


Article

# Fake Bitrate Detection of HEVC Videos Based on Prediction Process

Xiaoyun Liang <sup>1</sup>, Zhaohong Li <sup>1,2,\*</sup>, Zhonghao Li <sup>1</sup> and Zhenzhen Zhang <sup>3</sup>

<sup>1</sup> School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup> Guangdong Provincial Key Laboratory of Information Security Technology, Guangzhou 510000, China

<sup>3</sup> School of Information Engineering, Beijing Institute of Graphic Communication, Beijing 100029, China

\* Correspondence: zhli2@bjtu.edu.cn

Received: 29 June 2019; Accepted: 11 July 2019; Published: 15 July 2019



**Abstract:** In order to defraud click-through rate, some merchants recompress the low-bitrate video to a high-bitrate one without improving the video quality. This behavior deceives viewers and wastes network resources. Therefore, a stable algorithm that detects fake bitrate videos is urgently needed. High-Efficiency Video Coding (HEVC) is a worldwide popular video coding standard. Hence, in this paper, a robust algorithm is proposed to detect HEVC fake bitrate videos. Firstly, five effective feature sets are extracted from the prediction process of HEVC, including Coding Unit of I-picture/P-picture partitioning modes, Prediction Unit of I-picture/P-picture partitioning modes, Intra Prediction Modes of I-picture. Secondly, feature concatenation is adopted to enhance the expressiveness and improve the effectiveness of the features. Finally, five single feature sets and three concatenate feature sets are separately sent to the support vector machine for modeling and testing. The performance of the proposed algorithm is compared with state-of-the-art algorithms on HEVC videos of various resolutions and fake bitrates. The results show that the proposed algorithm can not only can better detect HEVC fake bitrate videos, but also has strong robustness against frame deletion, copy-paste, and shifted Group of Picture structure attacks.

**Keywords:** video forensics; HEVC; fake bitrate; prediction process

## 1. Introduction

Digital video has become an indispensable part of our daily lives nowadays. According to the statistics, about 65,000 videos are uploaded to YOUTUBE every day [1]. However, falsified videos have been exposed to sneak in [2], which will cause serious moral, ethical, and legal problems. Therefore, it is essential to verify the authenticity of videos before they can be trusted. One of the video tampering methods is to up-convert the bitrate of a video without introducing any additional information about the video content. However, this operation does not improve the quality of the video, so we call this claimed high bitrate ‘fake bitrate’. The abused fake bitrate videos will not only mislead the viewers, but also lead to a big waste of storage space. Hence, it is necessary to propose an effective algorithm to detect fake bitrate videos.

In the process of making a fake bitrate video, the encoded video is decompressed before upconverting the bitrate and then recompressed after that. The re-encoding process may be different from that in the original video, in terms of video coding format and/or video coding parameters. Hence, the fake bitrate video is compressed twice at least. Therefore, detecting whether the video has been recompressed is a key step in detecting fake bitrate video. However, very limited work has been reported on detecting High-Efficiency Video Coding (HEVC) recompressed videos. In summary, the recompression detection methods can be divided into two categories: transformation-process-based method and prediction-process-based method. Discrete Cosine Transform (DCT) is adopted in the

transformation process by most video codecs, thus it is logical to detect double compression by using DCT coefficients. References [3,4] detected double compression in MPEG (Moving Picture Experts Group) videos based on first digit distribution of DCT coefficients. Markov features calculated from the quantized DCT coefficients were used for double MPEG-4 compression detection [5]. After HEVC standard was published, Huang et al. [6–8] extracted the statistical characteristics such as Markov and co-occurrence matrix of DCT coefficients to detect recompressed HEVC videos. By analyzing the effects of Quantization Parameters (QP) on the distributions of DCT coefficients and TU size, Li et al. [9] proposed a method to detect HEVC video double compression with different QPs.

For prediction-process-based methods, Chen et al. [10] extracted the statistical feature from macroblock mode (MBM), which consisted of macroblock types and motion vectors in P-picture to detect double MPEG compression with the same QPs. Jiang et al. [11] concatenated the statistical features of rounding and truncation errors extracted from the intra-coding process and the macroblock-mode based features obtained from the inter-coding process to detect double recompressed MPEG and H.264 videos with the same coding parameters. By analyzing the characteristics of intra-coded macroblock and skipped macroblock of recompressed videos, Variation of Prediction Footprint (VPF) features were extracted to detect MPEG-x and H.264 recompression in [12]. After that, reference [13] combined block-artifacts of P-pictures with VPF features to detect recompressed MPEG-4 and H.264 videos. Compared with previous coding standards, HEVC introduces some new techniques in the prediction process to increase coding efficiency, especially the quadtree structure. According to the characteristics of the HEVC prediction process, Costanzo and Barni [14] counted up the motion prediction modes of P-pictures and B-pictures to distinguish whether the HEVC video has undergone AVC (Advanced Video Coding) first, and Jia et al. [15] applied the number of Prediction Unit (PU) blocks with the size of  $4 \times 4$  in I-pictures to detect double-compressed HEVC video with the same QPs. Sequence of Number of Prediction Unit of its Prediction Mode (SN-PUPM) was exploited to detect HEVC recompression with different Group of Picture (GOP) sizes in [16]. Horizontal co-occurrence matrixes of DCT coefficients and the horizontal co-occurrence of PU partitioning modes in I-picture were combined in [17] to detect double HEVC compression.

To wrap up the previous HEVC recompression detection algorithms, most of them [6–9,15] were under the same or different QPs. Solutions to detect HEVC recompression with fake bitrate are rarely reported. Inspired by this, we presented an algorithm for detecting recompressed HEVC videos with fake bitrates [18]. Our previous work [18] only focused on the PU partitioning modes of the first P-picture in each GOP, and used the histogram of 25 PU partitioning modes of P-picture as the classification feature. Reference [18] confirmed that the HEVC recompression does affect the P-PU partitioning modes, while it did not comprehensively analyze other prediction variables in HEVC prediction process. However, there are six prediction variables generated in HEVC prediction process: Coding Unit of I-picture (I-CU) partitioning modes, PU of I-picture (I-PU) partitioning modes, Intra Prediction Modes (IPM) of I-picture (I-IPM), CU of P-picture (P-CU) partitioning modes, PU of P-picture (P-PU) partitioning modes, and IPM of P-picture (P-IPM). Therefore, in this paper, we will comprehensively analyze the relevant prediction variables generated in the prediction process to explore the correlation and independence among all the prediction variables. Overall, we present an algorithm to detect recompressed HEVC videos with fake bitrate by classification features extracted from prediction process. Firstly, we extract five single features, and sent them to Support Vector Machine (SVM) to verify their effectiveness. In order to enhance the expressiveness of the classification feature, single features are fused in three ways. Finally, three concatenation features are separately sent to SVM to distinguish recompressed HEVC videos with fake bitrate from single-compressed HEVC videos.

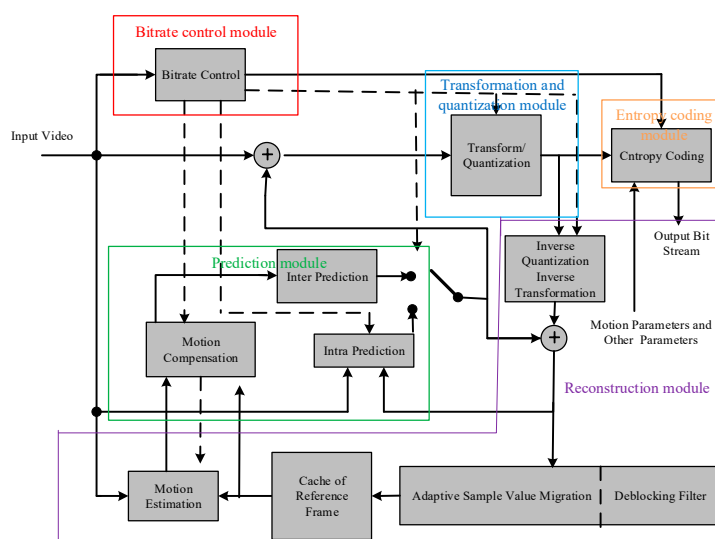
The main contributions of this paper are as follows. Firstly, we propose an efficient method to identify recompressed HEVC videos with fake bitrate, and its classification accuracy is much higher than the previous algorithms. Secondly, it is the first time, to the best of our knowledge, to comprehensively extract multiple effective prediction features from HEVC prediction process. Thirdly,

the robustness of the proposed method is considered, and experimental results show that it has strong robustness against frame deletion, copy-paste, and shifted GOP structure attacks.

The rest of the paper is organized as follows. Section 2 briefly reviews the prediction process in HEVC and continued with the analysis of several effective prediction features under different fake bitrates in Section 3. Experimental results are shown in Sections 4 and 5 concludes the paper.

## 2. Basics of Prediction Process in HEVC

In 2013, the HEVC standard was jointly ratified and published by ITU-T and ISO/IEC [19,20]. The video coding layer of HEVC employs the same hybrid approach (inter/intra prediction and 2-D transform coding) used in all video compression standards since H.261 [21]. The coding framework of HEVC is shown in Figure 1. It mainly includes bitrate control module, transformation and quantization module, prediction module, entropy coding module, and reconstruction module. It can be seen that the prediction module is a very important part of the coding process, and it is closely related to each module.



**Figure 1.** Coding framework of high-efficiency video coding (HEVC).

Prediction coding in HEVC uses the spatial and temporal correlation of the image signal to predict the current encoding pixel with its reconstructed pixel. There are two kinds of video signal prediction methods in HEVC: intra prediction and inter prediction. We use Figure 2 to describe the HEVC prediction process visually. In the prediction process of HEVC, each picture firstly performs CU and PU partitioning in a quadtree structure, and then selects the optimal IPM in each PU. The HEVC standard stipulates that I-picture only can opt intra prediction, while P-picture can adopt either intra or inter prediction.

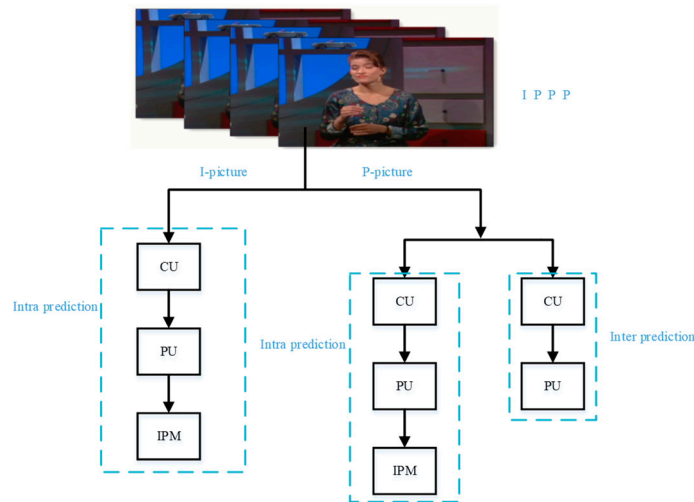


Figure 2. Prediction process of HEVC.

Using the quadtree partitioning method, the coding tree unit (CTU) is divided into CUs and then the PU partitioning modes in each CU block are determined. A  $64 \times 64$  luma of CTU can be split into multiple CUs with size from  $8 \times 8$  to  $64 \times 64$  (Figure 3a). The partitioning can be described by a quadtree structure, as shown in (Figure 3b), where the numbers indicate the index of the CUs in Figure 3a. Regardless of whether it is intra prediction or inter prediction, there are only four types of CU partitioning mode. The index of each CU partitioning modes is shown in Table 1.

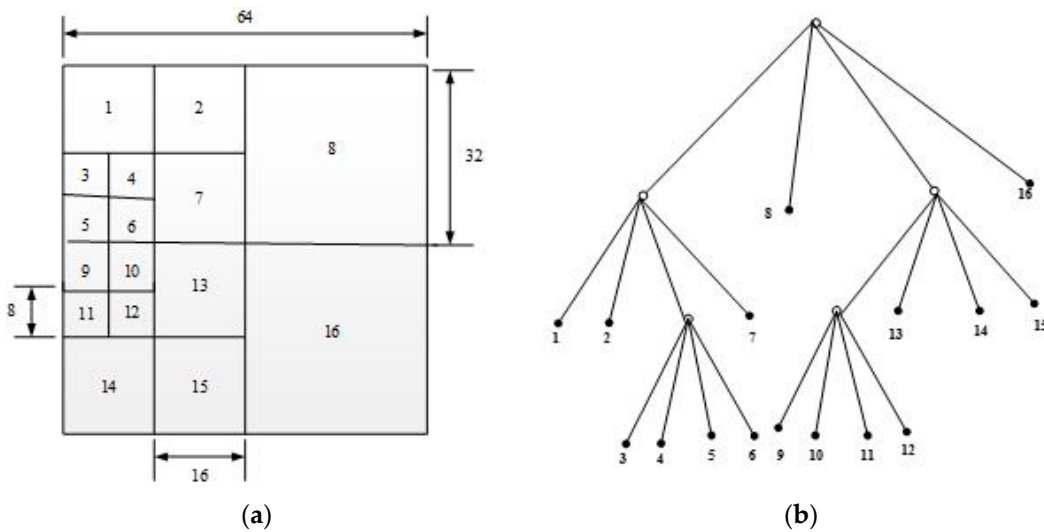


Figure 3. Example of coding tree unit (CTU) to coding unit (CU) quadtree partitioning. (a) CTU partitioning example. (b) Corresponding quadtree.

Table 1. Indexes of CU partitioning modes.

Index	Mode
1	$8 \times 8$
2	$16 \times 16$
3	$32 \times 32$
4	$64 \times 64$

The PU partitioning modes in intra prediction and inter prediction are shown in Figure 4. A CU can be symmetrically divided into one or four prediction blocks (PBs) for intra prediction. While in inter prediction, a CU can be divided into symmetric or asymmetric PBs. There are 5 PU partitioning

modes in intra prediction and 25 PU partitioning modes in inter prediction, and their indexes are listed in Tables 2 and 3.

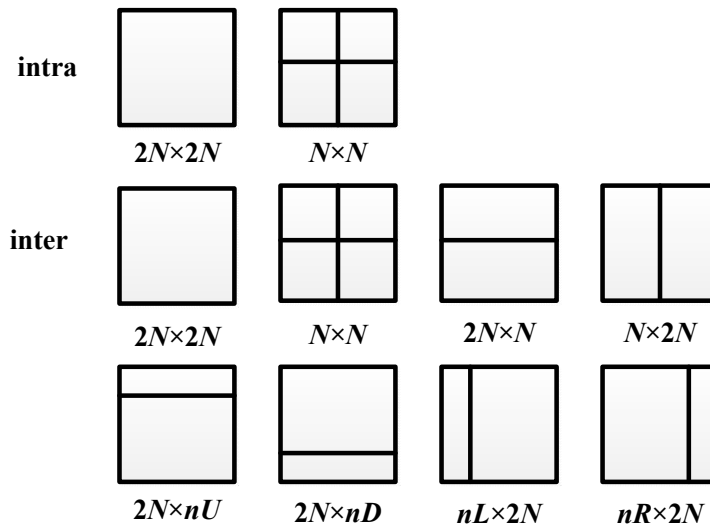


Figure 4. Prediction Unit (PU) partitioning modes in intra and inter prediction.

Table 2. Indexes of PU partitioning modes in intra prediction.

Index	Mode
1	4 × 4
2	8 × 8
3	16 × 16
4	32 × 32
5	64 × 64

Table 3. Indexes of PU partitioning modes in inter prediction.

Index	Mode	Index	Mode	Index	Mode	Index	Mode	Index	Mode
1	4 × 4	6	8 × 16	11	4 × 16	16	32 × 8	21	32 × 64
2	8 × 8	7	8 × 16	12	32 × 32	17	24 × 32	22	64 × 48
3	4 × 8	8	16 × 12	13	32 × 16	18	8 × 32	23	64 × 16
4	8 × 4	9	16 × 4	14	16 × 32	19	64 × 64	24	48 × 64
5	16 × 16	10	12 × 16	15	32 × 24	20	64 × 32	25	16 × 64

In the intra prediction process, the optimal IPM is selected in each PU. HEVC defined 35 IPM, which are DC mode, 33 angle modes, and planar mode, as shown in Figure 5. The increase of intra prediction angles makes intra prediction more accurate, thereby reducing the spatial redundancy of video more effectively.

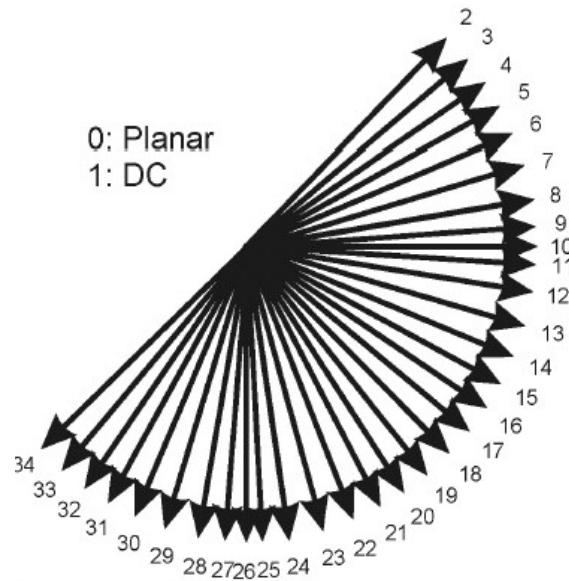


Figure 5. Intra prediction modes.

### 3. Prediction Features in HEVC Videos

In this paper, we focus on proposing an effective algorithm to distinguish recompressed HEVC videos with fake bitrate from original videos. In the HEVC video recompression process, reconstruction error and quantization error are generated in the reconstruction process and quantization process, respectively. These two kinds of errors make the decoding video lose a part of content information, which further influence prediction variables: I-CU partitioning modes, I-PU partitioning modes, I-IPM, P-CU partitioning modes, P-PU partitioning modes, and P-IPM. The theoretical model and case study of the influence of irreversible coding errors on prediction variables will be described in detail in this section.

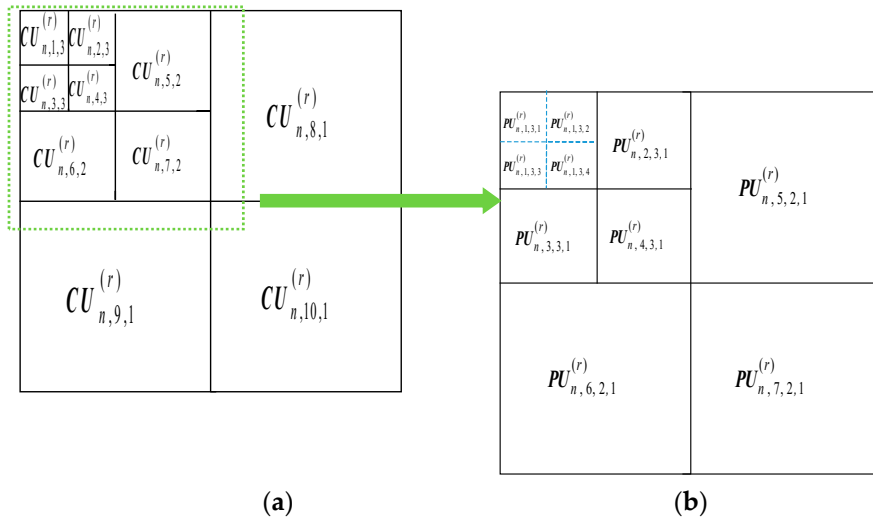
#### 3.1. Theoretical Analysis and Modeling

For an original YUV (Luma and Chroma) sequence  $V (V = \{F_n | n = 1, 2, \dots, N\})$ , where  $F_n$  denotes the  $n$ th frame and  $N$  is the total number of video frames. Given a bitrate  $r$ , the CU partitioning modes, PU partitioning modes, and IPM in I-picture and P-picture are successively determined. Let  $\rho(\cdot)$  represents the bit allocation process of rate control module, the amount of bits allocated to the  $n$ th frame can be written as  $b_n^{(r)} = \rho(n; V; r)$ . Please note that in this paper, a picture contains only one slice.

For intra prediction process  $\pi(\cdot)$ , it will select the optimal splitting of CTU into CUs, the optimal partitioning of CU into PUs, and the best IPM in PU to obtain the smallest rate-distortion value. Let  $k$  and  $K_n$  represent the  $k$ th CU in the  $n$ th frame and the total number of CUs in the  $n$ th frame, respectively. The CU partitioning sequence with  $d$  partitioning depth in the  $n$ th frame can be denoted as  $\mathbf{CU}_{\text{intra}}^{(r)} = \left\{ \mathbf{CU}_{n,k,d}^{(r)} \mid k = 0, 1, 2, \dots, K_n, d \in \{0, 1, 2, 3\} \right\}$ , where  $d = 0, 1, 2, 3$  means the CU size is  $64 \times 64, 32 \times 32, 16 \times 16, 8 \times 8$ , respectively. When the CU depth is 0 to 2, the PU partitioning mode is the same as the CU partitioning mode. When the CU depth is 3, the PU can either be  $8 \times 8$  or  $4 \times 4$ . Therefore, the PU partitioning sequence in the  $n$ th frame can be denoted as  $\mathbf{PU}_{\text{intra}}^{(r)} = \left\{ \mathbf{PU}_{n,k,d,i}^{(r)} \mid i = \{1, \dots, I\}, I = \begin{cases} 1; & \text{when } d = 0, 1, 2 \\ 1 \text{ or } 4; & \text{when } d = 3 \end{cases} \right\}$ , where  $i$  and  $I$  represent the  $i$ th PU in the  $k$ th CU and the total number of PUs contained in the  $k$ th CU, respectively. The IPM sequence in the  $i$ th PU can be represented as  $\mathbf{IPM}_{\text{intra}}^{(r)} = \left\{ \mathbf{IPM}_{n,k,d,i,j}^{(r)} \mid j \in \{1, 2, 3, \dots, 35\} \right\}$ , where  $\mathbf{IPM}_{n,k,d,i,j}^{(r)}$  denotes an IPM whose index is  $j$  of the  $i$ th PU in the  $k$ th CU whose depth is  $d$  in the  $n$ th frame.

In order to demonstrate the subscript more clearly, we draw a CTU with size of  $64 \times 64$  in the  $n$ th frame as an example. The CU sequence of the CTU is shown in Figure 6a, the PU sequence of the  $32 \times$

32 CU in the upper left corner of Figure 6a is shown in Figure 6b. The index of each CU is counted from up to down and from left to right.



**Figure 6.** CU sequence and PU sequence in the nth picture with bitrate r. (a) CU sequence in a 64 × 64 CTU. (b) PU sequence in a 32 × 32 CU.

The sequences of CU partitioning modes, PU partitioning modes, and IPMs in intra prediction process can be represented as Equation (1). We can see that CU partitioning modes, PU partitioning modes, and IPMs in intra prediction are mainly affected by the picture content and the bits allocated to the frame.

$$(IPM_{intra}^{(r)}, PU_{intra}^{(r)}, CU_{intra}^{(r)}) = \pi(F_n; b_n^{(r)}) \quad (1)$$

For inter prediction process  $f(\cdot)$ , the partition strategy is similar to intra prediction. The CU partitioning sequence with  $d$  depth in the nth frame can be represented as  $CU_{inter}^{(r)} = \{CU_{n,k,d}^{(r)} | k = 0, 1, 2, \dots, K_n, d \in \{0, 1, 2, 3\}\}$ . Different from the case in intra prediction, apart from symmetric partitioning modes, inter prediction also has asymmetric PU partitioning modes, as shown in Figure 4. Therefore, the inter PU partitioning sequence in the nth frame can be represented as  $PU_{inter}^{(r)} = \{PU_{n,k,d,i}^{(r)} | i = \{1, \dots, I\}, I = \begin{cases} 1 \text{ or } 2; & \text{when } d = 0, 1, 2 \\ 1 \text{ or } 4; & \text{when } d = 3 \end{cases}\}$ , where  $i$  and  $I$  represent the  $i$ th PU in the  $k$ th CU and the total number of PUs contained in the  $k$ th CU, respectively. When the size of the  $k$ th CU is  $64 \times 64$  or  $32 \times 32$  or  $16 \times 16$  ( $d = 0$  or  $1$  or  $2$ ), the  $k$ th CU can be symmetric or asymmetric partitioned, so  $I = 1$  or  $2$ . When  $d = 3$ , it is the same as intra prediction. There is no IPM exist in inter prediction. Thus, the CU partitioning sequence and PU partitioning sequence in inter prediction process can be represented as Equation (2), where  $C_n$  is the reference frame of  $F_n$ . That is to say, the CU partitioning modes and PU partitioning modes in inter prediction are not only determined by the content of the picture and the bits allocated to the frame, but also by the content of the reference frame.

$$(PU_{inter}^{(r)}, CU_{inter}^{(r)}) = f(F_n; b_n^{(r)}; C_n) \quad (2)$$

We use the P-PU partitioning sequence as an example to describe the difference between HEVC recompressed video with fake bitrate and single-compressed video. As shown in Figure 2, P-picture can use either intra prediction or inter prediction. That is to say, P-PU partitioning sequence in a HEVC video denoted as  $PPU^{(r)}$  is composed of PU partitioning sequence by adopting intra prediction and that by adopting inter prediction. Therefore, the P-PU partitioning sequence  $PPU^{(r)}$  can be expressed as Equation (3).

$$PPU^{(r)} = PU_{intra}^{(r)} \cup PU_{inter}^{(r)} \quad (3)$$



Assume that the first and second compression bitrates of the recompressed video are  $r_1$  and  $r_2$  ( $r_1 - r_2$  for abbreviation), respectively. Here, we only consider the fake bitrate recompression situation, that is, the bitrate up-converting, so  $r_1 < r_2$ . In the second compression, the amount of bits allocated to the  $n$ th P-frame can be written as Equation (4), where  $\hat{V}$  ( $\hat{V} = \{\hat{F}_n | n = 1, 2, \dots, N\}$ ) represents the decoded YUV sequence, and  $\hat{F}_n$  is the decompressed picture of  $F_n$ . The P-PU partitioning sequence of HEVC recompressed video is  $PPU^{(r_1, r_2)}$ .

$$b_n^{(r_1, r_2)} = \rho(n; \hat{V}; r_2) \quad (4)$$

$$PPU^{(r_1, r_2)} = PU_{\text{intra}}^{(r_1, r_2)} \cup PU_{\text{inter}}^{(r_1, r_2)} \quad (5)$$

For HEVC single-compressed video with bitrate  $r_2$ ,  $b_n^{(r_2)}$  is denoted as the amount of bits allocated to the  $n$ th P-picture, and  $PPU^{(r_2)}$  is the P-PU partitioning sequence of this single-compressed video. We can achieve

$$b_n^{(r_2)} = \rho(n; V; r_2), \quad (6)$$

$$PPU^{(r_2)} = PU_{\text{intra}}^{(r_2)} \cup PU_{\text{inter}}^{(r_2)}. \quad (7)$$

Use  $D(\cdot)$  to represent the difference between  $PPU^{(r_1, r_2)}$  and  $PPU^{(r_2)}$ . According to Equations (1)–(7), we can get the difference between the P-PU partitioning sequence of HEVC single-compressed videos and its corresponding recompressed videos with fake bitrate, as shown in Equation (8), where  $\hat{F}_n$  and  $\hat{C}_n$  are the decompressed version of  $F_n$  and  $C_n$ , respectively.

$$\begin{aligned} D(PPU^{(r_1, r_2)}, PPU^{(r_2)}) &= D((PU_{\text{intra}}^{(r_1, r_2)} \cup PU_{\text{inter}}^{(r_1, r_2)}), (PU_{\text{intra}}^{(r_2)} \cup PU_{\text{inter}}^{(r_2)})) \\ &= D((\pi(\hat{F}_n; b_n^{(r_1, r_2)}) \cup f(\hat{F}_n; b_n^{(r_1, r_2)}; \hat{C}_n)), (\pi(F_n; b_n^{(r_2)}) \cup f(F_n; b_n^{(r_2)}; C_n))) \\ &= D((\pi(\hat{F}_n; \rho(n; \{\hat{F}_n | n = 1, 2, \dots, N\}; r_2)) \cup f(\hat{F}_n; \rho(n; \{\hat{F}_n | n = 1, 2, \dots, N\}; r_2); \hat{C}_n)), \\ &\quad (\pi(F_n; \rho(n; \{F_n | n = 1, 2, \dots, N\}; r_2)) \cup f(F_n; \rho(n; \{F_n | n = 1, 2, \dots, N\}; r_2); C_n))) \end{aligned} \quad (8)$$

It can be concluded from Equation (8) that the main difference between  $PPU^{(r_1, r_2)}$  and  $PPU^{(r_2)}$  comes from the contents of  $F_n$ ,  $\hat{F}_n$ ,  $C_n$ , and  $\hat{C}_n$ . The relation between  $\hat{F}_n$  and  $F_n$  can be derived as Equation (9), where  $E(F_n)$  and  $E(C_n)$  are the quantization error of  $F_n$  and  $C_n$  under the given quantization step  $Q_p$ , respectively,  $[\ ]$  means rounding operation. Therefore, we can get  $\hat{F}_n - F_n = E(F_n) - E(C_n)$ , which means the difference between  $\hat{F}_n$  and  $F_n$  is caused by quantization error and reconstruction error. Quantization error contains rounding error and truncation error in the quantization process. Reconstruction error is caused by reference frame in the reconstruction process. It means that the difference between P-PU partitioning sequence in HEVC double-compressed video with fake bitrate and single-compressed video is mainly caused by quantization error and reconstruction error.

$$\begin{aligned} \hat{F}_n &= \text{IDCT}([\text{DCT}(F_n - C_n) / Q_p] \times Q_p) + C_n \\ &\approx \text{IDCT}([\text{DCT}(F_n) / Q_p] \times Q_p) - \text{IDCT}([\text{DCT}(C_n) / Q_p] \times Q_p) + C_n \\ &= F_n + E(F_n) - E(C_n) \end{aligned} \quad (9)$$

There are six kinds of prediction variables as illustrated in Figure 2. Similar as P-PU partitioning sequence  $PPU^{(r)}$ , we can get other prediction sequences: I-CU partitioning sequence  $ICU^{(r)}$ , I-PU partitioning sequence  $IPU^{(r)}$ , I-IPM sequence  $IIPM^{(r)}$ , P-CU partitioning sequence  $PCU^{(r)}$ , P-IPM sequence  $PIPM^{(r)}$ . And the difference between other prediction sequences of HEVC recompressed video and that of single-compressed video can be illustrated in Equations (10)–(13), respectively. Please



note that the theoretical model of  $IIPM^{(r)}$  and  $PIPM^{(r)}$  are the same because IPM only appears in the intra prediction process.

$$\begin{aligned} D(ICU^{(r_1, r_2)}, ICU^{(r_2)}) &= D(CU_{intra}^{(r_1, r_2)}, CU_{intra}^{(r_2)}) \\ &= D(\pi(\hat{F}_n; b_n^{(r_1, r_2)}), \pi(F_n; b_n^{(r_2)})) \\ &= D(\pi(\hat{F}_n; \rho(n; \{\hat{F}_n | n = 1, 2, \dots, N\}; r_2)), \pi(F_n; \rho(n; \{F_n | n = 1, 2, \dots, N\}; r_2))) \end{aligned} \quad (10)$$

$$\begin{aligned} D(PCU^{(r_1, r_2)}, PCU^{(r_2)}) &= D((CU_{intra}^{(r_1, r_2)} \cup CU_{inter}^{(r_1, r_2)}), (CU_{intra}^{(r_2)} \cup CU_{inter}^{(r_2)})) \\ &= D((\pi(\hat{F}_n; b_n^{(r_1, r_2)}) \cup f(\hat{F}_n; b_n^{(r_1, r_2)}; \hat{C}_n)), (\pi(F_n; b_n^{(r_2)}) \cup f(F_n; b_n^{(r_2)}; C_n))) \\ &= D((\pi(\hat{F}_n; \rho(n; \{\hat{F}_n | n = 1, 2, \dots, N\}; r_2)) \cup f(\hat{F}_n; \rho(n; \{\hat{F}_n | n = 1, 2, \dots, N\}; r_2); \hat{C}_n)), \\ &(\pi(F_n; \rho(n; \{F_n | n = 1, 2, \dots, N\}; r_2)) \cup f(F_n; \rho(n; \{F_n | n = 1, 2, \dots, N\}; r_2); C_n))) \end{aligned} \quad (11)$$

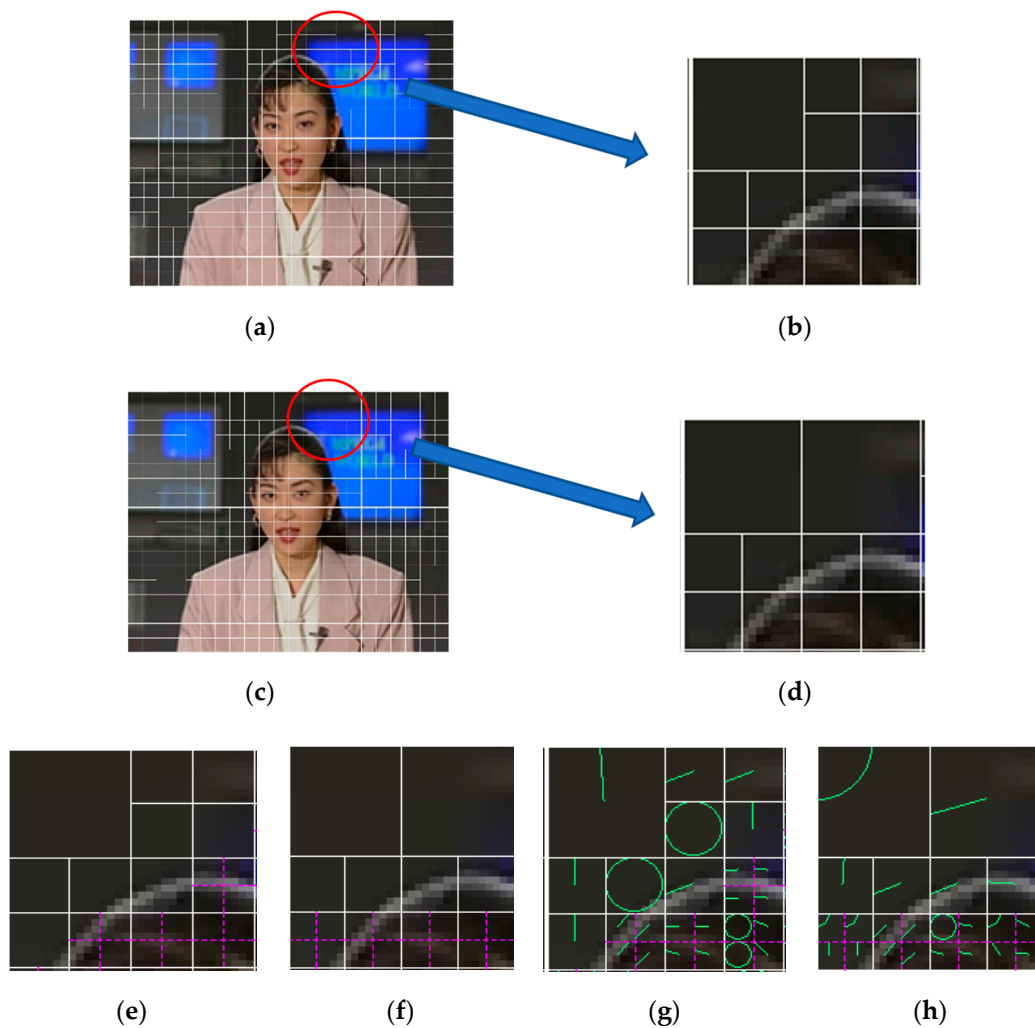
$$\begin{aligned} D(IPU^{(r_1, r_2)}, IPU^{(r_2)}) &= D(PU_{intra}^{(r_1, r_2)}, PU_{intra}^{(r_2)}) \\ &= D(\pi(\hat{F}_n; b_n^{(r_1, r_2)}), \pi(F_n; b_n^{(r_2)})) \\ &= D(\pi(\hat{F}_n; \rho(n; \{\hat{F}_n | n = 1, 2, \dots, N\}; r_2)), \pi(F_n; \rho(n; \{F_n | n = 1, 2, \dots, N\}; r_2))) \end{aligned} \quad (12)$$

$$\begin{aligned} D(IIPM^{(r_1, r_2)}, IIPM^{(r_2)}) &= D(IPM_{intra}^{(r_1, r_2)}, IPM_{intra}^{(r_2)}) \\ &= D(\pi(\hat{F}_n; b_n^{(r_1, r_2)}), \pi(F_n; b_n^{(r_2)})) \\ &= D(\pi(\hat{F}_n; \rho(n; \{\hat{F}_n | n = 1, 2, \dots, N\}; r_2)), \pi(F_n; \rho(n; \{F_n | n = 1, 2, \dots, N\}; r_2))) \end{aligned} \quad (13)$$

From Equations (10)–(13), we can conclude that the difference between the six prediction variable sequences in HEVC recompressed video and single-compressed video is mainly caused by irreversible quantization error and reconstruction error. Therefore, we consider that these six prediction variables can be used as classification features to distinguish HEVC recompression videos with fake bitrate and single ones.

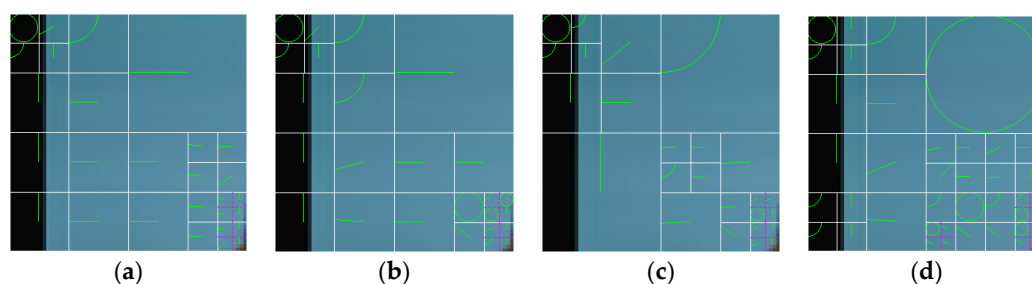
### 3.2. Feature Analysis and Example Description

To illustrate the difference between prediction variables in fake bitrate recompressed video and single-compressed video, we will extract the CU partitioning modes, PU partitioning modes, and IPM of the same I-frame within the recompressed video with fake bitrate 100–400 Kbps and within the corresponding single-compressed video with bitrate 400 Kbps, as shown in Figure 7. The 16th picture (I-picture) with CU partitioning is extracted from the double-compressed video with fake bitrate as shown in Figure 7a, and the same frame of the single-compressed video is shown in Figure 7c. Taking the  $32 \times 32$  block (shown in Figure 7b,d) surrounded by the red circle in the I-frame as an example for analysis. The block with PU partitioning modes in double- and single-compressed videos are shown in Figure 7e,f, respectively. Figure 7g,h exhibit the block with IPM in double- and single-compressed video, respectively. Observing Figure 7, it can be found that the three prediction variables (I-CU partitioning modes, I-PU partitioning modes, I-IPM) in the same I-picture of single-compressed video and double-compressed video with fake bitrate are quite different. Therefore, these three prediction variables of I-picture can be used as independent classification features for detecting fake bitrate recompression videos. In addition, even within the same I-CU partitioning block, the I-PU partitioning mode in single-compressed frame and recompressed frame is not necessarily the same, as is IPM, so these three prediction variables of I-picture are interdependent. As we can see, I-IPM is based on I-PU partitioning modes, and I-PU partitioning modes is based on I-CU partitioning modes, thus, the fusion method of concatenating these three single features can be used to enhance feature expression.



**Figure 7.** CU partitioning, PU partitioning, and intra prediction modes (IPM) of the 16th I-picture in single- and double-compressed videos. (a) CU partitioning of the recompressed 16th I-picture. (b) CU partitioning of recompressed  $32 \times 32$  block. (c) CU partitioning of the single-compressed 16th I-picture. (d) CU partitioning of single-compressed  $32 \times 32$  block. (e) PU partitioning of recompressed  $32 \times 32$  block. (f) PU partitioning of single-compressed  $32 \times 32$  block. (g) IPM of recompressed  $32 \times 32$  block. (h) IPM of single-compressed  $32 \times 32$  block.

Four  $64 \times 64$  blocks are extracted from the 7th I-picture in the “claire\_qcif.yuv” with different bitrates, as shown in Figure 8. The white line indicates the CU partitioning, the purple line indicates the PU partitioning under the CU, and the green line indicates the IPM under the PU. Figure 8a is extracted from the single-compressed video with bitrate 400 Kbps, Figure 8b is extracted from the recompressed video with bitrate 300–400 Kbps, Figure 8c is extracted from the recompressed video with bitrate 200–400 Kbps, and Figure 8d is extracted from the recompressed video with bitrate 100–400 Kbps. Compared with the single-compressed video block (Figure 8a), the I-CU partitioning, I-PU partitioning, and I-IPM in the recompressed video blocks will change, the number of changes is shown in Table 4. Both visually and statistically, it can be seen that when the difference between the first bitrate and the second bitrate of the recompressed video becomes smaller, the difference between the prediction variables of I-picture in single-compressed video and double-compressed video will become smaller, the same as the prediction variables of P-picture.

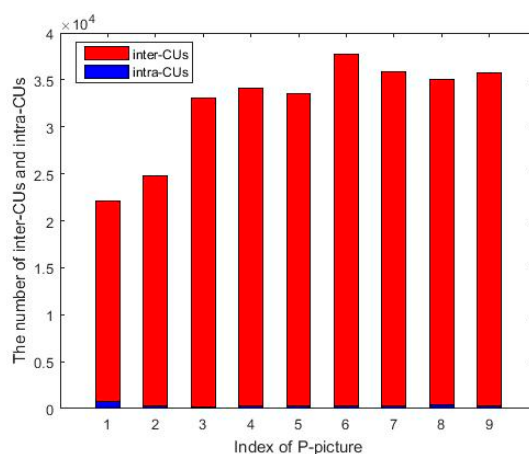


**Figure 8.** CU partitioning, PU partitioning, and IPM of the 7th I-picture in single- and double-compressed videos with different bitrates. (a) Variables in the single-compressed I-picture with bitrate 400 Kbps. (b) Variables in the recompressed I-picture with bitrate 300–400 Kbps. (c) Variables in the recompressed I-picture with bitrate 200–400 Kbps. (d) Variables in the recompressed I-picture with bitrate 100–400 Kbps.

**Table 4.** The number of changes in the variables of I-picture within the recompressed block compared to the single-compressed block from Figure 8.

	I-CU Partitioning	I-PU Partitioning	I-IPM
100–400 Kbps	3	4	18
200–400 Kbps	3	3	13
300–400 Kbps	1	1	12

In P-picture, CU block traverses the various modes of inter prediction and intra prediction, and then selects the mode with the lowest rate distortion cost as the final optimal prediction mode. While the temporal redundancy in the video is more than the spatial redundancy, the efficiency of the general inter prediction is higher than that of the intra prediction. Therefore, most of the P-pictures select inter prediction instead of intra prediction. This phenomenon is indeed ubiquitous in HEVC videos, for example, randomly selecting 9 P-pictures from a HEVC single-compressed video. Then, separately counting the number of CU blocks which adopt intra prediction (abbreviated as “intra-CUs”) and the number of CU blocks which adopt inter prediction (abbreviated as “inter-CUs”) in each P-pictures, the statistical histogram is shown in Figure 9. The red bars represent the inter-CUs and blue bars represent the intra-CUs. It can be seen that intra-CUs only take a tiny percentage of the total CUs and they are much smaller than inter-CUs. Furthermore, there are 35 IPM choices for each PU block, which makes the number of each IPM mode too small in P-pictures to be used as a single classification feature. Therefore, we ignore the P-IPM and only adopt the other two prediction variables for P-pictures: P-CU partitioning modes and P-PU partitioning modes. Similarly, the CU partitioning modes and the PU partitioning modes of P-picture are mutually dependent.



**Figure 9.** The number of CUs in random P-pictures.

In the following, we will use a line chart to show the difference between the five prediction variables in the HEVC double-compressed video and the single-compressed video, as shown in Figure 10. The bitrate of the HEVC single-compressed video is 400 Kbps (blue lines) and the corresponding recompressed video with the first compression bitrate  $r_1 = 100$  Kbps and the second compression bitrate  $r_2 = 400$  Kbps (red lines). The abscissa indicates the index of the mode used in the prediction variable, and the ordinate indicates the total number of times the mode used in the prediction variable appears in the video. It is clearly that the difference between the red and blue lines in Figure 10a–d is obvious, which means the four prediction variables (I-CU partitioning modes, I-PU partitioning modes, P-CU partitioning modes, P-PU partitioning modes) can effectively distinguish HEVC single- and double-compressed videos. In Figure 10e, we find that the number of I-IPM only differs greatly in the 1st, 2nd, 10th, 11th, 26th, and 27th intra prediction modes, and this phenomenon appears in most videos. Thus, we only use these six modes as the representatives of I-IPM.

### 3.3. The Proposed Target Features

From the above analysis, we know that the numbers of the five prediction variables are different between double-compressed videos with fake bitrate and single ones. In order to make the characteristics of the prediction variables more universal, we consider using the probability matrix of prediction variables as the classification features. The probability set can be expressed as Equations (14) and (15), where  $f_i$  denotes the probability matrix of the  $i$ th prediction variable, where  $i = 1, 2, 3, 4, 5$  represents 4-dimensional I-CU partitioning modes, 5-dimensional I-PU partitioning modes, 6-dimensional I-IPM, 4-dimensional P-CU partitioning modes, and 25-dimensional P-PU partitioning modes, respectively.  $M_i$  represents the dimensions of each prediction variable, so  $M_i = \{4, 5, 6, 4, 25 | i = 1, 2, 3, 4, 5\}$ .  $x_{ij}$  represents the number of the  $j$ th mode of the  $i$ th prediction variable in a HEVC video.  $p_{ij}$  represents the probability of the  $j$ th mode for  $i$ th prediction variable in the whole video. For example, for I-CU partitioning modes,  $i = 1$ ,  $M_1 = 4$ ,  $j = 1, 2, 3, 4$ . The number of each I-CU partition modes in the HEVC video are  $x_{1j}$ , respectively, and their probabilities are expressed as  $p_{1j} = x_{1j} / \sum_{j=1}^{M_1} x_{1j}$ , then, the probability matrix of I-CU partitioning modes is  $f_1$ . The probability matrix of other prediction variables is similar to that of I-CU partitioning modes. Finally, we can get five single features: the probability matrix of I-CU partitioning modes (PMoICU) with dimensions of 4, the probability matrix of I-PU partitioning modes (PMoIPU) with dimensions of 5, the probability matrix of I-IPM (PMoIIPM) with dimensions of 6, the probability matrix of P-CU partitioning modes (PMoPCU) with dimensions of 4, and the probability matrix of P-PU partitioning modes (PMoPPU) with dimensions of 25.

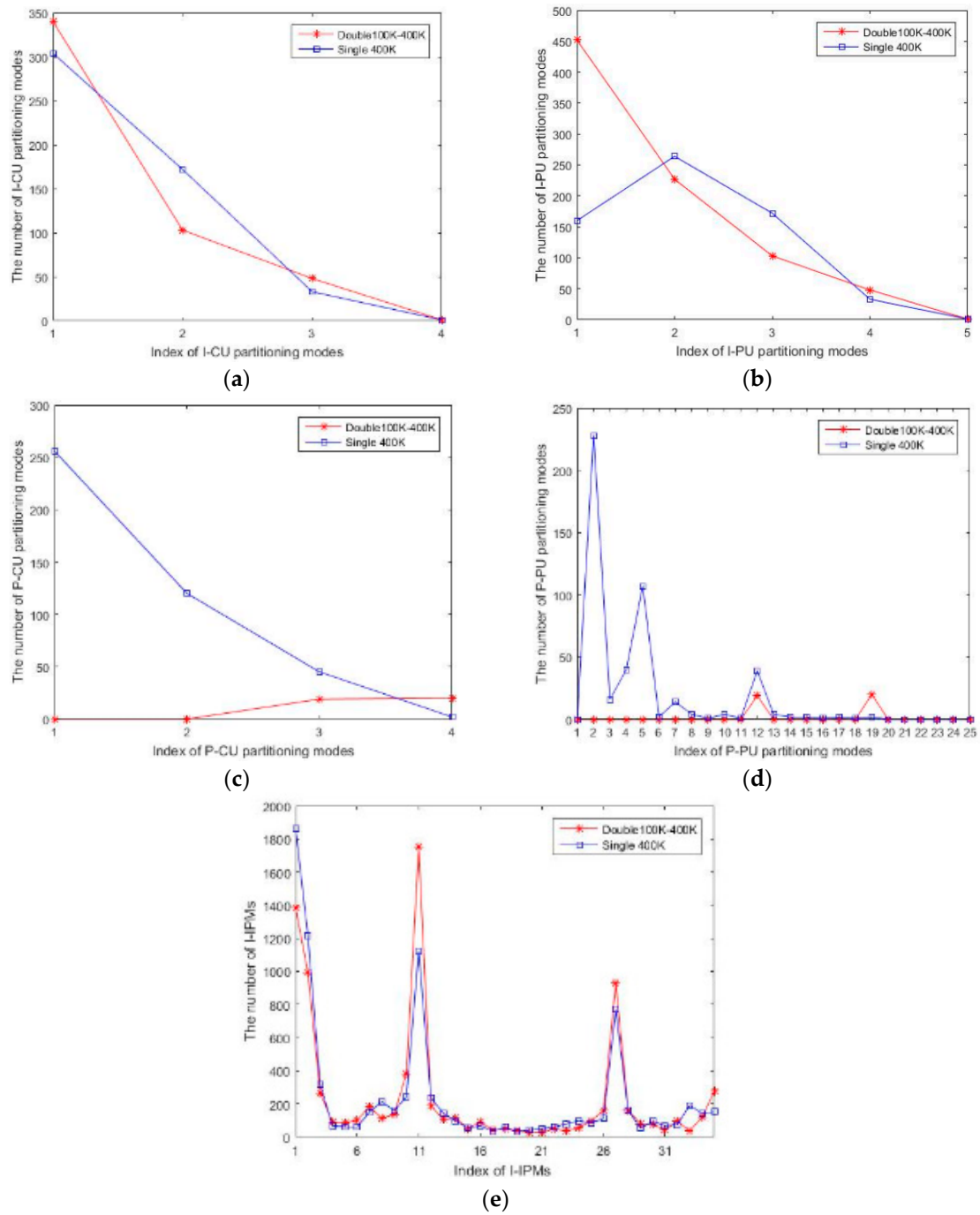
$$F = \{f_i | i = 1, 2, \dots, 5\} \quad (14)$$

$$f_i = \left\{ p_{ij} \middle| p_{ij} = x_{ij} / \sum_{j=1}^{M_i} x_{ij}, j = 1, 2, \dots, M_i \right\} \quad (15)$$

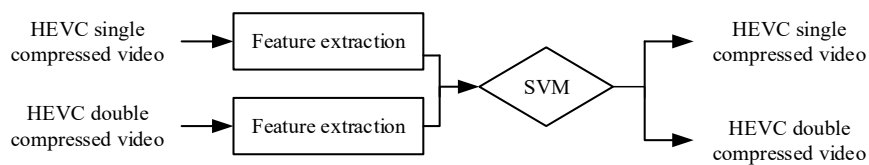
According to Section 3.2, CU partitioning modes, PU partitioning modes, and IPM are interdependent, the feature fusion method can enhance the expressiveness of single features. Therefore, we concatenate these single features from three aspects, which are called Concatenation Feature of I-picture (CFoI), Concatenation Feature of P-picture (CFoP), and Full Concatenation Feature (FCF). CFoI with dimensions of 15 is made up of PMoICU, PMoIPU, and PMoIIPM; CFoP with dimensions of 29 is composed of PMoPCU and PMoPPU, while FCF with dimensions of 44 is concatenated by all single features.

### 3.4. The Flow of the Proposed Algorithm

The flow of the proposed algorithm is shown in Figure 11. The classification feature is extracted from the HEVC video, and then the feature is sent to the SVM classifier for model training and testing to determine whether the video is single-compressed video or double-compressed video with fake bitrate.



**Figure 10.** The number of variables in a single-compressed video with bitrate 400 Kbps and its corresponding double-compressed version with bitrate 100–400 Kbps. (a) I-CU partitioning modes. (b) I-PU partitioning modes. (c) P-CU partitioning modes. (d) P-PU partitioning modes. (e) I-IPMs.



**Figure 11.** The flow of the proposed algorithm.

The process of feature extraction of the proposed algorithm is shown in Figure 12. Five prediction variables are extracted from the decoding stream of HEVC video. Based on the whole HEVC video, probabilistic statistics are performed on the five prediction variables to obtain five single features, respectively. Then, PMoICU, PMoIPU, and PMoIIPM are concatenated into CFoI, PMoPCU and PMoPPU are concatenated into CFoP, and finally, CFoI and CFoP are concatenated into the final classification feature FCF of the proposed algorithm.

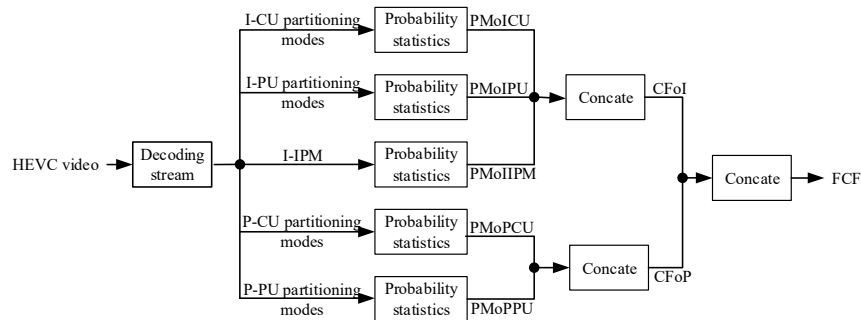


Figure 12. The process of feature extraction.

#### 4. Experimental Results

In this section, the performance of the proposed method is investigated on HEVC single- and double-compressed videos in QCIF (Quarter Common Interchange Format), CIF (Common Interchange Format), 720p, and 1080p video sets. The resolution of these four video sets are  $176 \times 144$ ,  $352 \times 288$ ,  $1280 \times 720$ , and  $1920 \times 1080$ , respectively. Each YUV sequence [22] is cut into several non-overlapping subsequences with 100 frames. Finally, we can obtain 36 QCIF videos, 43 CIF videos, 36 720p videos, and 32 1080p videos with YUV420P pixel format. The GOP size is set to 4 with the IPPP structure both in single- and double-compressed videos, and the rate control is enabled in the encoding process. Because this paper is for detecting recompressed video with fake bitrate, that is, the second bitrate of the recompressed video in our video sets is greater than the first bitrate (bitrate up-converting case), regardless of the cases of bitrate dropping or equal. To guarantee the quality of videos, for double compression video, the first compression bitrates is selected from  $r_1 = \{100 \text{ K}, 200 \text{ K}, 300 \text{ K}\}$  (bps) for QCIF and CIF video sets,  $r_1 = \{10 \text{ M}, 20 \text{ M}, 30 \text{ M}\}$  (bps) for 720p video set and  $r_1 = \{10 \text{ M}, 30 \text{ M}, 50 \text{ M}\}$  (bps) for 1080p video set, the second compression bitrate is selected from  $r_2 = \{200 \text{ K}, 300 \text{ K}, 400 \text{ K}\}$  (bps) for QCIF and CIF video sets,  $r_2 = \{20 \text{ M}, 30 \text{ M}, 40 \text{ M}\}$  (bps) for 720p video set, and  $r_2 = \{30 \text{ M}, 50 \text{ M}, 70 \text{ M}\}$  (bps) for 1080p video set. The bitrate of the corresponding single-compressed video is  $r_2$ .

The accuracy rate (AR) is used as the criterion and calculated by  $AR = (TPR + TNR)/2$ , where TPR and TNR mean true positive rate and true negative rate, respectively. For the LIBSVM [23] classifier, PolySVC kernel is selected and the optimal parameters of the classifier are obtained by a round-robin algorithm. The entire classification procedure is repeated for 20 times and the average AR is considered as the final classification accuracy. The ratio of videos in training and testing sets is set to be 5:1 for each video set, the videos are randomly assigned to training or testing set, and each training set and test set is non-overlapping. Each single prediction feature will be tested separately to verify its validity, and then the concatenation features CFoI, CFoP, and FCF will be tested immediately. The robustness of the proposed method against frame-deletion, copy-paste, and shifted GOP structure attacks will be also discussed.

##### 4.1. Single Features of I-Picture

From Section 3, we know that single features related to I-pictures are PMoICU, PMoIPU, and PMoIIPM, and their dimensions are 4, 5, and 6, respectively. In this section, their classification accuracies on four video sets are tested and shown in Tables 5–7. As we can see, though the dimensions of these three single features are low, but they are still effective in recompression detection with fake



bitrate, especially PMoIIPM, its detecting accuracies are above 85% for most situations. The results can be explained by Figure 7. Because PU partitioning mode is based on CU partitioning mode, and the IPM is based on PU partitioning mode, it means that IPM can capture the most image details among the three features. These three single features of I-picture can be considered as complementary features and can be fused to detect HEVC recompressed videos with fake bitrate. The concatenation features will be tested in Section 3.3.

**Table 5.** Classification accuracy of probability matrix of I-CU partitioning modes (PMoICU) on four kinds of video sets (in percentage).

(a) QCIF Video Set		(b) CIF Video Set		(c) 720p Video Set		(d) 1080p Video Set	
$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy
100–200 K	72.5	100–200 K	83.9	10–20 M	74.6	10–30 M	68.5
100–300 K	75.4	100–300 K	84.6	10–30 M	75.8	10–50 M	72.0
100–400 K	82.9	100–400 K	86.4	10–40 M	78.3	10–70 M	72.0
200–300 K	66.3	200–300 K	72.5	20–30 M	74.2	30–50 M	63.0
200–400 K	74.2	200–400 K	81.8	20–40 M	76.3	30–70 M	70.0
300–400 K	65.0	300–400 K	71.1	30–40 M	60.0	50–70 M	62.5

**Table 6.** Classification accuracy of probability matrix of I-PU partitioning modes (PMoIPU) on four kinds of video sets (in percentage).

(a) QCIF Video Set		(b) CIF Video Set		(c) 720p Video Set		(d) 1080p Video Set	
$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy
100–200 K	72.5	100–200 K	82.9	10–20 M	85.4	10–30 M	79.0
100–300 K	75.0	100–300 K	90.0	10–30 M	87.5	10–50 M	80.5
100–400 K	90.0	100–400 K	89.3	10–40 M	85.3	10–70 M	80.0
200–300 K	61.7	200–300 K	76.1	20–30 M	80.4	30–50 M	72.0
200–400 K	82.1	200–400 K	80.4	20–40 M	82.9	30–70 M	86.5
300–400 K	71.3	300–400 K	70.0	30–40 M	82.1	50–70 M	73.0

**Table 7.** Classification accuracy of probability matrix of I-IPM (PMoIIPM) on four kinds of video sets (in percentage).

(a) QCIF Video Set		(b) CIF Video Set		(c) 720p Video Set		(d) 1080p Video Set	
$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy
100–200 K	90.8	100–200 K	78.2	10–20 M	96.3	10–30 M	83.0
100–300 K	92.9	100–300 K	83.2	10–30 M	97.5	10–50 M	97.0
100–400 K	90.0	100–400 K	90.7	10–40 M	98.8	10–70 M	99.5
200–300 K	88.3	200–300 K	73.6	20–30 M	85.0	30–50 M	76.0
200–400 K	89.6	200–400 K	84.3	20–40 M	92.5	30–70 M	90.0
300–400 K	76.7	300–400 K	73.4	30–40 M	70.8	50–70 M	77.0

#### 4.2. Single Features of P-Picture

The single features related to P-pictures are 4-dimensional PMoPCU and 25-dimensional PMoPPU, and their classification accuracies are shown in Tables 8 and 9, respectively. For PMoPCU, its average classification accuracy on four video sets is 81%. PMoPPU performs better and its average classification accuracy is above 90% for most cases. These two single features of P-picture can be considered as complementary features and can be fused to detect HEVC recompressed videos with fake bitrate. The concatenation features will be tested in Section 4.3.



**Table 8.** Classification accuracy of probability matrix of P-CU partitioning modes (PMoPCU) on four kinds of video sets (in percentage).

(a) QCIF Video Set		(b) CIF Video Set		(c) 720p Video Set		(d) 1080p Video Set	
$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy
100–200 K	92.9	100–200 K	66.4	10–20 M	78.8	10–30 M	89.0
100–300 K	94.6	100–300 K	71.8	10–30 M	88.8	10–50 M	93.0
100–400 K	100	100–400 K	76.1	10–40 M	92.9	10–70 M	94.5
200–300 K	76.3	200–300 K	68.6	20–30 M	69.2	30–50 M	90.5
200–400 K	87.5	200–400 K	69.3	20–40 M	78.3	30–70 M	93.5
300–400 K	71.7	300–400 K	61.1	30–40 M	60.8	50–70 M	79.0

**Table 9.** Classification accuracy of probability matrix of P-PU partitioning (PMoPPU) on four kinds of video sets (in percentage).

(a) QCIF Video Set		(b) CIF Video Set		(c) 720p Video Set		(d) 1080p Video Set	
$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy
100–200 K	98.3	100–200 K	83.9	10–20 M	91.3	10–30 M	91.0
100–300 K	98.8	100–300 K	93.2	10–30 M	97.1	10–50 M	96.0
100–400 K	100	100–400 K	98.6	10–40 M	97.1	10–70 M	97.0
200–300 K	86.3	200–300 K	77.9	20–30 M	92.9	30–50 M	91.0
200–400 K	98.3	200–400 K	91.4	20–40 M	93.8	30–70 M	97.0
300–400 K	84.6	300–400 K	81.4	30–40 M	90.1	50–70 M	92.0

The statistics in Tables 5–9 show that the five single features proposed in this paper can effectively detect HEVC fake bitrate videos in both low resolution QCIF and CIF video sets and High-Definition (HD) 720p and 1080p video sets.

#### 4.3. The Concatenation Features

According to Section 3.2, concatenation can be used to enhance the expressiveness of classification feature. The classification accuracies of CFoI and CFoP are exhibited in Tables 10 and 11, respectively. It can be seen that all the classification accuracies are above 80% for both CFoI and CFoP, even reach 100% for some cases. Compared Table 10 with Tables 5–7, we can obtain that the classification accuracy of CFoI is 17.6%, 11.4% and 4.7% higher than that of single features: PMoICU, PMoIPU, and PMoIIPM, respectively. Similarly, comparing CFoP with single features PMoPCU and PMoPPU, and the average increments are 12.2% and 1.1%, respectively. It is obviously that the classification accuracy is improved a lot after the feature concatenation. Finally, the detecting results of the final classification feature of the proposed method FCF are shown in Table 12. We can see that all the classification accuracies of FCF are over 92% on four video sets, especially on the HD video sets (720p and 1080p) whose classification accuracies are all above 95%.

**Table 10.** Classification accuracy of concatenation feature of I-picture (CFoI) on four kinds of video sets (in percentage).

(a) QCIF Video Set		(b) CIF Video Set		(c) 720p Video Set		(d) 1080p Video Set	
$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy
100–200 K	88.8	100–200 K	91.1	10–20 M	95.4	10–30 M	94.5
100–300 K	88.3	100–300 K	93.6	10–30 M	97.5	10–50 M	100
100–400 K	94.6	100–400 K	92.1	10–40 M	97.9	10–70 M	100
200–300 K	83.8	200–300 K	85.7	20–30 M	88.8	30–50 M	87.5
200–400 K	92.9	200–400 K	88.6	20–40 M	94.6	30–70 M	96.0
300–400 K	80.0	300–400 K	82.5	30–40 M	87.5	50–70 M	85.5

**Table 11.** Classification accuracy of concatenation feature of P-picture (CFoP) on four kinds of video sets (in percentage).

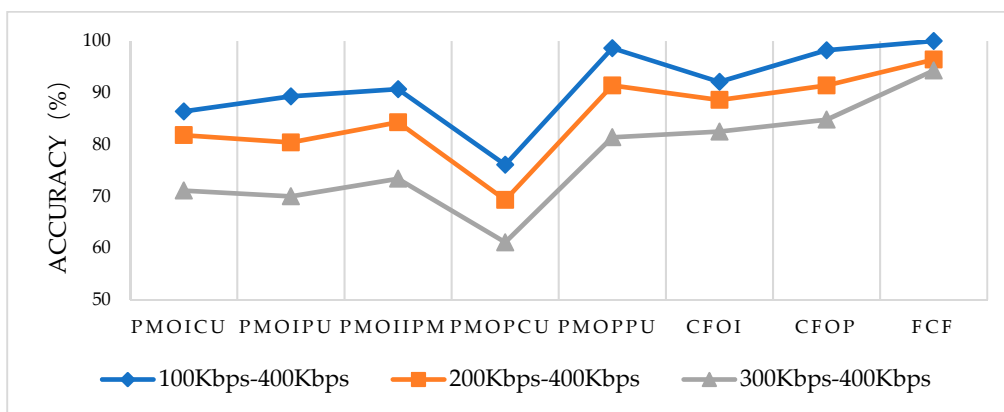
(a) QCIF Video Set		(b) CIF Video Set		(c) 720p Video Set		(d) 1080p Video Set	
$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy
100–200 K	99.2	100–200 K	88.2	10–20 M	92.9	10–30 M	92.0
100–300 K	98.3	100–300 K	93.9	10–30 M	97.1	10–50 M	96.0
100–400 K	85.0	100–400 K	98.2	10–40 M	100	10–70 M	98.5
200–300 K	97.5	200–300 K	80.0	20–30 M	97.5	30–50 M	92.0
200–400 K	98.8	200–400 K	91.4	20–40 M	97.1	30–70 M	98.0
300–400 K	83.8	300–400 K	84.8	30–40 M	93.8	50–70 M	92.5

**Table 12.** Classification accuracy of full concatenation feature (FCF) on four kinds of video sets (in percentage).

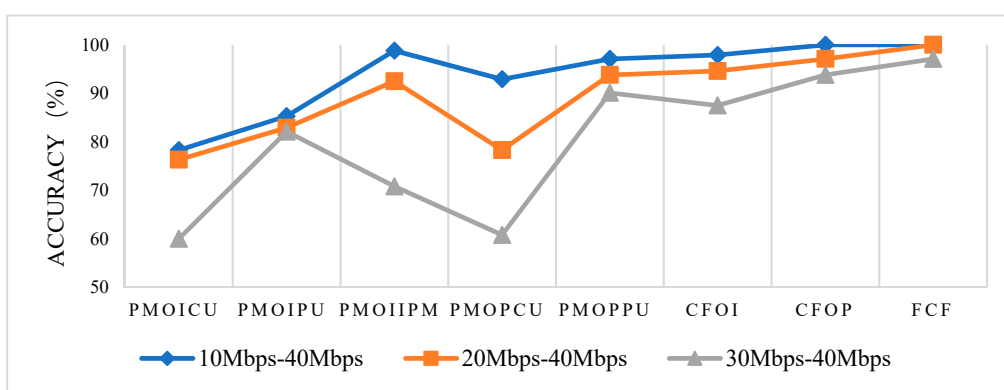
(a) QCIF Video Set		(b) CIF Video Set		(c) 720p Video Set		(d) 1080p Video Set	
$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy	$r_1 - r_2$ (bps)	Accuracy
100–200 K	100	100–200 K	95.7	10–20 M	96.7	10–30 M	96.0
100–300 K	100	100–300 K	100	10–30 M	100	10–50 M	100
100–400 K	100	100–400 K	100	10–40 M	100	10–70 M	100
200–300 K	92.0	200–300 K	93.9	20–30 M	95.8	30–50 M	98.5
200–400 K	100	200–400 K	96.4	20–40 M	100	30–70 M	100
300–400 K	92.0	300–400 K	94.3	30–40 M	97.1	50–70 M	96.0

Then, we will investigate the relationship between the classification accuracy and the bitrate difference. In CIF video sets, we compared the classification accuracies with three different bitrates, which are 100–400 Kbps, 200–400 Kbps, and 300–400 Kbps, as shown in Figure 13a. Simultaneously, the classification accuracies with three different bitrates, which are 10–40 Mbps, 20–40 Mbps, and 30–40 Mbps in 720p video sets is shown in Figure 13b. It can be found that, whether in the low-resolution video set or the high-resolution video set, the smaller the difference between the first bitrate and the second bitrate of the recompressed video, the lower the classification accuracy of each feature. It is because in this case, the difference between the prediction variables of single-compressed video and recompressed video becomes smaller, as in the legend analysis in Section 3.2.

Regarding the double compression detection of HEVC videos, some algorithms are specific to the detection of recompression with the same quantization parameters, e.g., reference [15]. Some algorithms like references [6–9] are proposed for detecting recompression with different quantization parameters. Reference [16] is proposed for detecting recompression with different GOP, and references [17,18] are proposed for detecting fake bitrate videos. In order to better demonstrate the performance of the proposed method, we will compare the proposed method with previous excellent algorithms [8,9,16–18] in different aspects, respectively. The feature extraction part for previous algorithms are implemented by using their own codes, and the SVM classifier is set as the above demonstration. Since the references [16–18] have analyzed the robustness of their detecting algorithms, so the comparison with these three algorithms will be in the subsequent robustness analysis.



(a)



(b)

**Figure 13.** The classification accuracies of five single features and three concatenation features in the video sets with different bitrates, respectively. (a) CIF video sets; (b) 720p video sets.

The team of Ningbo University proposed three classical HEVC recompressed video detection algorithms before and after 2015, the reference [8] is one of them. Reference [9] is the latest algorithm for HEVC recompression detection in 2018. The classification accuracies of the proposed method and the excellent references [8,9] under the QCIF video set are shown in Table 13. When  $r_1 - r_2$  equals 100–200 Kbps, the classification accuracy of FCF is 11.4% and 5.8% higher than references [8,9], respectively. The classification accuracy of FCF is also higher than references [8,9] under other cases. Therefore, we can conclude that the proposed method has higher classification accuracy in HEVC fake bitrate video detection than the classical algorithm [8] and the latest algorithm [9]. Furthermore, we will compare both of the classification accuracy and different kinds of robustness among the proposed method and previous works [16–18] in next three sections. Since most of the previous works only test robustness to one specific attack on one specific resolution video set In order to maintain fairness, we will compare the proposed method with the previous work by using the same attack in the same resolution video set, respectively.

**Table 13.** Classification accuracy comparison: Refs. [8,9] versus the proposed method on QCIF video set (in percentage).

$r_1 - r_2$ (bps)	100–200 K	100–300 K	100–400 K	200–300 K	200–400 K	300–400 K
[8]	88.6	92.2	92.5	85.3	90.8	76.7
[9]	94.2	95.4	98.3	90.0	92.5	86.7
FCF	100	100	100	92.0	100	92.0

#### 4.4. Robustness to Frame-Deletion

Frame-deletion is a widely used technique in digital video tampering, thus we construct two frame-deletion video sets and test the robustness of the proposed method to frame-deletion in this section. Firstly, we compress the videos with bitrate of  $r_1$  and decompress them. Then, we delete the 30th–59th frames from each decompressed video and recompress each video with bitrate of  $r_2$  ( $r_1 < r_2$ ). Finally, the frame-deleted videos can be obtained. Table 14 presents the classification accuracy of our proposed feature FCF for detecting frame-deleted videos on 1080p video set. It can be seen that all the classification accuracies of FCF reach 100% in identifying recompressed frame-deleted videos with fake bitrate. From Section 3, we know that the difference of prediction variables between single-compressed videos and recompressed ones are caused by the difference of  $\hat{F}_n$  and  $F_n$ . For frame-deleted videos, besides the difference between  $\hat{F}_n$  and  $F_n$ , frame deletion operation will reduce the number of video frames and influence the bitrate allocated to each picture, which further lead to the change of prediction variables. Thus, the recompressed frame-deleted video exhibits more significant changes in terms of FCF compared to the fake bitrate video without frame-deletion.

**Table 14.** Classification accuracy of proposed method on 1080p frame-deleted video set (in percentage).

$r_1 - r_2$ (bps)	FCF
10–30 M	100
10–50 M	100
10–70 M	100
30–50 M	100
30–70 M	100
50–70 M	100

Reference [17] showed its experimental results on the QCIF recompressed video set and QCIF frame-deletion video set, thus we compare it with the proposed method on these two video sets and the comparison results are shown in Table 15. It indicates that the accuracies of FCF are as high as that of reference [17], except the cases when the bitrates are 200–300 Kbps and 300–400 Kbps. But the classification accuracies of the proposed method on the QCIF frame-deleted video set are all 100%, which are much higher than that of reference [17]. Therefore, it can be said that the proposed method maintains high fake bitrate detection accuracy and has stronger robustness against frame-deletion attack.

**Table 15.** Classification accuracy comparison: Ref. [17] versus the proposed method on QCIF frame-deleted video set (in percentage).

$r_1 - r_2$ (bps)	QCIF Video Set		QCIF Frame-Deleted Video Set	
	[17]	FCF	[17]	FCF
100–200 K	100	100	97.1	100
100–300 K	100	100	97.5	100
100–400 K	100	100	98.3	100
200–300 K	98.8	92.0	89.2	100
200–400 K	100	100	93.3	100
300–400 K	96.7	92.0	77.1	100

#### 4.5. Robustness to Shifted GOP Structure

GOP structure is an important setting in encoding videos, and it will influence the CU partitioning modes, PU partitioning modes, and IPM. Therefore, we will test the robustness of the proposed method against shifted GOP structure attack. Reference [16] is specifically proposed for recompression detection with different GOP structures. In order to facilitate the comparison with it, we do comparison experiments on the 1080p video set, which is the same as reference [16]. The encoding

parameters of the 1080p single-compressed videos are: GOP structure IPPPPPPP and fake bitrate  $r_2$ . The recompressed 1080p video set with shifted GOP structure can be obtained as follows: after decompressing single-compressed 1080p videos with parameters: GOP structure IPPP and bitrate  $r_1$ , we recompress them with the parameters: GOP structure IPPPPPPP and recompressed fake bitrate  $r_2$  ( $r_1 < r_2$ ). The classification accuracies of the proposed method and the reference [16] on unshifted and shifted GOP structure video sets are shown in Table 16. It is obvious that whether on unshifted GOP structure video sets or shifted GOP structure video set, the accuracies of the proposed method are all much higher than that of reference [16]. Therefore, compared with reference [16], the proposed method has stronger performance in fake bitrate detection and resistance to shifted GOP structural attack.

**Table 16.** Classification accuracy comparison: Ref. [16] versus the proposed method on 1080p unshifted/shifted GOP structure video set (in percentage).

$r_1 - r_2$ (bps)	Unshifted GOP Structure		Shifted GOP Structure	
	[16]	FCF	[16]	FCF
10–30 M	78.5	96.0	99.0	100
10–50 M	81.5	100	100	100
10–70 M	74.5	100	94.5	100
30–50 M	80.0	98.5	94.0	97.0
30–70 M	80.0	100	92.5	97.5
50–70 M	65.5	96.0	92.0	97.0

#### 4.6. Robustness to Copy-Paste Tampering

Copy-paste is also a common tampering operation for digital videos. Thus, we test the robustness of the proposed method to copy-paste tampering. A copy-paste video set for 1080p video set is constructed as follows. We copy a region of the first frame in decompressed video and paste it to the 30th–59th frames, and then recompress the video at fake bitrate of  $r_2$  ( $r_1 < r_2$ ). The copied region size takes 30% of the first frame.

Our previous work [18] is the latest paper on HEVC recompression detection published in 2018, it used only one prediction variable: PU partitioning modes of P-picture. However, in the prediction process of HEVC, the upper layer of PU is CU and the lower layer of PU is IPM. Therefore, the reference [18] ignored the CU partitioning modes, which can better describe the overall complexity of the image block, and ignored the IPM, which contains the most detail information of the video content, more importantly, it ignored the information of I-picture, which is vital for HEVC videos. Therefore, by comprehensively analyzing the prediction process of HEVC, the proposed method extracts I-CU partitioning modes, I-PU partitioning modes, I-IPM, P-CU partitioning modes, and P-PU partitioning modes simultaneously. Considering that each prediction variable has different expressiveness on different video content. For example, IPM is more accurate for texture-complex regions, and CU is more suitable for smooth region, but we know that almost every video has smooth regions and complex regions, so we use the fusion method to enhance feature expression and further improve classification accuracy of the proposed method. In the following, we will comprehensively compare our method with reference [18], including computational complexity, the classification accuracy in recompressed video sets, and robustness to different video attacks.

In the proposed method, the feature extraction time and SVM training time are the main consuming time, so they are used to represent the computational complexity, and compared with that of [18]. The feature dimension of the proposed method is 44 D, and that of the reference [18] is 25 D. The feature extraction time and the time required for model training in four resolution videos are shown in Table 17. Because the feature dimension of the proposed method is slightly higher, the time complexity of the proposed method is slightly higher.

**Table 17.** Computational complexity: Ref. [18] versus the proposed method on videos.

Video	Feature Extraction Time		Training Time	
	[18]	FCF	[18]	FCF
Akiyo_qcif.yuv	0.260 s	0.379 s	0.222 s	0.247 s
bowing_cif.yuv	1.413 s	1.520 s	0.244 s	0.272 s
mobcal_720p.yuv	6.022 s	6.141 s	0.222 s	0.247 s
BasketballDrive_1080p.yuv	12.204 s	12.326 s	0.147 s	0.188 s

Reference [18] has strong robustness on resisting copy-paste, frame-deletion, and shifted GOP structure attacks. In order to better demonstrate the superiority of the proposed method, we compare it with reference [18] in four kinds of 1080p video sets, as shown in Table 18. The comparison results show that whether on the recompressed video sets or on the three kinds of attacked video sets, the classification accuracies of the proposed method are all higher than that of the reference [18], which illustrates that the proposed method not only improves the accuracy of detecting HEVC fake-bitrate recompressed videos, but also improves the robustness against frame-deletion, copy-paste, and shifted GOP structure attacks.

**Table 18.** Classification accuracy comparison: Ref. [18] versus the proposed method on 1080p video sets (in percentage).

$r_1 - r_2$ (bps)	Recompressed Video Set		Frame-Deleted Video Set		Shifted GOP Structure		Copy-Paste Video Set	
	[18]	FCF	[18]	FCF	[18]	FCF	[18]	FCF
10–30 M	92.0	96.0	97.0	100	98.5	100	92.5	97.5
10–50 M	100	100	97.5	100	99.0	100	99.5	100
10–70 M	96.5	100	98.5	100	100	100	98.5	100
30–50 M	90.0	98.5	98.0	100	97.0	97.0	94.5	98.5
30–70 M	94.0	100	97.5	100	96.0	97.5	97.0	100
50–70 M	84.5	96.0	96.5	100	94.5	97.0	93.0	98.5

## 5. Conclusions

In this paper, we have proposed a novel method to detect double-compressed HEVC videos with fake bitrates. After systematically and comprehensively analyzing the prediction process of HEVC, five effective single features are extracted from prediction process: the probability matrixes of I-CU partitioning modes, I-PU partitioning modes, I-IPM, P-CU partitioning modes, and P-PU partitioning modes. Considering the complementarity between these five single features, concatenation is adopted and three concatenation features are obtained: CFoI, CFoP, and FCF. Each kind of classification features is sent to SVM for training and testing. Experimental results show that both the five single features and the concatenation features are effective. Furthermore, compared with the state-of-art works, the proposed method not only has higher classification accuracy for detecting HEVC recompressed video with fake bitrates, but also has stronger robustness against frame-deletion, copy-paste, and shifted GOP structure attacks. In future, we will further study other HEVC coding modules and look for more effective classification features.

**Author Contributions:** Conceptualization, Z.L. (Zhaohong Li) and Z.Z.; methodology, X.L.; software, X.L.; formal analysis, Z.L. (Zhaohong Li), Z.Z. and Z.L. (Zhonghao Li); data curation, Z.L. (Zhonghao Li); writing—original draft preparation, X.L.; writing—review and editing, Z.L. (Zhaohong Li), Z.Z. and Z.L. (Zhonghao Li); supervision, Z.L. (Zhaohong Li); project administration, Z.L. (Zhaohong Li); funding acquisition, Z.L. (Zhaohong Li).

**Funding:** This research was funded by the National Natural Science Foundation of China, grant number “61702034” and the Opening Project of Guangdong Province Key Laboratory of Information Security Technology, grant number “2017B030314131”.

**Conflicts of Interest:** The authors declare no conflict of interest.



## References

1. Chowdhury, S.A.; Makaroff, D. Characterizing videos and users in YouTube: A survey. In Proceedings of the Seventh International Conference on Broadband, Wireless Computing, Communication and Applications (BWCCA), Victoria, BC, Canada, 12–14 November 2012; pp. 244–251.
2. Dezfoli, F.N.; Dehghantanha, A.; Mahmoud, R.; Sani, N.F.B.M.; Daryabar, F. Digital forensic trends and future. *Int. J. Cyber Secur. Digit. Forensics* **2013**, *3*, 48–77.
3. Chen, W.; Shi, Y.Q. Detection of double MPEG compression based on first digit statistics. In Proceedings of the International Workshop on Digital Watermarking (IWDW 2008), Busan, Korea, 10–12 November 2008; pp. 16–30.
4. Sun, T.F.; Wang, W.; Jiang, X.H. Exposing video forgeries by detecting MPEG double compression. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2012), Kyoto, Japan, 25–30 March 2012; pp. 1389–1392.
5. Jiang, X.H.; Wang, W.; Sun, T.F.; Shi, Y.Q.; Wang, S. Detection of double compression in MPEG-4 videos based on Markov statistics. *IEEE Signal Process. Lett.* **2013**, *20*, 447–450. [[CrossRef](#)]
6. Huang, M.; Wang, R.; Jian, X.; Li, Q.; Xu, D. Detection of double compression for HEVC videos based on statistical properties of DCT coefficient. *J. Optoelectron. Laser* **2015**, *26*, 733–739.
7. Huang, M.; Wang, R.; Xu, J.; Xu, D.; Li, Q. Detection of double compression for HEVC videos based on the Markov feature optimization. In Proceedings of the 12th China Inf. Hiding Multimedia Security Workshop (CIHW 2015), Wuhan, China, 7–10 October 2015; pp. 475–481.
8. Huang, M.; Wang, R.; Xu, J.; Xu, D.; Li, Q. Detection of double compression for HEVC videos based on the co-occurrence matrix of DCT coefficients. In Proceedings of the 14th Int. Workshop Digital Water-Marking (IWDW 2015), Tokyo, Japan, 7–10 October 2015; pp. 61–71.
9. Qian, L.; Wang, R.; Xu, D. Double Compression in HEVC Videos Based on TU Size and Quantized DCT Coefficients. *IET Inf. Secur.* **2018**, *7*. [[CrossRef](#)]
10. Chen, J.; Jiang, X.; Sun, T.; He, P.; Wang, S. Detecting double MPEG compression with the same quantiser scale based on MBM feature. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 2064–2068.
11. Jiang, X.; He, P.; Sun, T.; Xie, F.; Wang, S. Detection of Double Compression with the Same Coding Parameters Based on Quality Degradation Mechanism Analysis. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 170–185. [[CrossRef](#)]
12. Vazquez-Padin, D.; Fintani, M.; Bianchi, T. Detection of video double encoding with GOP size estimation. In Proceedings of the IEEE International Workshop on Information Forensics and Security, Tenerife, Spain, 2–5 December 2012; pp. 151–156.
13. He, P.S.; Jiang, X.H.; Sun, T.F.; Wang, S.L. Detection of double compression in MPEG-4 videos based on block artifact measurement. In Proceedings of the International Conference on Intelligent Computing (ICIC), FuZhou, China, 20–23 August 2015; pp. 84–96.
14. Andrea, C.; Mauro, B. Detection of double AVC/HEVC encoding. In Proceedings of the Signal Processing Conference (EUSIPCO) 2016 24th European, Budapest, Hungary, 29 August–2 September 2016; pp. 2245–2249.
15. Jia, R.; Li, Z.; Zhang, Z.; Li, D. Double HEVC compression detection with the same QPs based on the PU number. In Proceedings of the 3rd Annual International Conference Information Technology Application (ITA 2016), Hangzhou, China, 29–31 July 2016; pp. 1–4.
16. Xu, Q.; Sun, T.; Jiang, X.; Dong, Y. HEVC double compression detection based on SN-PUPM feature. In *Digital Forensics and Watermarking (IWDW 2017), Lecture Notes in Computer Science*; Kraetzer, C., Shi, Y.Q., Dittmann, J., Kim, H., Eds.; Springer: Cham, Switzerland, 2017; Volume 10431, pp. 3–17.
17. Li, Z.H.; Jia, R.S.; Zhang, Z.Z.; Liang, X.Y.; Wang, J.W. Double HEVC compression detection with different bitrates based on cooccurrence matrix of PU types and DCT coefficients. *ITM Web Conf.* **2017**, *12*. [[CrossRef](#)]
18. Liang, X.Y.; Li, Z.H.; Yang, Y.Y.; Zhang, Z.Z.; Zhang, Y. Detection of Double Compression for HEVC Videos with Fake Bitrate. *IEEE Access* **2018**, *6*, 2169–3536. [[CrossRef](#)]
19. Rec. ITU-T H.265: *High Efficiency Video Coding*. 2016. Available online: <https://www.itu.int/itu-t/recommendations/rec.aspx?rec=12905> (accessed on 12 July 2019).



20. ISO/IEC 23008-2: 2017, Information—High Efficiency Video Coding and Media Delivery in Heterogeneous Environments—Part 2: High Efficiency Video Coding. 2017. Available online: <https://www.iso.org/standard/69668.html> (accessed on 12 July 2019).
21. Sullivan, G.J.; Ohm, J.R.; Han, W.J.; Wiegand, T. Overview of the high efficiency video coding (HEVC) standard. *IEEE Trans. Circuits Syst.* **2012**, *22*, 1649–1668. [CrossRef]
22. Xiph.org Video Test Media [derf's collection]. Available online: <https://media.xiph.org/video/derf/> (accessed on 12 July 2019).
23. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2015**, *2*, 27. Available online: <http://www.csie.ntu.edu.tw/~{cjlin/libsvm> (accessed on 2 August 2015). [CrossRef]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).