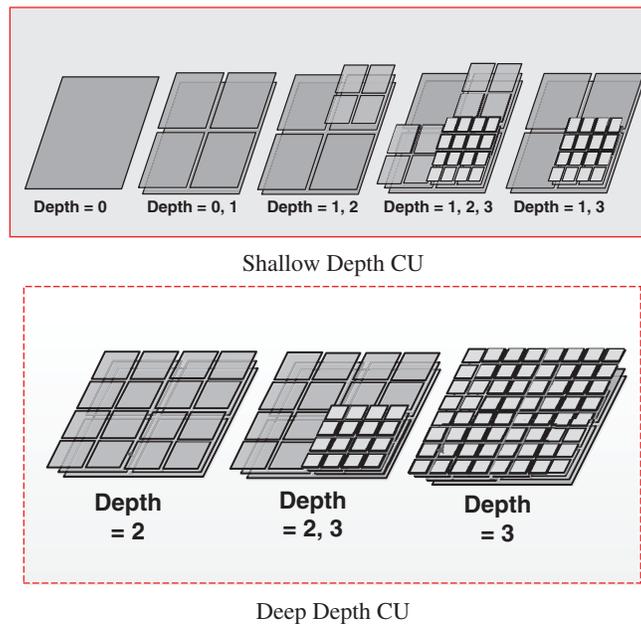
Therefore, the scholars take advantage of machine learning techniques to effectually decrease the calculational time of HEVC. The process of dividing CUs is regarded as a classified problem using classical classification algorithms in machine learning. HEVC can automatically estimate the division depth of CU without any traversal calculation. For example, [8] proposed an effective depth decision method for CUs based on Bayesian classifier, whose classification features were composed of HEVC intermediate coding information, and Bayers classifier was used to estimate whether to stop the ergodic process of the CUs. However, the traversal calculation still has to begin from depth 0 though the depth of the coding video is 2 or 3. We can directly estimate the beginning depth of the CU division and greatly reduce the coding time if we use the "top skip" method in the depth selection method. At the same time, what features are selected as input to the classifier has a great impact on the division accuracy [9–11].

**2.1 Effect of Frame Texture complexity, Coding Quantization Parameters on Depth Division of CUs**

Fig. 3 indicates that there are four types of depth division in CTU. And CU can be specifically divided into two types: shallow depth CU and deep depth CU. Shallow depth CU refers to the CU with initial division depth of 0 and 1. Deep depth CU refers to the CU with an initial division depth of 2. Shallow depth CU and deep depth CU are shown in Fig. 4.



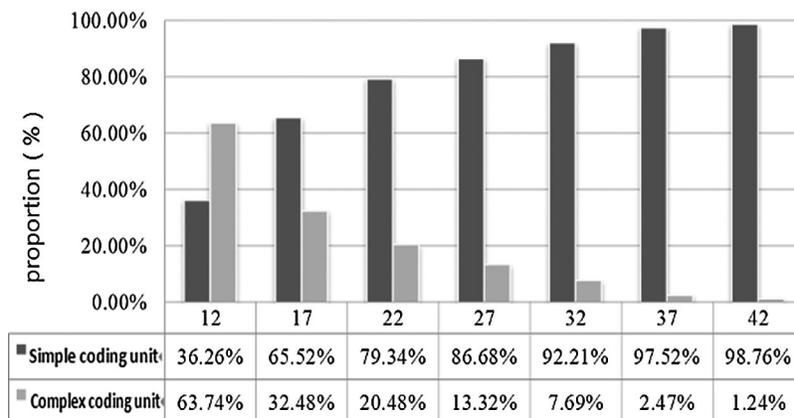
**Figure 3:** Four depth division types



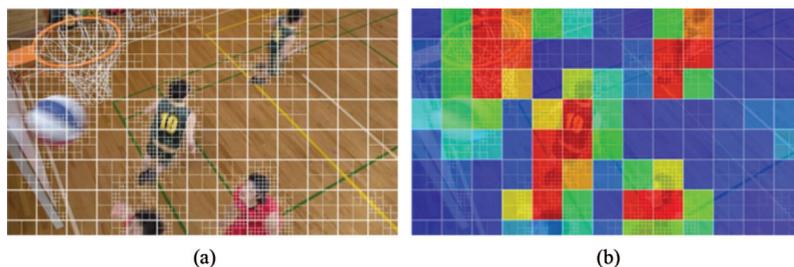
**Figure 4:** Demonstration of shallow and deep depth CU (a) Shallow Depth CU (b) Deep Depth CU

As shown in Fig. 6a, the areas with flat content contain more low-frequency components, and the areas with complex textures contain more high-frequency components. In order to achieve higher compression efficiency, larger-size CU will be selected to encode areas with more low-frequency components during the video compression. For regions with large high-frequency components, smaller-size CU blocks are usually selected for encoding in order to preserve more details of the decoded video. It shows that texture complexity has a great impact on the depth division of CU. Therefore, texture complexity can be utilized as one of the effective inputs to decide the original partition depth of CU [12].

As shown in Fig. 5, when the test video sequence is coded with different QPs in the same coding configuration, it can be observed that the ratio of deep depth CU reduces with the increase of QP. As a result, the QPs can be utilized to decide whether the current CU is a shallow depth CU. On the other hand, under the same compression configuration, the larger the QP setting is, the rougher the compression is, so there are relatively more shallow depth CUs. The smaller the QP setting is, the finer the compression is, so there are relatively more deep depth CUs [13]. All these showed that we can use texture complexity and QPs as the inputs of depth division method.



**Figure 5:** Shallow and deep depth CU with different QPs



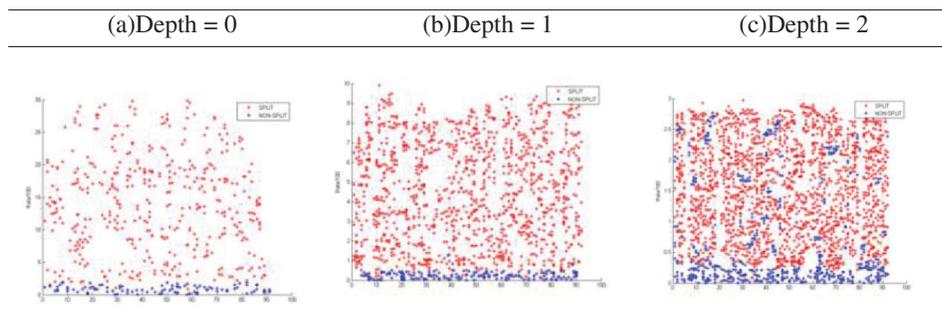
**Figure 6:** CU partition and corresponding heatmap (a) CU Partition Results (b) Heatmap

**2.2 Effect of Code Rate, Distortion on Depth Division of CUs**

Selecting a proper set of features as inputs greatly affect the classifier since it can potentially reduce the training and prediction time while also improving the accuracy of depth selection. In order to maintain the efficient coding efficiency of HEVC, coding rate and distortion can be used as reference conditions to select the correct coding method. Therefore, we will explore the influence of bit rate and distortion on the CU depth division and the validity of utilizing them as input of the classifier.

As can be seen from Fig. 6, CUs with deeper depth cost higher bit rate. Otherwise, CUs with shallower depth cost lower rate. As shown in Tab. 1, on the basis of video sequence BasketballDrill, the relationship of the CU depth and the code rate under the LD configuration is displayed. It can be found that the higher the bit rate cost, the greater the possibility of CU division. Moreover, the line between “Divide” and “Non-divide” is much clearer. Therefore, we can treat bit rate as the need information of the classifier.

**Table 1:** Correlation between CU division depth and bit rate

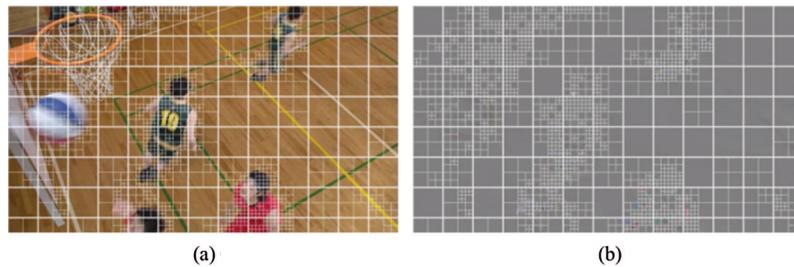


As can be seen from Fig. 7, the CU partition frame and distortion frame of the video sequence BasketballDrill, in which the distorted frame is obtained by the coding frame and the initial one. The encoder will generally choose deep coding in the area of intense movement of the object, or in the area of more detail. On the contrary, shallow coding in these regions may lead to lower accuracy of motion prediction, thus increasing coding distortion.

As shown in Tab. 2, the correlation of the distortion and the depth of the CU under the LD configuration is displayed. Hence, we can treat distortion as the need information of the classifier.

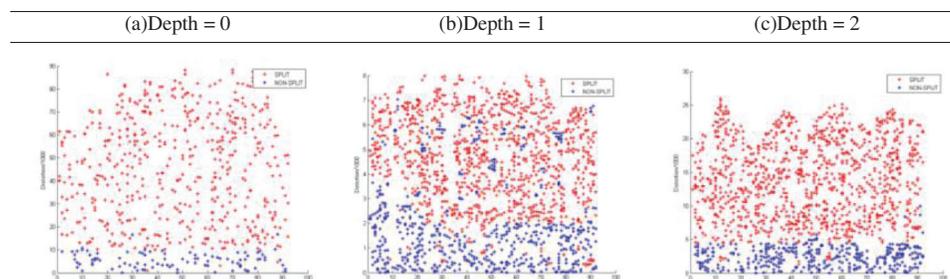
The NN (nearest neighbor) [14] method is extensively applied in the field of machine learning. The core idea of neural network algorithm is “close to win”, that is, the classification type of the closest sample determines the classification result of each sample. If the nearest neighbor of the target sample in its eigenspace belongs to a specific category, the target sample should also belong to the same category and

has the basic characteristics of the samples in that category. Up to now, the nearest neighbor classifier, as a supervised classification method, has been widely and successfully applied in text classification, image classification, face recognition, pattern recognition, and other fields.



**Figure 7:** CU partition and distortion map (a) CU Partition Results (b) Distortion map

**Table 2:** Correlation between CU depth and coding distortion



The NN classifier will be used to decide the depth division of the CU. The used video sequences contain rich content and scenes at different resolutions, that is, they not only contain a large number of shallow depth division CU in the region with simple texture and flat content, but also contain deep depth division CU in the region with intricate pattern or violent motion. [Tab. 3](#) indicates that the selected video sequences conform to the principles of accuracy, representativeness, and statistics.

### 3 An Effective Depth Prediction Method for CU

According to the idea of “top skip” and “early terminate”, this paper proposes a machine learning-based effective depth decision method for CU with two important part. The first part is forecasting the original partition depth of the CU, which primarily uses the texture complexity and QPs to predict. The second part is determining the end division depth of the CU, which primarily decides whether to terminate the depth division early.

#### 3.1 A. Initial Partition Depth Prediction of CU

Here we first introduce the “top skip” mechanism used in our fast depth selection algorithm. By designing a method to use texture complexity and QPs to predict the initial partition CU depth, the problem of time waste caused by uniformly dividing all CU from depth 0 is improved. The following is the detailed process:

**Table 3:** The features of each using sequences

Sequence	Class	Resulation	Diagram	Characteristics
BQTerrace	B	1920x1080		slight movement complex textures flat areas
BasketballDrill	C	832x480		vigorous movement complex textures flat areas
BQSquare	D	416x240		movement complex textures flat areas
FourPeople	E	1280x720		background complex textures flat areas

Four test sequences, BQTerrace, BasketballDrill, BQSquare, and FourPeople, were selected with different QPs and entropy values. The quantitative parameter is the X-axis. The Y-axis expresses the entropy, which is in the range of 0.5~7.5. It is a very practical method to distinguish shallow and deep depth CUs simply and effectively by constructing a prediction dictionary.

**Design of the original depth division forecast dictionary:** As shown in Fig. 8, the red circle ( $\odot$ ) means that under the setting of the QP and entropy, there are more deep depth CUs in the test results, so the CU is considered to be a deep depth CU. The blue mark ( $\times$ ) means that under the setting of the QP and entropy, there are more shallow depth CUs in the test results, so this CU is considered to be a shallow coding unit.

**CU type decision:** According to the QPs and the entropy, determine whether the current CU is a deep depth CU by the dictionary;

**Depth division of CU:** The quadtree traverses from depth 0 if the current Cu is a shallow depth CU. Otherwise, the quadtree traverses from depth 2.

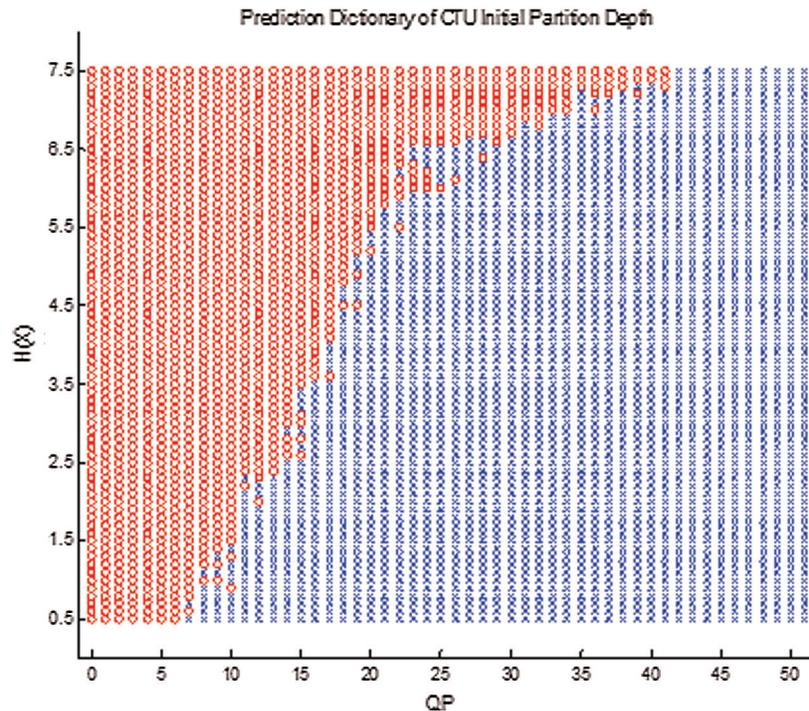
### 3.2 Terminational Partition Depth Prediction of CU

In order to forecast whether the current CU needs to be divided, this paper feeds the NN classifier with the rate and distortion as the needed information. Here we introduce the details of different components involved in this process.

**Training the NN classifier:** We train an off-line NN classifier which uses rate and distortion to judge whether the current CU needs to be divided.

**CU division judgement:** The distortion and bit rate are regarded as the needed information of the NN classifier to get the classified result.

**CU final depth:** Based on the classified result, determine whether the current CU needs to be divided.



**Figure 8:** Forecast dictionary of CU initial division depth

### 3.3 Overall Framework

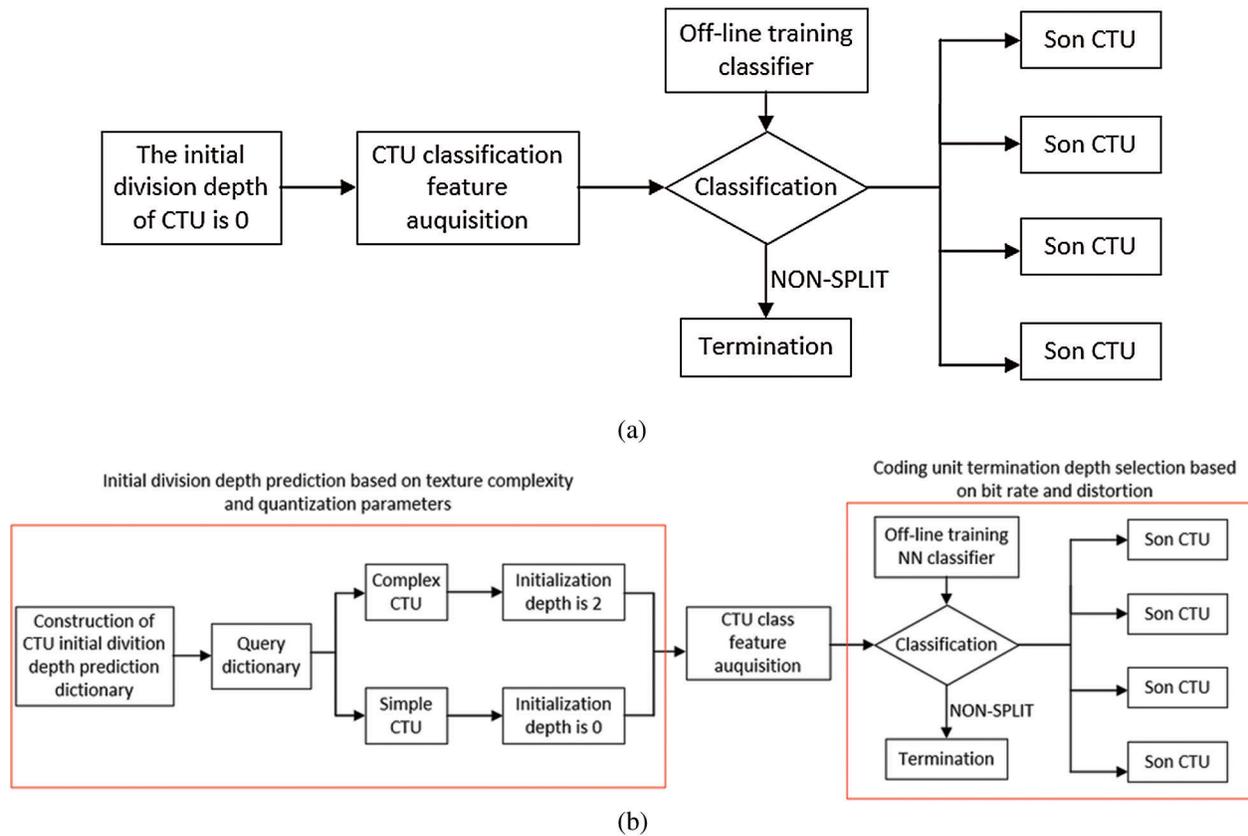
Fig. 9b shows the framework process of the overall method. In the following, we introduce the effective depth decision method for CU in detail.

First, we train the original dividing depth forecast dictionary by four sequences. Second, the correct type of CTU is conducted by looking up a dictionary based on the QPs and the minimum entropy value. Third, the quadtree traverses from depth 0 if the current CU is a shallow depth CU. Otherwise, the quadtree traverses from depth 2. Fourth, we train the NN classifier through the code rate, distortion, and division result of abundant types of CU. Fifth, using distortion and bit rate as the needed information to output the classified division result. Sixth, continue to divide the current CU and repeat last step until the largest depth is reached. Otherwise, do not continue dividing.

## 4 Experimental Part

Video sequences with rich content and scenes were used for experiments to test the performance of the proposed depth determination method. Fig. 10 shows the details of the video sequences used.

This experiment utilizes BDBR (Bjontegaard Delta Bit-rate) and BDPSNR (Bjontegaard Delta Peak Signal-to-Noise Rate) to express the coding ability of the proposed method. As shown in Fig. 11, the proposed method can decrease the coding time by 34.56% on average comparing with the AI (All Intra) configuration of HM-16.9. Meanwhile, the bit rate is only increased by 1.21%~2.29%, and the coding quality is only lost by 0.08 dB~0.26 dB. Our goal is to reduce the HEVC coding complexity without compromising the quality of image reconstruction.



**Figure 9:** The overall process: (a) Initial flow (b) Modified flow

Class	Resolution	Sequence	Frame Rate
B	1920x1080	BQTerrace, BasketballDrive	60/50fps
C	832x480	BQMall, BasketballDrill	60/50fps
D	416x240	BQSquare, BasketballPass	60/50fps
E	1280x720	FourPeople, Johnny	60/60fps

**Figure 10:** Video sequences

This paper compares the ability of the proposed method with the fast coding method [15], as shown in Figs. 12 and 13. The proposed method can decrease the coding time by 4.11% compared with [15] while ensuring good video reconstruction quality.

To prove that the termination depth partition prediction method is effective, we tested the classification accuracy when the depth is 0, 1 and 2, respectively. As shown in Fig. 14, the average classification accuracy is close to 92%, which indicates that the proposed method can predict the final division of coding units quite accurately.

Class	Resolution	Sequence	Proposed		
			BR (%)	PSNR (dB)	$\Delta T(\%)$
Class B	1920x1080	BQTTerrace	2.10	-0.20	36.57
		BasketballDrive	1.78	-0.26	28.64
Class C	832x480	BQMall	1.58	-0.10	32.56
		BasketballDrill	1.62	-0.18	29.28
Class D	416x240	BQSquare	1.21	-0.11	32.65
		BasketballPass	2.29	-0.15	28.22
Class E	1280x720	FourPeople	1.35	-0.08	42.78
		Johnny	1.46	-0.08	45.82
Average			<b>1.67</b>	<b>-0.15</b>	<b>34.56</b>

Figure 11: Ability of proposed method (AI) under AI

Class	Resolution	Sequence	Proposed		
			BR (%)	PSNR (dB)	$\Delta T(\%)$
Class B	1920x1080	BQTTerrace	0.08	0.02	2.92
		BasketballDrive	0.05	-0.06	5.91
Class C	832x480	BQMall	-0.07	0.04	3.43
		BasketballDrill	0.04	0.02	6.93
Class D	416x240	BQSquare	0.25	0.01	2.42
		BasketballPass	0.52	-0.01	5.82
Class E	1280x720	FourPeople	-0.23	-0.02	3.16
		Johnny	0.04	0.03	2.29
Average			<b>0.08</b>	<b>0.00</b>	<b>4.11</b>

Figure 12: Ability of proposed method (LD) [15]

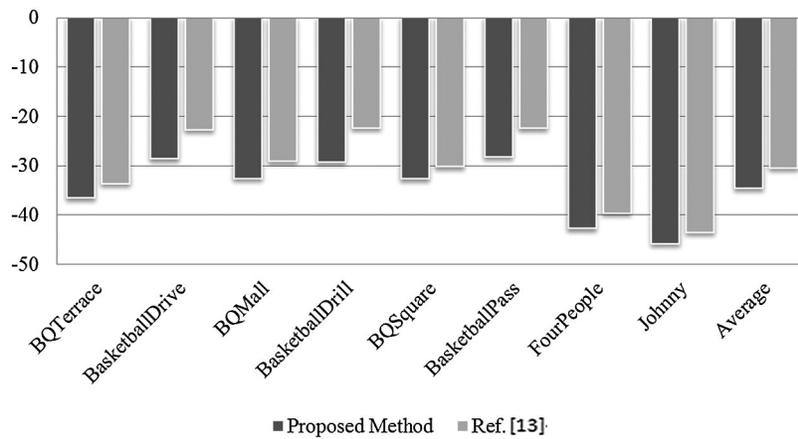


Figure 13: Coding time reducing of proposed method (LD) [15]

Class	Resolution	Accuracy of CU Partition under Different Depths			
		Depth=0	Depth=1	Depth=2	Average
Class B	1920x1080	93.32	89.95	87.75	<b>90.34</b>
Class C	832x480	92.76	90.57	90.04	<b>91.12</b>
Class D	416x240	93.80	91.17	87.58	<b>90.85</b>
Class E	1280x720	96.02	95.58	95.10	<b>95.56</b>
<b>Average</b>		<b>93.98</b>	<b>91.82</b>	<b>90.12</b>	<b>91.97</b>

**Figure 14:** Classification accuracy of different depths

## 5 Conclusions

This paper proposes an effective CU depth decision method which combines the thought of top skip and early termination, by solving the intricate problem result from the deep traversal of the CU in HEVC. Firstly, based on texture complexity and QPs, a scheme for forecasting the original segmentation depth of CU is designed. According to the idea of top skip, this method can help deep depth CUs reduce the rate-distortion cost calculation with depths of 0 and 1, thus saving coding time. Secondly, an end depth decision method based on the neural network classifier is designed. This method is proposed to model the division problem as two classification problems and releases early termination. It reduces the calculation cost for shallow depth CUs with depth of 2 or 3, and reduces the complexity of encoder. It shows that the proposed method saves an average of 34.56% coding time compared with the original HEVC encoder, and achieves a balance between encoding video quality and encoding bit rate cost.

**Funding Statement:** This paper is supported by the National Natural Science Foundation of China (61672064), Basic Research Program of Qinghai Province (No. 2020-ZJ-709), and the project for advanced information network Beijing laboratory (PXM2019\_014204\_500029).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] G. J. Sullivan, J. R. Ohm, W. J. Han and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [2] K. Misra, A. Segall, M. Horowitz and S. Xu, "An overview of tiles in HEVC," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 969–977, 2013.
- [3] K. Kim and W. W. Ro, "Fast CU depth decision for HEVC using neural networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 5, pp. 1462–1473, 2019.
- [4] W. Zhu, Y. Yi, H. Zhang, P. Chen and H. Zhang, "Fast mode decision algorithm for HEVC intra coding based on texture partition and direction," *Journal of Real-Time Image Processing*, vol. 17, no. 2, pp. 275–292, 2020.
- [5] Y. Gao, P. Y. Liu, Y. Y. Wu and K. B. Jia, "Quadtree degeneration for HEVC," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2321–2330, 2016.
- [6] D. Y. Wang, Y. Sun, C. Zhu, W. S. Li and F. Dufaux, "Fast depth and inter mode prediction for quality scalable high efficiency video coding," *IEEE Transactions on Multimedia*, vol. 22, no. 4, pp. 833–845, 2020.
- [7] S. Ahn, B. Lee and M. Kim, "A novel fast CU encoding scheme based on spatiotemporal encoding parameters for HEVC inter coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 422–435, 2015.
- [8] J. Lee, S. Kim, K. Lim and S. Lee, "A fast CU size decision algorithm for HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 411–421, 2015.

- [9] K. Duan, P. Y. Liu, Z. Q. Feng and K. B. Jia, "Fast PU intra mode decision in intra HEVC coding," in *Proc. Data Compression Conf. (DCC)*, Snowbird, UT, USA, pp. 570, 2019.
- [10] B. J. Hyun and S. M. Hoon, "Adaptive early termination algorithm using coding unit depth history in HEVC," *Journal of Signal Processing Systems*, vol. 91, no. 8, pp. 863–873, 2019.
- [11] Y. Gao, P. Y. Liu, Y. Y. Wu and K. B. Jia, "HEVC fast CU encoding based quadtree prediction," in *Proc. Data Compression Conf. (DCC)*, Snowbird, UT, USA, pp. 594, 2016.
- [12] T. H. Tsai, S. S. Su and T. Y. Lee, "Fast mode decision method based on edge feature for HEVC inter-prediction," *IET Image Processing*, vol. 12, no. 5, pp. 644–651, 2018.
- [13] Y. T. Kuo, P. Y. Chen and H. C. Lin, "A spatiotemporal content-based CU size decision algorithm for HEVC," *IEEE Transactions on Broadcasting*, vol. 66, no. 1, pp. 100–112, 2020.
- [14] B. V. Dasarathy, "Nearest neighbor (NN) norms: NN pattern classification techniques," *Los Alamitos IEEE Computer Society Press*, vol. 13, no. 100, pp. 21–27, 1990.
- [15] F. S. Mu, L. Song, X. K. Yang and Z. Y. Luo, "Fast coding unit depth decision for HEVC," in *Proc. IEEE Int. Conf. on Multimedia and Expo Workshops*, Chengdu, China, pp. 1–6, 2014.